MSSEG-2 challenge proceedings: Multiple sclerosis new lesions segmentation challenge using a data management and processing infrastructure

Olivier Commowick, Frédéric Cervenansky, François Cotton, Michel Dojat

MSSEG-2 challenge proceedings: Multiple sclerosis new lesions segmentation challenge using a data management and processing infrastructure

Editors

Olivier Commowick, Frédéric Cervenansky, François Cotton, Michel Dojat

http://portal.fli-iam.irisa.fr/msseg-2/

Preface

This proceedings book gathers methodological papers describing the segmentation methods evaluated at the second MICCAI Challenge on Multiple Sclerosis new lesions segmentation challenge using a data management and processing infrastructure. This challenge took place as part of an effort of the OFSEP[1] (French registry on multiple sclerosis aiming at gathering, for research purposes, imaging data, clinical data and biological samples from the French population of multiple sclerosis subjects) and FLI[2] (France Life Imaging, devoted to setup a national distributed e-infrastructure to manage and process medical imaging data). These joint efforts are directed towards automatic segmentation of MRI scans of MS patients to help clinicians in their daily practice. This challenge took place at the MICCAI 2021 conference, on September 23rd 2021.

More precisely, the problem addressed in this challenge is as follows. Conventional MRI is widely used for disease diagnosis, patient follow-up, monitoring of therapies, and more generally for the understanding of the natural history of MS. A growing literature is interested in the delineation of new MS lesions on T2/FLAIR by comparing one time point to another. This marker is even more crucial than the total number and volume of lesions as the accumulation of new lesions allows clinicians to know if a given anti-inflammatory DMD (disease modifying drug) works for the patient. The only indicator of drug efficacy is indeed the absence of new T2 lesions within the central nervous system. Performing this new lesions count by hand is however a very complex and time consuming task. Automating the detection of these new lesions would therefore be a major advance for evaluating the patient disease activity.

Based on the success of the first MSSEG challenge, we have organized a MICCAI sponsored online challenge, this time on new MS lesions detection[3]. This challenge has allowed to 1) estimate the progress performed during the 2016 - 2021 period, 2) extend the number of patients, and 3) focus on the new lesions crucial clinical marker. We have performed the evaluation task on a large database (100 patients, each with two time points) compiled from the OFSEP cohort with 3D FLAIR images from different centers and scanners. As in our previous challenge, we have conducted the evaluation on a dedicated platform (FLI-IAM) to automate the evaluation and remove the potential biases due to challengers seeing the images on which the evaluation is made.

Acknowledgments

This challenge workshop was partially supported by the France Life Imaging national project in France (ANR-11-INBS-006), and by the OFSEP national project in France (ANR-10-COHO-002). This challenge would also not have been possible without the great work of FLI, OFSEP and VIP engineers, namely: Sorina Camarasu-Pop, Axel Bonnet, Julien Louis and Michaël Kain.

[1] OFSEP: https://www.ofsep.org/en
[2] FLI: https://portal.fli-iam.irisa.fr/
[3] Challenge website: https://portal.fli-iam.irisa.fr/msseg-2/

Organization

The challenge was jointly organized by France Life Imaging (for methodological and technical resources) and the OFSEP French national cohort in MS (for data provision and medical expertise). The organization was based on three specific boards: a scientific, technical and clinical committees.

Scientific committee

Benoit Combes	Inria, Rennes, France
Olivier Commowick	Inria, Rennes, France
François Cotton	University Hospital of Lyon, France
Michel Dojat	Inserm - GIN, Grenoble, France

Technical committee

Sorina Camarasu-Pop	Université Claude Bernard, Lyon, France
Frédéric Cervenansky	Université Claude Bernard, Lyon, France
Michel Dojat	Inserm - GIN, Grenoble, France
Stéphanie Lion	OFSEP, Lyon, France
Michael Kain	FLI-IAM, Inria, Rennes, France
Arthur Masson	Inria, Rennes, France

Medical committee

François Cotton	University Hospital of Lyon, France
Jean-Christophe Ferré	University Hospital of Rennes, France
Anne Kerbrat	University Hospital of Rennes, France
Thomas Tourdias	University Hospital of Bordeaux, France

Table of Contents

Detecting new multiple sclerosis lesions using a mixed approach

Ferran Prados[1,2,3] and Baris Kanber[1,2]

[1] Centre for Medical Image Computing (CMIC), Department of Medical Physics and Biomedical Engineering, University College London, 90 High Holborn, London, WC1V 6LJ, UK
[2] NMR Research Unit, Queen Square MS Centre, Department of Neuroinflammation, UCL Queen Square Institute of Neurology, 1st Floor, Russell Square House, 10-12 Russell Square, London, WC1B 5EH, UK
[3] e-Health Center, Universitat Oberta de Catalunya, Barcelona, Spain

Abstract. MICCAI 2021 new multiple sclerosis lesions segmentation challenge (MSSEG-2) was a follow-up to the first MSSEG challenge of 2016, this time focusing on the detection of newly appearing multiple sclerosis lesions. Our solution, presented here, uses the boundary shift integral pipeline to detect new lesions, which are refined by a gradient boosting classifier. This approach gives a mean, adjusted dice score coefficient of 0.557 (standard deviation: 0.351) over the training set upon leave-one-out cross-validation.

Keywords: lesions, multiple sclerosis, challenge, segmentation

1 Introduction

Multiple sclerosis (MS) is an immune mediated, demyelinating disease of the central nervous system (CNS) often characterised by lesions observed on magnetic resonance imaging. While lesional parameters such as the number of lesions, and the total lesion volume are of great importance, other parameters such as the accumulation of new lesions are also of significant interest as they might be an indicator of whether a disease modifying therapy is working or not. While it is possible to detect MS lesions cross-sectionally and use the change in the number of lesions/the detected lesion mask as a proxy for the appearance of new lesions, this method is known to be not very accurate.

MICCAI 2021 new multiple sclerosis lesions segmentation challenge (MSSEG-2) was a follow-up to the first MSSEG challenge of 2016, focusing on the detection of newly appearing multiple sclerosis lesions when two FLAIR MRI scans belonging to two different time points are compared.

Our solution presented here uses the boundary shift integral pipeline to detect new lesions, which are refined by a LightGBM classifier.

2 Methods

The dataset of the MSSEG-2 challenge comprised MRI data of 100 MS patients. Each patient had a baseline 3D FLAIR (fluid attenuated inversion recovery) scan, as well as a 3D FLAIR follow-up scan obtained 1 to 3 years after the first one. A total of 15 MRI scanners were represented, including those operating at 1.5T and 3T:

- Three GE scanners: Optima MR450w 1.5T, SIGNA HDx 3T, SIGNA HDxt 1.5T
- Six Philips scanners: Ingenia 1.5T, 2 Ingenia 3T, 1 Achieva dStream 3T, 1 Achieva 1.5T, 1 Achieva 3T
- Six Siemens scanners: 1 Aera 1.5T, 1 Skyra 3T, 1 Verio 3T, 1 Prisma 3T, 2 Avanto 1.5T

The partitioning of the data into training and test sets was done prior to the challenge and data from 40 patients was available for algorithm development. As for the first MSSEG challenge, all images from GE scanners were withheld from the training set to assess the generalizability of the solutions. Ground-truth data was provided for 4 experts, as well as in the form of a consensus new lesions mask. We only used the latter. The 3D FLAIR scans belonging to the two points were rigidly registered to their half-way space, which was also the space of the ground truth masks that were provided.

2.1 First stage new lesions detection

The first stage of our method was performed in MNI space using all the intensity correction steps done within the boundary shift integral pipeline: N4 bias field correction (using convergence=0.0001, FWHM=0.05 and 1000 iterations), denoising (using default parameters) and multiple time-point bias field correction (using default parameters)[1]. Moreover, we computed three prior masks that we combined to define the area where we could find a new lesion: hyperintensities at baseline mask (one standard deviation above the mean brain intensity), the ventricles (using a two times eroded standard ventricles mask) and the brain tissue (from a one time eroded MNI brain mask). After all these steps, we compute the subtraction between the two images, and we keep as potential new lesions all the clusters that survive one erosion, when they were bigger than 5 voxels, that are one standard deviation above the mean difference and within the computed inclusion mask.

2.2 Mask refinement (second stage of new lesions detection)

The new lesions masks obtained in the first stage were further refined using a LightGBM gradient boosting decision tree algorithm [2]. The 3D FLAIR scan of the first time point (IMG_1), and the second time point (IMG_2), as well as the new lesions mask obtained in the first stage ($MASK_{BSI}$) were inputs

of this stage. The former two were normalised to have a zero mean and unit standard deviation. Each voxel was treated separately, and the classifier target was a binary flag indicating whether that voxel was marked as lesional in the ground truth mask. The features vector had a length of 3 (input images) $\times$ 7 (voxel features) where the latter were the normalised 3D FLAIR signal intensities for IMG_1 and IMG_2, and the first stage detection results for $MASK_{BSI}$ for the voxel of interest and its 6 immediate neighbours in three-dimensional space. A similar method was recently used to integrate multi-modal MRI data for the detection of subtle lesions in MRI-negative focal epilepsy[3]. Leave-one-out cross validation was used, while LightGBM was instructed to do early stopping based on 5-fold nested cross validation and early stopping rounds of 50. We used a modified form of the dice score coefficient (DSC_m), which incorporated a smoothing parameter (s) to gracefully handle cases where the ground truth mask was blank (see Eq. 1).

$$DSC_m = \frac{2 * \sum(y_{groundtruth} * y_{predicted}) + s}{\sum y_{groundtruth} + \sum y_{predicted} + s} \tag{1}$$

Here, y is a voxel within a mask (predicted or ground truth). Smoothing parameter s was set equal to an arbitrary small value of 10^{-6}. Post-processing was performed by removing any patches with minor axis length less than 2 voxels or total number of voxels less than 27 (3^3 voxels). Finally, the detected new-lesions mask was binarised at a threshold of 20%.

3 Results

Upon first-stage detection, the mean DSC_m was 0.373 (standard deviation: 0.329). Following refinement by the second stage and post-processing, the corresponding values were 0.557 (standard deviation: 0.351). An example case (patient 024 of the training dataset) is shown in Figure 1.

References

1. Prados F, Moccia M, Johnson A, Yiannakas M, Grussu F, Cardoso MJ, Ciccarelli O, Ourselin S, Barkhof F, Kingshott CW. Generalised boundary shift integral for longitudinal assessment of spinal cord atrophy. Neuroimage; 2020: 209.
2. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems 30 [Internet]; 2017: 3146–3154.
3. Kanber B, Vos SB, de Tisi J. Wood TC, Barker GJ, Rodionov R, Chowdhury FA, Thom M, Alexander DC, Duncan JS, Winston GP. Detection of covert lesions in focal epilepsy using computational analysis of multimodal magnetic resonance imaging data. Epilepsia; 2021; 62(3):807-816.

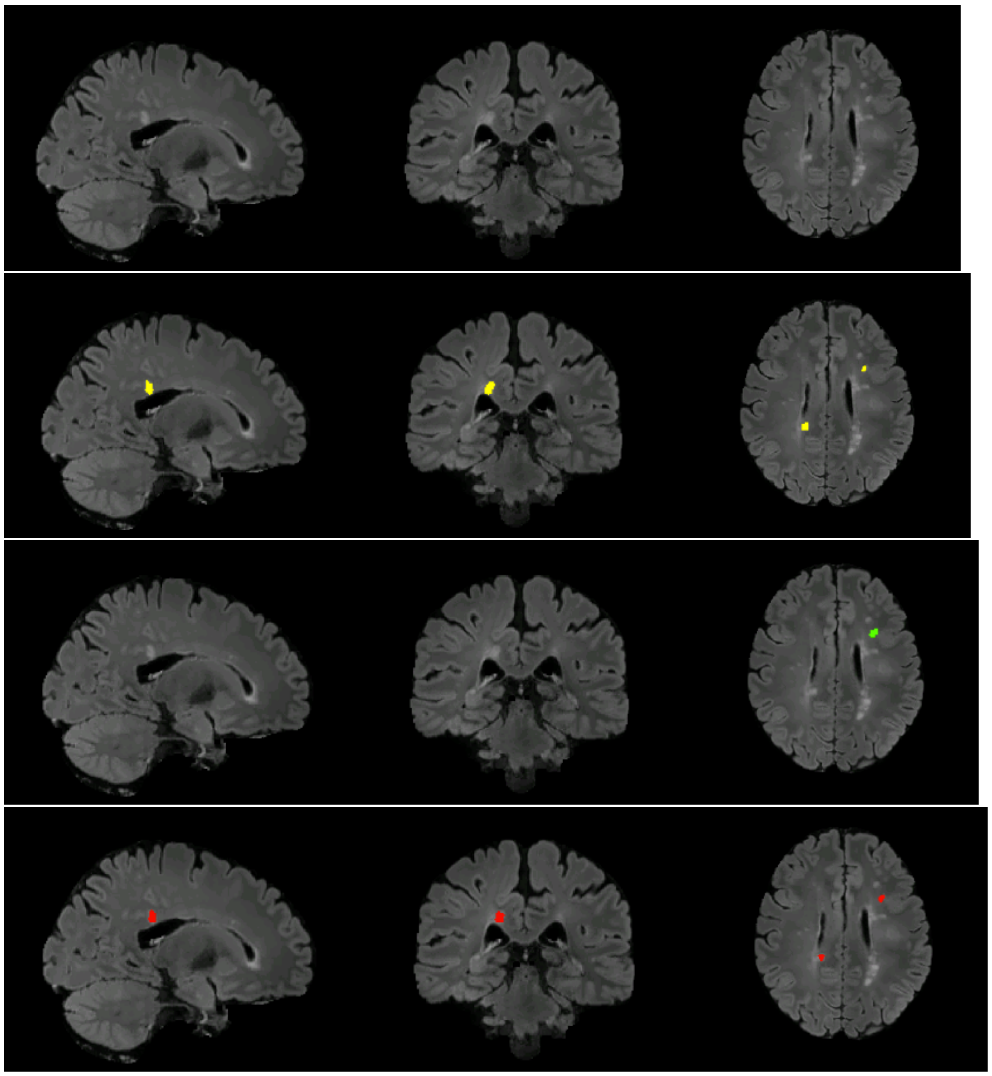

Fig. 1. From top to down: the second time-point 3D FLAIR, the ground truth (consensus new lesions mask in yellow), the output from the first stage (in green) and predicted/refined mask output by the second stage (in red).

A nnUnet implementation of new lesions segmentation from serial FLAIR images of MS patients

Arthur Masson[1], Brandon Le Bon[1], Anne Kerbrat[1,2], Gilles Edan[1,2], Francesca Galassi[1], and Benoit Combès[1,2]

[1] Univ Rennes, Inria, CNRS, Inserm IRISA UMR 6074, Empenn ERL U 1228, F-35000 Rennes, France
[2] Neurology Department, Rennes University Hospital, France

Abstract. In this short paper, we present a workflow -relying on the nnUNet framework- to segment new MS lesions from serial FLAIR images. In particular, we briefly discuss our dataset, our preprocessing steps, our nnUNet setting and our postprocessing steps.

1 Introduction

In this work, we use the nnUNet framework [2] to train a U-net [4] for the segmentation of new multiple sclerosis lesions from FLAIR MR imaging data acquired at different time steps. Overall, our choices were led to optimize detection at the lesion scale, taking only few attention to lesion delineation (that can be considered as of minor importance in a clinical context). Moreover for similar reasons, we focused our interest on patients with few new lesions. The resulting implementation choices are presented below.

2 Description of the method

In Section 2.1, we present the training data set. In Section 2.2, we present a preprocessing pipeline so that data of a given patient are appropriately aligned and signal intensity is comparable from the first FLAIR acquisition to the second one. In Section 2.3, we describe the nnUNet [2] setting we used to produce a trained 3D U-Net. The resulting network uses the pair of preprocessed FLAIR images as inputs, and outputs the softmax map associated with the presence of new lesion at each voxel. Finally in Section 2.4, we describe the post-processing steps aiming at producing a binary segmentation mask from the network softmax output.

2.1 Dataset

Our training data originates from two sources. First, we used 37 of the 40 data provided for the challenge. The remaining 3 subjects were discarded from the

training data set for assessment purposes and were not included in the final model due to time constraints. Potential missing new lesions in the dataset were identified by visually reviewing all lesions detected by a first model, learned independently from this data, and added if needed. Second, we used a supplemental set of similar data from 53 patients from our research projects from 3T Siemens and Philips scanners. Overall, our resulting training dataset contains 283 new lesions.

2.2 Preprocessing

Our preprocessing steps aim at preparing the data so that voxel-wise differences between consecutive images are as meaningful as possible. It consists in the following six steps:

1. MR volumes are reoriented in RAS coordinates,
2. skulls and skin tissues are removed using a robust registration-based brain extraction method (`animaAtlasBasedBrainExtraction`[3]),
3. the FLAIR image for the second time-point is rigidly registered on the FLAIR for the first time-point using a block matching registration method, (`animaPyramidalBMRegistration`, [1]),
4. images are all cropped using the FLAIR first time-point as a mask,
5. bias due to spatial inhomogeneity is estimated using the N4 algorithm [5] and removed from the data (`animaN4BiasCorrection`),
6. voxel intensities from acquisitions at the two time steps are jointly corrected using a twofold procedure: i) first, their joint histogram are linearly rescaled so that it best fits the y=x line in a least square sense, ii) second, a Nyul standardisation [3] on an in-house FLAIR template is applied on each acquisition independently (`animaNyulStandardization`).

2.3 Model architecture and learning

Our segmentation model consists of a 3D U-Net that was trained to segment new lesions from a pair of longitudinal preprocessed FLAIR acquisitions. This network was trained using the nnUNet framework [2]. Precisely, our 3D U-Net has 2 input channels of size [64, 160, 192] with spacing [1.0mm, 0.5mm, 0.5mm] (median training image spacing). This network was trained to minimise the sum of the Cross-Entropy loss and the Dice loss. Training was performed using a stochastic gradient descent, including a drop-out based regularization (removalRate = 0.2), run over 1000 epochs. Each epoch consists of 250 batches. Training was conducted on a GPU NVIDIA Quadro P6000, 24 GB and lasted 7 days.
Data augmentation consists of the following four alterations:

1. isotropic rescaling (ranging from 0.85 to 1.25),
2. 3D rotation (ranging from -15° to 15°),

[3] `anima.irisa.fr`

3. mirroring with respect to the sagittal plane,
4. intensity enhancements on new lesion voxels (modeling the diversity of signal change due to lesions).

2.4 Postprocessing

Our postprocessing step aims at producing binary segmentation map in the original data slab and consists of the three following steps:

1. softmax outputs map is binarized using a threshold of $p = 0.2$,
2. only connected components (26-connexity) with volume larger than 4mm^3 are kept,
3. segmentation mask is resampled to the slab and resolution from the original FLAIR image at first time step.

3 Conclusion

We provided a short description of a possible nnUNet implementation aiming at producing new MS lesions segmentation from serial FLAIR images. This implementation mainly stands on available software programs and libraries.

References

1. Commowick, O., Wiest-Daesslé, N., Prima, S.: Block-matching strategies for rigid registration of multimodal medical images. In: 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI). pp. 700–703. IEEE (2012)
2. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
3. Nyúl, L.G., Udupa, J.K., Zhang, X.: New variants of a method of mri scale standardization. IEEE transactions on medical imaging **19**(2), 143–150 (2000)
4. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
5. Tustison, N.J., Avants, B.B., Cook, P.A., Gee, J.C.: N4itk: improved n3 bias correction with robust b-spline approximation. IEEE Trans Med Imaging **29**(6), 708–711 (2010)

Segmentation of New Multiple Sclerosis Lesions in Longitudinal MRI Analysis Using a Multi-Stage 3D patch-wise Deep Learning Algorithm

Junghwa Kang[1], Siyun Jung[1], Jeongmin Yim[2], Hyebin Lee[3], Jinhee Jang[3], Yoonho Nam[1]

[1] [1]Division of Biomedical Engineering, Hankuk University of Foreign Studies, Yongin, KOREA, RPUBLIC OF
[2] College of Medicine, The Catholic University of Korea, Seoul, KOREA, REPUBLIC OF
[3] Department of Radiology, Seoul St. Mary's Hospital, Seoul, KOREA, REPUBLIC OF
kangjung9592@gmail.com

Abstract. Detecting changes in multiple sclerosis (MS) lesions through follow-up MR images is an important but time-consuming and subjective process. In this work, we propose a deep learning-based algorithm to automatically detect new MS lesions from the baseline and follow-up MR images. First, the brain and spinal cord regions were segmented. Second, the new MS lesions were segmented and classified. Finally, false positives were reduced by comparing the lesion masks from each time point. For model development, the model was trained and tuned using 40 patients (36 for training and 4 for validation). The proposed method showed the possibility in segmenting new MS lesions automatically.
Keywords: Multiple sclerosis, Lesion activity, New lesion, FLAIR

1 Introduction

Multiple Sclerosis (MS) is an autoimmune disease of the central nervous system [1]. MS is said to be disseminated in time and space, so segmentation of MS lesions in MRI scans is important in the diagnosis and follow-up of MS patients [2]. However, manual segmentation is time-consuming and subjective [3]. In this paper, we introduce a multi-step automated algorithm for the segmentation of new MS lesions on MRI scans.

2 Method

2.1 Data and Preprocessing

The data of 40 patients (36 for training and 4 for validation) were provided by the MICCAI 2021 MSSEG-2 challenge.[1] Each data consists of two 3D FLAIR images taken at different timepoints and a ground truth segmentation. To reduce the non-uniform bias field, the N4 bias field correction was conducted for preprocessing [4].

[1] https://portal.fli-iam.irisa.fr/msseg-2/data/

2.2 New MS lesion Segmentation

The pipeline of the proposed method is summarized in Fig. 1.

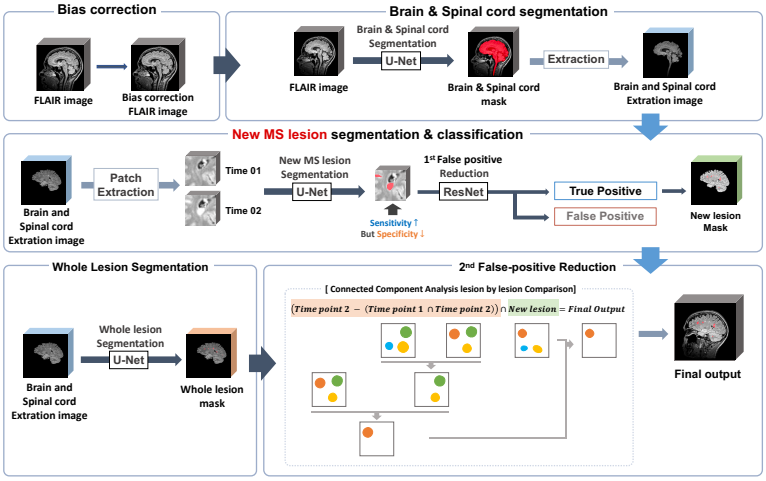

Fig. 1. Pipeline of our method

Step 1: Brain and Spinal Cord Segmentation. In the first step, we narrowed down our field of view to exclude the regions outside the brain and spinal cord. The brain mask was generated by BET from FSL and then the spinal cord regions were added by region-growing from the brain mask [5]. A 3D U-Net was used for training and input images were cropped as size of 96×96×96 3D patches [6].

Step 2: New MS Lesion Segmentation & first False-Positive Reduction. The new lesions were segmented using the patch images of two time points. Each patch size was considered to contain enough new lesions (96×96×96). As the input for model training, the baseline and follow-up patch images were concatenated, and the provided mask was used as the ground truth. As the network architecture, we used the 3D U-net [6]. Since the model was tuned to be highly sensitive during our training stage, it was necessary to reduce the false positives (FP). These FPs were firstly reduced through 3D ResNet for patch-wise classification.

Step 3: Whole MS Lesion Segmentation & Second False-Positive Reduction. The next step was segmenting whole MS lesions from each time point. The labels for the candidate lesions were initially generated by the lesion segmentation tool (LST) and then manually corrected [7]. A 3D U-Net was used for training and patch-wise segmentations were performed [6]. Comparing the new lesion mask from step 2 and the whole lesion masks from step 3, the remaining FPs were reduced and the final new lesion masks were generated. In this step, the clusters labeled by connected component analysis were compared as described in Figure 1.

3 Result

The performance of our method for new lesion segmentation was evaluated using the dice and F1 scoring method[2]. For validation cases, the proposed method showed dice and F1 scores of 0.75 and 0.68 in a patient who had the largest number of the new MS lesions (Fig. 2), and 0.74 and 0.8 in a patient who had only one new lesion, respectively.

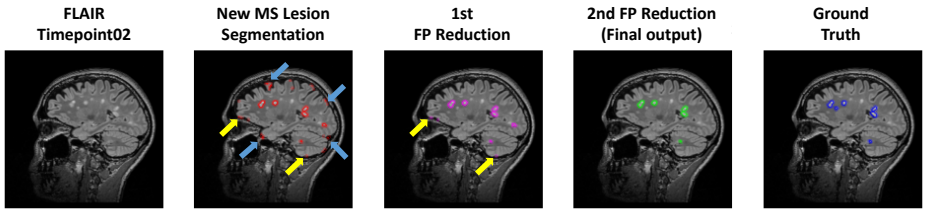

Fig. 2. Representative example. FPs were reduced by multi-stage processing (arrows).

4 Conclusion

We proposed a fully automated algorithm for detecting new MS lesions which consists of multi-stage CNN that helped focus on the area of interest by removing unnecessary areas step by step. The FP reduction method that not only focusing for the new MS legions but also focusing for the probable locations helped to classify and segment the new MS legions more robustly.

References

1. Gaetani, L., et al.: 2017 revisions of McDonald criteria shorten the time to diagnosis of multiple sclerosis in clinically isolated syndromes. Journal of neurology, 265(11), 2684-2687 (2018).
2. Christine, E., et al.: MRI FLAIR lesion segmentation in multiple sclerosis: Does automated segmentation hold up with manual annotation? NeuroImage: Clinical, 13, 264-270 (2017).
3. Edward A., et al.: Accuracy and reproducibility of manual and semiautomated quantification of MS lesions by MRI. Journal of Magnetic Resonance Imaging, 17(3), 300-308 (2003).
4. Tustison, N J., et al.: "N4ITK: improved N3 bias correction.". IEEE transactions on medical imaging, 29(6), 1310-1320 (2010).
5. Jenkinson, M., et al.: FSL. NeuroImage, 62, 782-90 (2012).
6. Ronneberger, O., et al.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J. (eds.) International Conference on Medical image computing and computer-assisted intervention – MICCAI 2015, vol.9351, pp. 234-241. Springer, LNCS (2015).
7. Schmidt, P., et al.: An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. Neuroimage, 59(4), 3774-3783 (2012).

[2] https://portal.fli-iam.irisa.fr/msseg-2/evaluation/

Siamese convolutional neural network for new multiple sclerosis lesion segmentation

Alexandre Fenneteau[1,2,3,4][0000−0001−7109−3548], Pascal Bourdon[2,3], David Helbert[2,3], Christine Fernandez-Maloigne[2,3], Christophe Habas[4,3,7], and Rémy Guillevin[3,5,6]

[1] Siemens Healthcare, Saint Denis, France
[2] XLIM Laboratory, University of Poitiers, UMR CNRS 7252; Poitiers, France
[3] I3M, Common Laboratory CNRS-Siemens, University and Hospital of Poitiers ; Poitiers, France
[4] Neuroimaging Department, Quinze Vingts Hospital; Paris, France
[5] Poitiers University Hospital, CHU; Poitiers, France
[6] DACTIM-MIS/LMA Laboratory University of Poitiers, UMR CNRS 7348; Poitiers, France
[7] University of Versailles Saint-Quentin; Versailles, France

Keywords: CNN · Segmentation · Multiple Sclerosis.

Introduction

Multiple Sclerosis (MS) is an autoimmune and inflammatory disease causing lesions in the brain and the spinal cord. The activity monitoring is usually performed visually by the radiologist but the visual detection of new lesions is subject to inter-rater variability [1].

The automatic detection of new and enlarged lesions have been conducted on intensity and deformation analysis [8]. Schmidt *et al.* proposed a method to revise lesion segmentation of two successive Magnetic Resonance (MR) exams to produce a lesion change map [12]. McKinley *et al.* used their Convolutional Neural Network (CNN) to generate segmentation confidence map of each time points to identify significant lesion change [9]. Gessert *et al.* investigated the use of adapted U-net [11] to segment lesion activity [5,4].

We decided to propose a solution designed for the segmentation of only new MS lesions based on the use of a U-net like architecture as Gessert *et al.* methods. Inspired by the work of Schmidt *et al.* and McKinley *et al.* we wanted to use both time point image segmentations for the prediction of new lesions. We conceived a model with two CNN part: one Siamese segmentation block and a refinement block.

Material and methods

Architecture

To segment new lesions, the architecture is conceived to have two parts: one segmentation block and one refinement block see Fig.1. The segmentation block

performs segmentation on both T2-fluid-attenuated inversion recovery (FLAIR) images separately, making the neural network Siamese. The refinement block concatenates segmentation maps produced for each FLAIR images to predict the final segmentation of new lesions. In practice, we cannot be sure that the neural network works as described because of the complexity of such models.

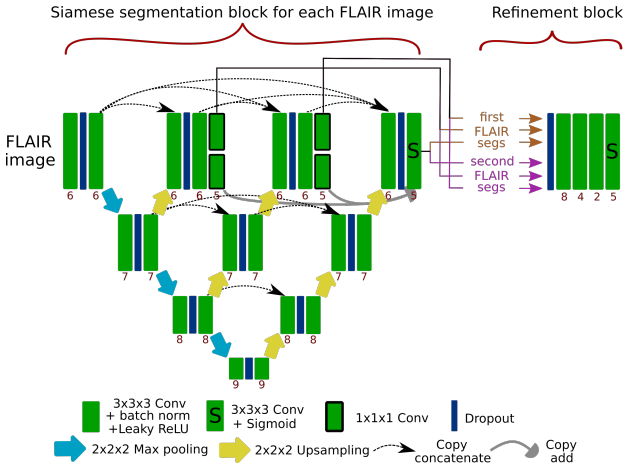

Fig. 1. Proposed architecture with the Siamese improved MPU-net in segmentation block and the convolutional refinement block.

Segmentation block

To generate lesion segmentation, the MPU-net [3] has been selected. It is a U-net [11] like architecture with only 3D convolutions, max-pooling and upsampling with approximately 30,000 learnable parameters. The light nature of this architecture is interesting since the number of learnable parameters is subject to the curse of dimensionality [2]. The restricted number of parameters also limits the training and inference costs in time and energy. Big and complex models can fit random labels [15], and so, can over-fit a whole data set. We expect light models to have a reduced capacity of memorizing a whole data set and thus can be constrained to generalize by this fact. To improve the MPU-net, batch normalization [6] has been added after each convolution and dropout layers [13] has been added in each block. Those changes improved both stability and performance of the model.

We wanted the refinement block to work with different levels of abstraction for a better analysis. The implementation of this idea has been made by fusing the MPU-net and the U-net++ architectures [16]. The U-net++ consists in adding intermediate decoders with skip-connections between them making the model deeply supervised. To limit the number of convolutional layers in the

model and to keep it light, we decided to reduce the number of consecutive convolutions in one level of encoder and decoder to 2.

Refinement block

The refinement block takes as input the segmentation maps of all decoders from the segmentation block on each FLAIR images. It is composed of a dropout layer and 4 convolution-Batch normalization layers couples.

Pre-processing

Images were previously rigidly co-registered bringing both FLAIR images to a middle point. In addition to that, the Anima MS longitudinal preprocessing script[1] was used to extract brain and apply N4 bias field correction [14]. Since images did not have the same resolution, we opt for a resampling to 0.8mm isometric. The FLAIR image histograms were then standardized [10] and each image intensity was normalized by subtracting mean and dividing by standard deviation.

Learning specifications

The architecture has been trained with FLAIR volumes from two time points to reconstruct each given segmentation (the consensus and 4 different radiologist segmentation of new lesions). We found that training with all 5 segmentation ground truths helped the neural network to converge and to find its own consensus. The volumes sent were $56 \times 56 \times 56$ voxels patches. Patches were randomly extracted in brain: uniformly for patient with no new lesions; and with a probability of 80% to contain new lesion voxels for patient with new lesions. The batch size was set to 40 and epochs were defined to contain 872 patches from each patient. The model was trained with the Adam optimizer [7] and a learning rate of 0.004 for 40 epochs. We gradually decreased learning rate by 10% after each epoch. The Dice loss is used as loss function as in [3].

Testing

On testing time, the neural network predicts 5 segmentation maps for each patch, but only the consensus prediction is kept. Patches are regularly extracted all over the volume with an overlap of 28 and spatially averaged. The prediction map is then resampled at the original resolution and thresholded to keep only voxels > 0.5 as new lesion voxel.

[1] Anima scripts: RRID:SCR_017072 https://anima.irisa.fr

References

1. Altay, E.E., Fisher, E., Jones, S.E., Hara-Cleaver, C., Lee, J.C., Rudick, R.A.: Reliability of classifying multiple sclerosis disease activity using magnetic resonance imaging in a multiple sclerosis clinic. JAMA neurology **70**(3), 338–344 (2013)
2. Bengio, Y., LeCun, Y., et al.: Scaling learning algorithms towards ai. Large-scale kernel machines **34**(5), 1–41 (2007)
3. Fenneteau, A., Bourdon, P., Helbert, D., Fernandez-Maloigne, C., Habas, C., Guillevin, R.: Investigating efficient cnn architecture for multiple sclerosis lesion segmentation. Journal of Medical Imaging **8**(1), 014504 (2021)
4. Gessert, N., Bengs, M., Krüger, J., Opfer, R., Ostwaldt, A.C., Manogaran, P., Schippling, S., Schlaefer, A.: 4d deep learning for multiple sclerosis lesion activity segmentation. arXiv preprint arXiv:2004.09216 (2020)
5. Gessert, N., Krüger, J., Opfer, R., Ostwaldt, A.C., Manogaran, P., Kitzler, H.H., Schippling, S., Schlaefer, A.: Multiple sclerosis lesion activity segmentation with attention-guided two-path cnns. Computerized Medical Imaging and Graphics **84**, 101772 (2020)
6. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
7. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
8. Lladó, X., Ganiler, O., Oliver, A., Martí, R., Freixenet, J., Valls, L., Vilanova, J.C., Ramió-Torrentà, L., Rovira, À.: Automated detection of multiple sclerosis lesions in serial brain mri. Neuroradiology **54**(8), 787–807 (2012)
9. McKinley, R., Wepfer, R., Grunder, L., Aschwanden, F., Fischer, T., Friedli, C., Muri, R., Rummel, C., Verma, R., Weisstanner, C., et al.: Automatic detection of lesion load change in multiple sclerosis using convolutional neural networks with segmentation confidence. NeuroImage: Clinical **25**, 102104 (2020)
10. Nyúl, L.G., Udupa, J.K., Zhang, X.: New variants of a method of mri scale standardization. IEEE transactions on medical imaging **19**(2), 143–150 (2000)
11. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. LNCS, vol. 9351, pp. 234–241. Springer (2015), http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a, (available on arXiv:1505.04597 [cs.CV])
12. Schmidt, P., Pongratz, V., Küster, P., Meier, D., Wuerfel, J., Lukas, C., Bellenberg, B., Zipp, F., Groppa, S., Sämann, P.G., et al.: Automated segmentation of changes in flair-hyperintense white matter lesions in multiple sclerosis on serial magnetic resonance imaging. NeuroImage: Clinical **23**, 101849 (2019)
13. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research **15**(1), 1929–1958 (2014)
14. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4itk: Improved n3 bias correction. IEEE Transactions on Medical Imaging **29**(6), 1310–1320 (June 2010). https://doi.org/10.1109/TMI.2010.2046908
15. Zhang, C., Bengio, S., Hardt, M., Recht, B., Vinyals, O.: Understanding deep learning requires rethinking generalization. arXiv preprint arXiv:1611.03530 (2016)
16. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging **39**(6), 1856–1867 (2019)

Segmentation of New Multiple Sclerosis Lesions using an Ensemble of SC U-Nets with Multi Channel Patch-Based Inputs*

Adam Gibicar[1], Samir Mitha[1], and April Khademi[1,2,3]

[1] Electrical, Computer and Biomedical Engineering Dept., Ryerson University, Toronto, ON, Canada
[2] Keenan Research Center for Biomedical Science, St. Michael's Hospital, Unity Health Network, Toronto, ON, Canada
[3] Institute for Biomedical Engineering, Science and Technology (iBEST), a partnership between St. Michael's Hospital and Ryerson University, Toronto, ON, Canada

Abstract. In this paper, we propose a pipeline to segment new lesions in 3D FLAIR sequences of multiple sclerosis (MS) patients. New lesion identification is an important part of monitoring MS progression over time and would benefit greatly from being automated due to the time-consuming nature of identifying new lesions. An ensemble of 5 SC U-Nets is used with 4-channel patch-based inputs which consist of the time1 and time2 volumes and attention maps.

Keywords: FLAIR MRI · Multiple Sclerosis · Deep learning.

1 Introduction

Multiple sclerosis (MS) is a chronic autoimmune neurological disease that attacks myelinated axons in the central nervous system (CNS). These attacks on the CNS lead to the appearance of focal lesions in the white matter. Longitudinal magnetic resonance imaging (MRI) is routinely used to detect lesions and monitor disease progression over time. Accurate detection of new MS lesions is important in determining clinical treatments and future options [4]. Identifying new lesions in MRIs can be very challenging due to their small size and subtle appearance, which leads to intra/inter observer variability in manual segmentations [2].

In this work, multi channel 2D patches are extracted from 3D FLAIR MRI, using longitudinal scans and computed attention maps. Attention maps are generated using subtraction imaging between time 1 and time 2 scans, which assists the model in learning temporal differences in MS lesions. Multi channel inputs are randomly augmented during training of 5 randomly initialized SC U-Net models [9], and predictions are aggregated to boost performance.

*Thank you to organizers of the MSSEG-2 Challenge.

2 Methods

2.1 Pre-Processing

Pre-processing was used to increase similarity between volumes. Anima pre-processing scripts [†] were used with Gaussian intensity normalization to produce brain extracted, bias corrected, and intensity normalized volumes. Negative values were removed to increase alignment and aid in removal of empty patches.

2.2 Sampling Strategy

A positive difference map is computed by subtracting time1 from time2, and thresholding the difference above 0. The difference map is then multiplied by both time points, to create corresponding attention maps. This is inspired by the work of Sepahvand et al [8], where attention is applied to time1 scans. In this work, attention is also applied to time2 scans. The use of multiple attention maps provides a multi-modal input, which has been seen to improve performance [6], and increases the quantity of training data. Both time-points and their respective attention maps are sampled using 64x64x4 patches with 50% overlap. This localizes small lesions and maintains contextual information [1]. 50% overlap was used as it provides a good trade off between performance and computational expense [3]. To deal with class imbalance, patches were sampled equally as positive or negative. All volumes were sampled in the sagittal plane. During batch generation, random affine transformations were applied.

2.3 Models

In this work, a 2D SC U-Net architecture is the base model [9]. 5 randomly initialized SC U-Nets are trained with random augmentations applied. SC U-Net implements a modified skip connection from the U-Net to help localize lesions, alleviate vanishing gradients and improve convergence speed and segmentation performance as shown by [9]. U-Net ensembles have had success in past white matter hyperintensity (WMH) segmentation works [5] [7], thus using an improved architecture may improve performance. All 5 SC U-Nets were trained using generalized dice loss.

The segmentation results from each of the models were aggregated using the probability map proposed by Li et al [5], and then binarized by thresholding at 0.5.

3 Preliminary Results

The metrics used to determine the performance of preliminary results were the metrics from the Anima segmentation performance analyzer. The specific formulas for these can be found in the MSSEG-2 evaluation criteria. The metrics are listed in Table 1. Sample pipeline predictions are shown in Fig. 1. An average of 9 false positives were predicted for volumes with no new lesions.

[†] https://anima.irisa.fr/

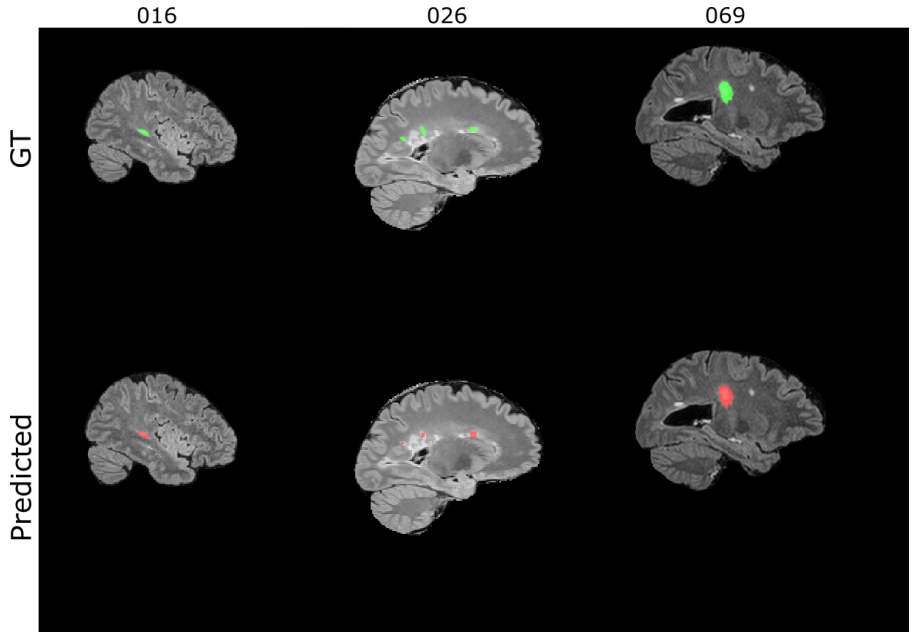

Fig. 1. Sample segmentations compared to the ground truth, for volumes 016, 026 and 069. See Table 1 for corresponding results

4 Discussion

SC U-Nets have been shown to be useful for WMH segmentation as shown by Wu et al [9]. However, this particular challenge entails detection of new WML. Based on preliminary results, we found that using a multi channel patch-based approach with SC U-Nets yields promising segmentations. The use of attention maps guided the model to learn new lesions. One limitation of using attention maps derived from the difference between volumes is that growing lesions or CSF flow-through may appear, resulting in more false positives. Another limitation is the use of brain extraction, which causes new lesions in the brain stem and/or spinal cord to be missed.

Table 1. Table containing metrics for test volumes with ground truths.

Case	DSC	PPV	Se	Sp	S	SeL	F1 Score
16	0.735	0.799	0.680	1.000	0.259	1.000	0.714
26	0.556	0.728	0.449	1.000	0.178	0.667	0.381
69	0.873	0.873	0.874	1.000	0.001	1.000	0.727
74	0.749	0.864	0.661	1.000	0.300	0.889	0.941
Mean	0.728	0.816	0.666	1.000	0.185	0.889	0.691

5 Conclusion

In conclusion, we present a method for detecting new lesions in longitudinal 3D FLAIR MRI sequences using an ensemble of 5 SC U-Nets and 2D patch-based sampling. Preliminary results showed a mean DSC of 0.728 and a limited number of false positives. The use of multiple attention maps makes the most of the limited data set by providing spatial and temporal features. Future work investigating varying loss and sampling along with post-processing may improve performance.

References

1. Bernal, J., Kushibar, K., Cabezas, M., Valverde, S., Oliver, A., Lladó, X.: Quantitative analysis of patch-based fully convolutional neural networks for tissue segmentation on brain magnetic resonance imaging. IEEE Access **7**, 89986–90002 (2019)
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. Hashemi, S.R., Salehi, S.S.M., Erdogmus, D., Prabhu, S.P., Warfield, S.K., Gholipour, A.: Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: Application to multiple sclerosis lesion detection. IEEE Access **7**, 1721–1735 (2018)
4. Kaunzner, U.W., Al-Kawaz, M., Gauthier, S.A.: Defining disease activity and response to therapy in ms. Current treatment options in neurology **19**(5), 20 (2017)
5. Li, H., Zhang, J., Muehlau, M., Kirschke, J., Menze, B.: Multi-scale Convolutional-Stack Aggregation for Robust White Matter Hyperintensities Segmentation. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 199–207. Springer (2019)
6. Narayana, P.A., Coronado, I., Robinson, M., Sujit, S.J., Datta, S., Sun, X., Lublin, F.D., Wolinsky, J.S., Gabr, R.E.: Multimodal mri segmentation of brain tissue and t2-hyperintense white matter lesions in multiple sclerosis using deep convolutional neural networks and a large multi-center image database. In: 2018 9th Cairo International Biomedical Engineering Conference (CIBEC). pp. 13–16. IEEE (2018)
7. Park, G., Hong, J., Lee, J.M.: White matter hyperintensities segmentation using unet with highlighted foreground. Tech. rep., Tech. Rep., 2020.[Online]. Available: https://wmh. isi. uu. nl/wp-content ... (2020)
8. Sepahvand, N.M., Arnold, D.L., Arbel, T.: Cnn detection of new and enlarging multiple sclerosis lesions from longitudinal mri using subtraction images. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). pp. 127–130. IEEE (2020)
9. Wu, J., Zhang, Y., Wang, K., Tang, X.: Skip connection u-net for white matter hyperintensities segmentation from mri. IEEE Access **7**, 155194–155202 (2019)

MSSEG-2 *new MS lesions* detection and segmentation challenge using a data management and processing infrastructure

Timo Löhr[1,2], Johannes C. Paetzold[1,2,3], Anjany Sekobouyina[1,5], Suprosanna Shit[1,2], Ivan Ezhov[1,2], Benedikt Wiestler[2,4], and Bjoern H. Menze[1,2,5]

[1] Departments of Informatics, Technical University Munich, Germany
[2] TranslaTUM Center for Translational Cancer Research, Munich, Germany
[3] ITERM Institute Helmholtz Zentrum München, Neuherberg, Germany
[4] Neuroradiology Department, Klinikum Rechts der Isar, Munich, Germany
[5] Department of Quantitative Biomedicine of UZH, Zurich, Switzerland
timo.loehr@tum.de

Abstract. We propose an automatic solution for the MSSEG-2 challenge to detect and segment new multiple sclerosis (MS) lesions in T2-FLAIR magnetic resonance images (MRI). New lesions arising between two time points are one of the most relevant biomarkers for clinical diagnosis and prognosis of MS disease progression. Our method relies on a 3D U-net with data augmentation; it segments all MS lesions for both given FLAIR images. We then subtract the segmentations from both time points and post-process to achieve our new lesion count and segmentation.

Keywords: MS Lesion · Detection · Segmentation.

1 Our Contribution

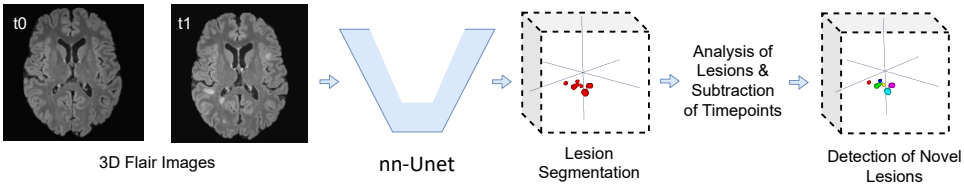

Fig. 1. Schematic overview of our proposed pipeline for segmenting new MS lesions. The 3D MRI images of both timepoints (t0 and t1) are segmented using the nnU-net. Then the segmentation results are post processed to remove small components. Subsequently, we subtract the lesions in both timepoints to detect and segment the novel lesions.

This report describes our contribution to the MSSEG-2, MS lesion segmentation challenge. The challenge has two tasks: 1) the detection of individual new lesions disregarding the accuracy of their contours and 2) the segmentation accuracy of these new lesions. At the core of our solution to both tasks is a traditional three dimensional U-Net segmentation. We treat all MS lesions in the images as a binary segmentation problem and subsequently identify which lesions appear in the later timepoint and count these individual lesions. Our network architecture is a 3D extension of the traditional and successful U-Net [1,10,9,7]. Precisely, we use the nnU-net implementation which has gathered fame in its own [2] and adapt it to our application [4,5].

2 Implementation Details

Our pipeline for segmenting new MS lesions based on the FLAIR image at two time-points consists of three stages: image pre-processing, segmentation (nnU-Net), and post-processing.

2.1 Pre-Processing

Aligning the input images, normalizing the intensity distribution and standardizing image values is the main intention of the pre-processing step. First, for each patient we co-register its images into the same spatial space. Second, we co-register the images of all patients to the MNI atlas resulting in a solution of 1x1x1mm and a size of 193x224x193 pixels. Therefore, we are able to compare between different patients and enforcing similar data characteristics for an easier training of the model. In order to focus on the brain region of the image, we skull strip every image with the HD-BET tool [3] and extract the brain region. After aligning the data and removing irrelevant data, we normalize the intensities in each image to zero mean and one standard deviation. Finally, before providing the data to our segmentation network, we standardize each image to the range [0,1].

2.2 Segmentation U-Net

Our 3D U-Net for segmenting all MS lesions has five stages with each stage having a convolutional block. Additionally, stages on the same level are connected via a skip connection. We used instance normalization and a leaky ReLU as the activation function. As a loss for the network, we used an equally weighted sum of the Dice Loss and the weighted cross entropy loss. We used on the fly data augmentations such as random flipping, random rotation along all three axes and shearing. All models are implemented in Keras using the tensorflow backend. We use the Adam optimizer with a learning rate of 10^{-4}. The network was trained for 400 epochs with a batch size of 2. Since Dice loss provides an edge at handling class imbalance [8,6] and cross-entropy loss is beneficial for smooth training convergence [11,7], we use both. The total loss function is as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{Dice} + \mathcal{L}_{CrossEnt} \tag{1}$$

2.3 Post-Processing

Based on the segmentation of both *t0* (segmentation *S0*) and *t1* (segmentation *S1*) we deduce our new lesions. We achieve this by simply subtracting the segmented lesions from *t1* with the lesions already present in *t0*. To match *S0* and *S1* we run a connected component analysis on both segmentation. In addition, we remove all components which are smaller than $10mm^3$ to ensure a relevant minimal size of each lesion.

3 Experiments and Results

We train and validate our U-Net on two different datasets.

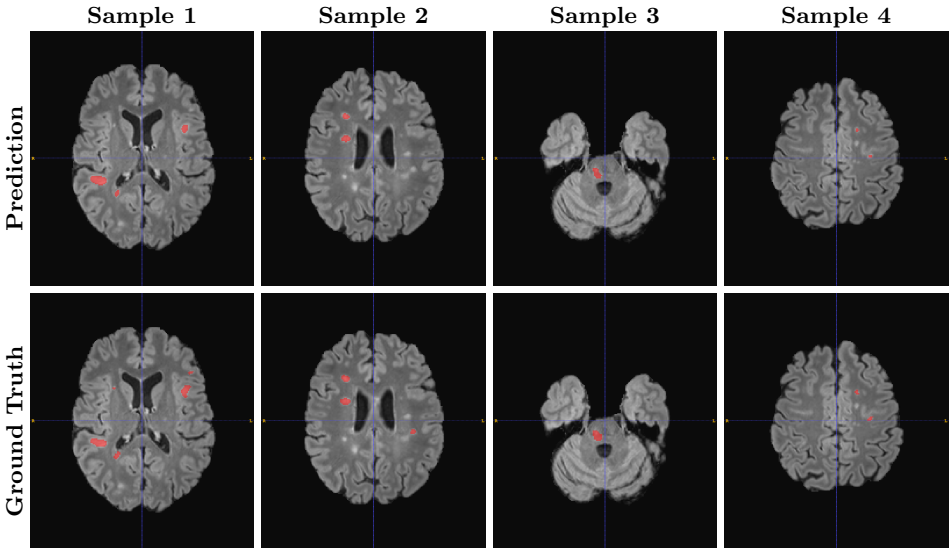

Fig. 2. Qualitative results: The first, second, and third columns represent three different examples for new lesions from different patients. In the first row, we present the prediction of our segmentation model while the second row shows the ground truth from the challenge data. We observe that our model effectively learns to detect multiple lesions at the same time.

3.1 Datasets

We obtained one dataset from the challenge including 40 patients with MRI scans at two timepoints (*t0*, *t1*), manual segmentation masks of new lesions done by four different experts and a ground truth which is a majority voting of all experts. The scans were performed on 15 different scanners (three GE scanners, six Philips and six Siemens scanner). For each patient the scan *t1* was performed between 1 to 3 years after *t0*. In addition, we train our model on another dataset with 489 patients which was collected at the University Hospital rechts der Isar in Munich. In this set each patient was scanned multiple times, an expert segmented all lesions in the first timepoint *t0* and only new lesions in the follow up scans. All scans were done with a Philips Achieva scanner.

3.2 Experiment

We train our model on the second dataset aquired at the hospital in Munich and fine-tune the weights with the provided challenge dataset. The validation set consists of 10 patients from the challenge dataset. We chose our model based on the best dice score on the validation set. We observe a dice score of 42.7 and a detection accuracy of 46.5 for our best model on the validation set. Fig. 2 shows some qualitative results on the validation dataset.

References

1. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: International conference

on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016)
2. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., et al.: nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486 (2018)
3. Isensee, F., Schell, M., Tursunova, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H., Heiland, S., Wick, W., Bendszus, M., Maier-Hein, K.H., Kickingereder, P.: Automated brain extraction of multi-sequence MRI using artificial neural networks. CoRR (2019)
4. Li, H., Loehr, T., Wiestler, B., Zhang, J., Menze, B.H.: e-uda: Efficient unsupervised domain adaptation for cross-site medical image segmentation. CoRR (2020)
5. Loehr, T., Li, H., Menze, B.: A multi-view approach for automatic segmentation of intracranial aneurysms from time of flight mras
6. Navarro, F., Shit, S., Ezhov, I., Paetzold, J., Gafita, A., Peeken, J.C., Combs, S.E., Menze, B.H.: Shape-aware complementary-task learning for multi-organ segmentation. In: International Workshop on Machine Learning in Medical Imaging. pp. 620–627. Springer (2019)
7. Paetzold, J.C., Schoppe, O., Al-Maskari, R., Tetteh, G., Efremov, V., Todorov, M.I., Cai, R., et al.: Transfer learning from synthetic data reduces need for labels to segment brain vasculature and neural pathways in 3D. In: International Conference on Medical Imaging with Deep Learning–Extended Abstract Track (2019)
8. Qasim, A.B., Ezhov, I., Shit, S., Schoppe, O., Paetzold, J.C., Sekuboyina, A., Kofler, F., Lipkova, J., Li, H., Menze, B.: Red-GAN: Attacking class imbalance via conditioned generation. yet another medical imaging perspective. vol. 1, p. 13 (2020)
9. Shit, S., Ezhov, I., Paetzold, J.C., Menze, B.H.: Aν-net: Automatic detection and segmentation of aneurysm. In: CADA@ MICCAI. pp. 51–57 (2020)
10. Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P., Bauer, U., Menze, B.H.: cldice-a novel topology-preserving loss function for tubular structure segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16560–16569 (2021)
11. Tetteh, G., Efremov, V., Forkert, N.D., Schneider, M., Kirschke, J., et al.: Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes. arXiv preprint arXiv:1803.09340 (2018)

New MS Lesion Segmentation using Deep Residual Attention Gate U-Net using 2D slices of 3D MR Images

Beytullah Sarica[1][0000-0002-6660-8508] and Dursun Zafer Seker[2][0000-0001-7498-1540]

[1,*] Istanbul Technical University, Graduate School, Department of Applied Informatics Istanbul, Turkey
[2] Istanbul Technical University, Faculty of Civil Engineering, Department of Geomatics Engineering, Istanbul, Turkey
*saricab@itu.edu.tr, seker@itu.edu.tr

Abstract. Multiple sclerosis (MS) is a disease that damages the brain and central nervous system. In MS, the myelin sheath covering nerve fibers is attacked by the immune system which leads to communication issues between the brain and the rest of the body. Thus, the nerves can be either damaged or deteriorated by MS disease. MRI data has been widely used for disease diagnosis, patient follow-up, and monitoring of therapies, namely, having information about MS history. Comparing one time point to another using MRI data for delineating new MS lesions has been an important marker rather than finding the total number or volume of MS lesions in recent years. Thus, automatic segmentation of these new lesions emerging on the second time-point would help physicians to evaluate the patient disease activity. Recently, Deep Learning algorithms have been showing a great performance for the segmentation task in the medical field. In this study, an encoder-decoder architecture combining the standard U-Net architecture, deep residual learning, and Attention gate (AG) was proposed to detect new lesions on the second time-point MRI using 2D slices obtained from 3D MR images. The MSSEG-2 challenge provides a total of 100 MS patients' 3D MR images for the FLAIR modality where 40 of them are in the training set and 60 of them are in the test set, respectively. To process these 3D MR images, the proposed method and a slice-based approach were used together to segment new lesions on the target time point rather than using a 3D segmentation approach requiring more computational power and learning parameters.

Keywords: Deep residual learning, U-Net, Attention gate, Convolutional neural networks, MS activity lesion segmentation, Multiple sclerosis.

1 Data Preparation

The MSSEG-2 challenge dataset provides two raw 3D MR images for each patient where each has different dimensions. Firstly, the brain skull was removed from the raw data since MS lesions appear in the white matter of the brain. Then, early fusion was performed on the feature-level coming from both time points as an input for the used network architecture. 3D MRI consists of orthogonal plane orientations which provide

three views based on its plane orientations. The axial, sagittal, and coronal along the x, y, and z axes can be obtained as 2D slices. Each 3D MRI has different resolutions, so zero-padding was used to make 512x512 2D slice dimensions. The first time-point and second-time point were stacked together to create 2-channel feature mapping from each plane orientation. After that, all 2D slices extracted from all directions were concatenated together to create a single training input and to use the contextual information from all directions. Fig. 1 shows the raw and preprocessed input data for two time points with delineated mask data.

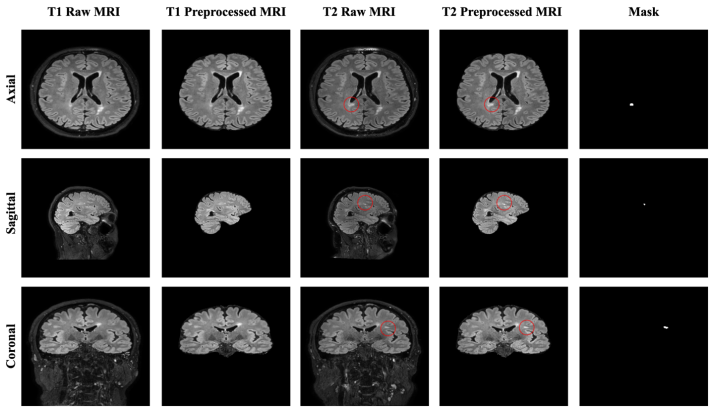

Fig. 1. The raw, preprocessed, and delineated mask images including two-time points for the MS segmentation task.

2 Implementation Details

There are 40 3D MR images in the training set and only 29 of them have new lesions in their second time-point. We split 29 images into the training and validation sets by 24 and 5, respectively. To prepare input data, each 3D image was divided into its orthogonal plane orientations, namely, axial, sagittal, and coronal views. 2 channel input feature data was created using each corresponding 2D slice from both time points as discussed in the previous section. Keras [1] and TensorFlow [2] libraries were used for the model development. The Google Colaboratory [3], which has a Tesla K80 GPU, was used for the training procedure. Focal and Dice loss were used together in order to handle unbalance labeled data between lesion and background since lesion pixels are a very small portion of the whole image. Data augmentation such as flipping and rotating was used on the fly using Keras data generator. Fig. 2 shows the whole pipeline for new lesion segmentation for MS activity. Firstly, 3D MRIs were converted into their plane orientations along the x, y, and z axes. Then, 2D slices of two-time points were fused together to create a single input training data for the deep residual U-Net model with Attention gate (AG). Predicted 2D slices based on the axial, sagittal, and coronal views were converted into the 3D binary segmentation output using the majority voting for each view.

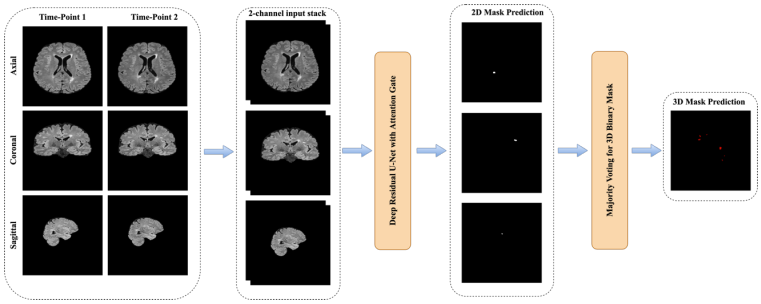

Fig. 2. The pipeline of new MS lesion segmentation using a slice-based approach.

3 The Proposed Model Implementation

The standard U-Net [4], which is an encoder-decoder network with skip connections, has shown competitive results in the medical field. Also, Deep residual learning [5] uses several residual blocks together where an identity mapping is created earlier to handle the performance problem, and skip connections are used to overcome vanishing gradients as well. Attention gate [6] is used during concatenating skip connection and up sampling to focus more features related to sizes and shapes on the target structure. [7] , which concatenates the standard U-Net and Deep residual learning, is one of the main sources of inspiration of this study. In this study, the standard U-Net, Deep residual learning, and Attention gate were used together for the MS lesion activity task. This U-Net architecture compromises encoding, bridge, and decoding parts. In the encoding part, the feature map from each layer is down sampling by halving the size. In the decoding part, the corresponding encoding part, which has a higher dimension, is concatenated with the up sampling of the feature maps, which comes from the lower dimension. The bridge part is responsible for connecting these parts. AG was added between the corresponding encoding part and the up sampling of features maps coming from the lower level. Figure 3 shows the detail of the designed network using the input data formed by the axial, sagittal, and coronal views extracted from each 3D MRI from two-time points.

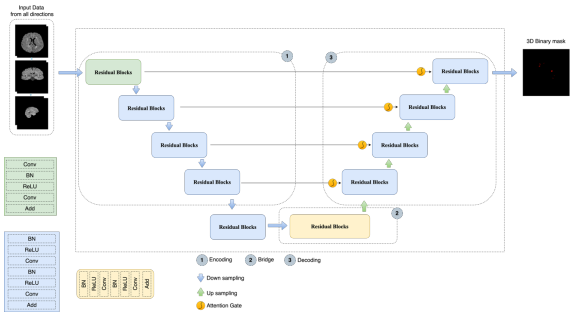

Fig. 3. The proposed model using the standard U-Net, Deep residual learning, and Attention gate.

References

1. Chollet, F., et al. (2015) Keras. https://github.com/fchollet/keras.
2. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16); 2016; pp. 265–283.
3. Bisong, E. Google Colaboratory. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*; Apress: Berkeley, CA, 2019; pp. 59–64 ISBN 978-1-4842-4470-8.
4. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical image computing and computer-assisted intervention; Springer, 2015; pp. 234–241.
5. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition; 2016; pp. 770–778.
6. Oktay, O.; Schlemper, J.; Le Folgoc, L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning where to look for the pancreas. *arXiv* **2018**.
7. Zhang, Z.; Liu, Q.; Wang, Y. Road extraction by deep residual u-net. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 749–753.

New MS lesion Segmentation with Lesion-wise Metrics Learning

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, the performance of automatic Multiple Sclerosis segmentation methods has been evaluated using lesion-wise detection metrics (*e.g.,* lesion-wise true positive rate). However, due to the complexity of implementing a differentiable loss function based on such metrics, most deep learning methods are trained with traditional voxel-wise metrics such as the Dice similarity coefficient. In this paper, to address this issue, we propose to use a convolutional neural network to estimate lesion-wise metrics from ground truth and predicted masks. Combined with the Dice Similarity Coefficient, this lesion-wise metric estimation model is used to optimize our new lesions segmentation method.

Keywords: Metric Learning · New lesion segmentation · Detection Metrics.

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. Automating new lesion detection can alleviate the clinicians workflow. To assess such automatic methods and compare them to expert segmentation, the research community relies on several segmentation metrics. Recent works [3] question the relevance of voxel-wise metrics (such as Dice) compared to detection metrics, which are used for MS diagnostic and clinical evaluation of the patient evolution. Besides, other works [6] suggest that multiple complementary metrics are needed to provide a better understanding of the automatic method performance. Deep learning methods have shown encouraging results in the task of MS segmentation [1, 2]. Many segmentation methods are optimized by maximizing the Dice between the prediction and ground truth. To the best of our knowledge, no prior work has successfully implemented a deep learning loss function based on lesion-wise metrics. In this paper, we propose a novel metric learning framework designed for lesion-wise metric optimization.

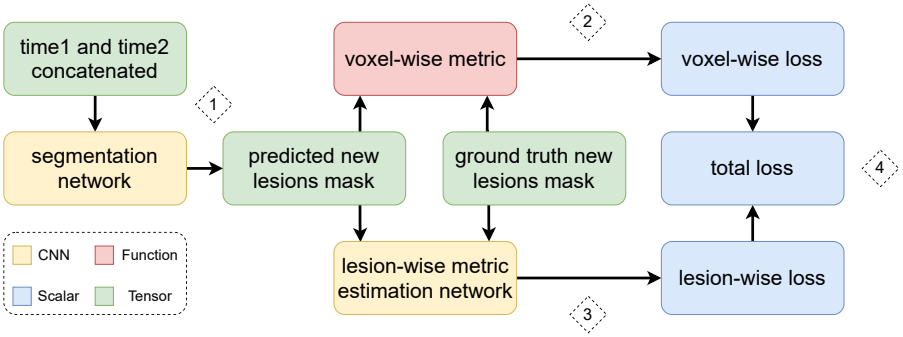

Fig. 1. The training framework

2 Method and Material

2.1 Method overview

The proposed method uses a Convolutional Neural Network (CNN) to estimate lesion-wise metrics. The CNN is used in addition to classic voxel-wise metrics to train a segmentation model. As shown in Fig. 1, the training of our new MS lesion segmentation network is performed with a 4 step cycle. First, the segmentation network is fed with a concatenation of 3D FLAIR patches relative to the two time points. As an output, the network predicts the new lesions mask. Second, a voxel-wise loss is computed from the predicted and the ground-truth new lesions masks. This segmentation error is estimated using the Dice similarity coefficient:

$$VoxelWise_{loss} = 1 - Dice(pred, true) \tag{1}$$

Third, the lesion-wise metric estimation network produce a lesion-wise loss term from the predicted and the ground-truth masks. The training of the metric estimation network is performed separately (see lesion-wise metric estimation network). Indeed, the weights of this network are frozen during the training of the segmentation network.

$$LesionWise_{loss} = 1 - F(pred, true) \tag{2}$$

where F(pred,true) is the metric estimation for the prediction and ground-truth masks.

Fourth, both voxel-wise and lesion-wise loss terms are combined into a total loss:

$$Total_{loss} = VoxelWise_{loss} + \alpha \times LesionWise_{loss} \tag{3}$$

This aggregated term is back propagated through the metric estimation network and back to the segmentation network. The gradient estimated at the level of the segmentation model can be used to update its weights using an optimization method (e.g. gradient descent).

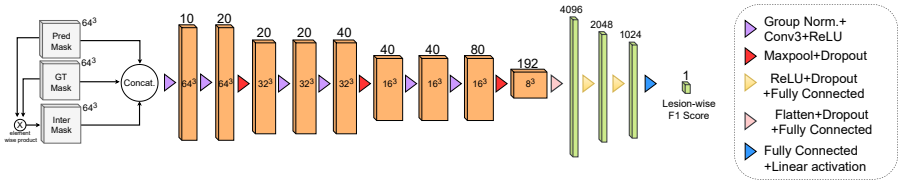

Fig. 2. The lesion-wise metric estimation block

2.2 Lesion-wise metric estimation network

The proposed network is a CNN that predicts an approximation of lesion-wise metrics from probability masks. As shown in the Fig. 2, the network is composed of convolutional layers and fully connected layers. This model is trained with mask pairs (*i.e.* representing the predicted mask and the ground-truth mask) and their respective lesion-wise metric. For better network optimization and metric estimation, the mask pairs should be selected to produce a balanced distribution of output values in the metric defined range (*i.e.* in our case, a balanced distribution of values in the range [0,1]). It is worth mentioning that the masks are augmented to simulate continuous probability maps. Specifically, masks are altered with gaussian blur and small random intensity changes. For this work, the estimation network predicts the lesion-wise F1 score (defined by the challenge organizers [9]) as the following:

$$LF_1 = \frac{2 \times LTPR \times LPPV}{LTPR + LPPV} \tag{4}$$

where LTPR (lesion-wise true positive rate) is the rate of ground truth lesions that intersect with predicted lesions and LPPV (lesion-wise positive predictive value) is the rate of predicted lesions that intersect with ground truth lesions. The mean squared error is used to train this estimation network.

2.3 Data

The dataset provided by the MICCAI 2021 - Longitudinal Multiple Sclerosis Lesion Segmentation Challenge [4] was used to train our method. For preprocessing, our strategy used the docker [5] built with the Anima scripts[3] proposed by the challenge organizers. It includes bias correction, denoising and skull stripping.

3 Implementation details

For the new lesion segmentation, a 3D U-Net architecture similar to the one proposed by [7] has been selected. As input, the network receives a concatenation of FLAIR patches of size 64^3 from the two times points. The model output

[3] anima.irisa.fr

predicts the new lesion mask. Image quality data augmentation [7] is used when training the new lesion segmentation model. The models are optimized with Adam [8] using a learning rate of 0.0001 and a momentum of 0.9. We found empirically that $\alpha = 0.2$ is a good tradeoff between Dice and the estimation of LF_1.

References

1. 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)
2. Commowick, O., Cervenansky, F., Ameli, R.: Msseg challenge proceedings: Multiple sclerosis lesions segmentation challenge using a data management and processing infrastructure (2016)
3. Commowick, O., Istace, A., Kain, M., Laurent, B., Leray, F., Simon, M., Pop, S.C., Girard, P., Ameli, R., Ferré, J.C., et al.: Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. Scientific reports **8**(1), 1–17 (2018)
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. García-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D.L., Collins, D.L.: Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. Medical image analysis **17**(1), 1–18 (2013)
7. 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)
8. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
9. the evaluation criteria for the MSSEG-2 challenge: https://portal.fli-iam.irisa.fr/files/2021/06/MS_Challenge_Evaluation_Challengers.pdf

Draw and Erase to Learn Better

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. In order to segment new Multiple Sclerosis (MS) lesions, automatic methods need longitudinal data to learn the joint information contained in both time points. However, longitudinal dataset for MS are expensive and difficult to build. In this paper, we propose an innovative pipeline to simulate two realistic time points with new MS lesion from a single FLAIR scan.

Keywords: Synthetic data · New lesion segmentation · Lesion generation.

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. Recently, the automation of MS segmentation has shown encouraging results. In some conditions, many techniques showed performance comparable to clinicians. Those methods use a single time point scan to segment all appearing lesions at the time of the acquisition. These techniques are not specialized in the detection of new lesions. Indeed, they repeatedly run the segmentation process for each time-point to detect the new appearing lesions. Unlike the human reader, these methods do not jointly exploit the information contained at each time point. In order to propose automatic methods specifically designed for the scenario of new lesions detection, it requires a longitudinal dataset of MS lesions with an evolution in their lesions. The organizers of the MICCAI 2021 - Longitudinal Multiple Sclerosis Lesion Segmentation Challenge [5] provided such a dataset. However, training state-of-the-art deep learning algorithms and achieving generalizing results on unseen domains [10, 1, 11] may require more data. Thus, we propose in this paper an innovative pipeline for generating synthetic training data suited to the scenario of new lesions detections. First, we propose a lesion generator model to simulate new lesions on healthy white matter regions. Second, we propose a lesion inpainting model to simulate healthy tissue on regions with existing lesions. Finally, we combine the lesion generator and the lesion inpainting to inpaint two realistic time points from a single FLAIR scan.

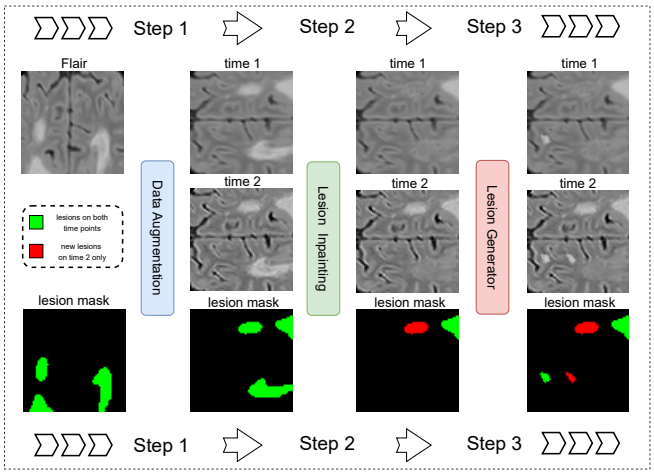

Fig. 1. Synthetic time points with new MS lesion generation pipeline

2 Method and Material

2.1 Method Overview

The proposed method is based on the simulation of new MS lesions between two time points. As shown in Fig1, our pipeline generates "on the fly" synthetic 3D FLAIR patches corresponding to two time points and their respective new lesion masks from a single FLAIR image. The synthetic data is generated in 3 steps. In the first step, a 3D FLAIR patch and its MS lesion mask are randomly sampled from MS lesion segmentation datasets [3, 4, 2]. Then, this FLAIR patch and lesion mask are randomly augmented with flipping and rotations. A copy of the FLAIR patch is performed to represent the two time points. Then, both identical patches are altered with different data augmentation [7], where noise, blur, edge enhancement, and subsampling distortion randomly differentiate the two patches. At this point, the lesion mask from the two synthetic time points are identical. Thus, there are no new lesions. In the second step, a connected component operation is used to separate each independent lesion from the lesion mask. Each lesion is either inpainted (*i.e.* removed) from one of the time points or both of them or neither time point. The lesion inpainting model is used to inpaint the lesion region with hallucinated healthy tissue (see 2.2). Next, the new lesion mask is constructed from lesion regions that have been kept in the second time point but not the first one. In the third step, the lesion generator model is used to simulate synthetic lesion at realistic locations. Synthetic lesions are generated for one of the time points or both of them (see 2.3). Similarly to the previous step, the new lesion mask is updated to include only the generated lesions on the second time point.

2.2 Lesion Inpainting Model

The lesion inpainting model is trained, independently and priorly to our proposed pipeline, with randomly selected 3D FLAIR patches which do not contain MS lesions or white matter hyperintensities. Similarly to [9], the 3D U-Net network is optimised to reconstruct altered input image. The input patch is corrupted with inpainting (random regions are replaced with noise) and other alterations. When the model is trained, it can be used to synthesize healthy regions in lesion locations that are replaced with random noise.

2.3 Lesion Generator Model

The lesion generator is trained before our proposed pipeline to simulate realistic lesions. The generator is a 3D U-Net network with two input channels and one output channel. The first input channel receives an augmented version of 3D FLAIR patch containing MS lesions where lesions are replaced with random noise. The second input channel receives the MS lesion mask of the original 3D FLAIR patch. The output channels predict the original 3D FLAIR patch with lesions. Thus, the trained model is able to simulate synthetic MS lesion from a 3D patch of FLAIR and its corresponding lesion mask.

2.4 New Lesions Segmentation

For the new lesion segmentation, a 3D U-Net architecture similar to the one proposed by [7] has been selected. As input, the network receives a concatenation of FLAIR patches from the two time points. The model output predicts the new lesion mask. The Dice loss is used during training to quantify similarity between the ground truth and the predicted mask.

The new lesion model is trained with mini-batches from both the generated data using the proposed pipeline and the challenge dataset [5]. Specifically in our implementation, the ratio of the real data from the challenge dataset over synthetic data in a mini-batch is 1/3. The total loss of the mini-batch is a weighted combination of the Dice loss of each element. Since challenge data are considered more reliable compared to synthetic data, a weight of 1/3 is attributed to synthetic data Dice loss whereas a weight of 1 is selected for challenge data.

2.5 Data

The dataset provided by the MICCAI 2021 - Longitudinal Multiple Sclerosis Lesion Segmentation Challenge [5] was used to train our method. For preprocessing, our strategy used the docker [6] built with the Anima scripts proposed by the challenge organizers. It includes bias correction, denoising and skull stripping.

3 Implementation Details

The networks are trained with FLAIR patches of size 64^3. Image quality data augmentation [7] is used when training the new lesion segmentation model. The models are optimized with Adam [8] using a learning rate of 0.0001 and a momentum of 0.9.

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. Coupé, P., Tourdias, T., Linck, P., Romero, J.E., Manjón, J.V.: Lesionbrain: an online tool for white matter lesion segmentation. In: International Workshop on Patch-based Techniques in Medical Imaging. pp. 95–103. Springer (2018)
5. the challenge dataset: https://portal.fli-iam.irisa.fr/msseg-2/data/
6. the pre-processing docker: https://github.com/Inria-Empenn/lesion-segmentation-challenge-miccai21/
7. 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)
8. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
9. Manjón, J.V., Romero, J.E., Vivo-Hernando, R., Rubio, G., Aparici, F., de la Iglesia-Vaya, M., Tourdias, T., Coupé, P.: Blind mri brain lesion inpainting using deep learning. In: International Workshop on Simulation and Synthesis in Medical Imaging. pp. 41–49. Springer (2020)
10. 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)
11. 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)

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.

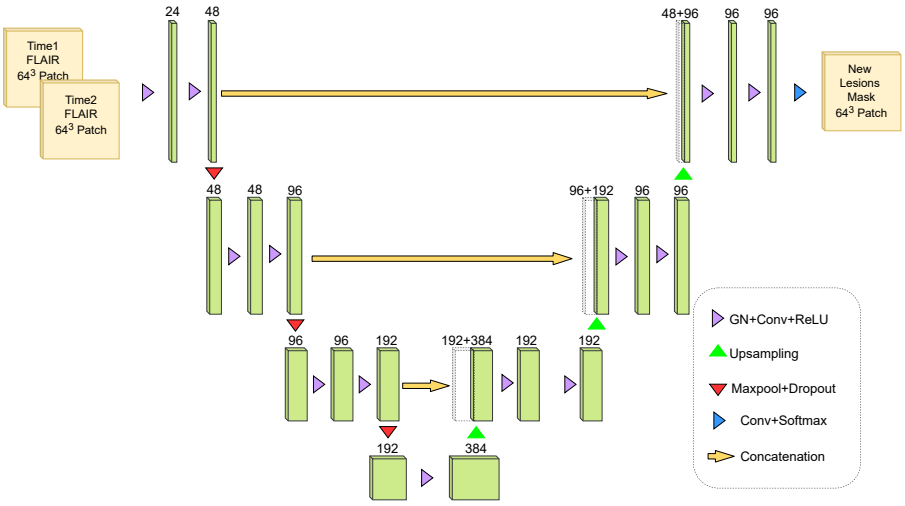

Fig. 1. Illustration of the used U-Net architecture. Each block is composed of group normalization (GN), Convolution (Conv) and Rectified Linear Unit (ReLU) activation.

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)

Double Pathway Method For MSSEG-2 Challenge

Domen Preloznik[1], Žiga Špiclin[1]

[1] Laboratory of Imaging Technologies, Faculty of Electrical Engineering, University of Ljubljana

Abstract. New lesion prevalence in patients brains during two time points shows a significant importance when evaluating patient's anti-inflammatory clinical therapy. Therefore, allowing clinicians to examine new lesion burden in an ongoing patient's therapy would be a major advancement in clinical practice. We carried out an automated detection of new brain lesions using convolutional neural network with double pathway architecture. We propose double pathway architecture due to its ability to evaluate input image through both – a highly detailed as well as voluminous processing. Double pathway architecture concurrently examines miniature but detailed and expanded but ambiguous input segments and intertwines the result through a fully connected convolutional neural network block. For new brain lesion detection application, we evaluated and determined numerous parameter and architecture assemblies and finished with the most prominent. Model training and testing was carried out on training dataset provided by MICCAI 2021 - Longitudinal Multiple Sclerosis Lesion Segmentation Challenge [1], [2]. Extensive evaluations demonstrated promising results and therefore a chance to participate for MSSEG-2 Challenge.

Keywords: convolutional neural network, double pathway, lesion detection.

1 Introduction

Brain disorders are conditions that have been taken into thorough examination due to high social and economic impact they have. Progressive exploration and fact-finding analysis regarding brain structure has been made possible with exploitation of new technology and advancements in medicine and machine learning.

Multiple sclerosis (MS) is a neurodegenerative disease, characterized by lesions, presented as focal areas of demyelination and inflammation, usually visible in brains' white matter. Imaging biomarkers, such as: brain lesion number, volume or load; represent a crucial criterion while assessing MS disease progression [3], [4]. MR imaging in conjunction with deep learning applications enables one of the most important aspects in diagnosing disease progression by performing brain lesion detection and segmentation. Providing clinical practitioners with automated lesion detection in patients brain ensures that patients diagnosis, and treatment, is backed up by measurable data and is not entirely reliant on doctor's knowledge and experience.

While lesion segmentation on its own is an important aspect of MS progression, an even more viable imaging biomarker is new lesion formation between two timepoints,

a baseline and a follow-up [5]. Improving patients' condition through a MS treatment would result in less – or none, new lesion making. Contrary, improper MS therapy could yield more lesion creation, resulting in serious disease progression. Therefore, detailed, and automated baseline-to-follow-up analysis could be a major advancement in clinical practice regarding MS disease treatment.

2 Method and materials

2.1 Model overview

Proposed convolutional neural network (CNN) consists of parallel pathways combining both high and low resolution for input image segmentation. Two parallel pathways - hence the name double pathway CNN [6].

Double pathway (DP) CNN architecture enables concurrent examination of input image through two branches and joining the results with the final fully connected block. An input to DP is provided by extracting image patches, much smaller than original image size. Extracted patch is then processed to fit both high- and low- resolution pathways. For convenience name, mentioned extracted patch will be referenced as a large patch.

First branch of DP architecture is considered a high-resolution one. Input to the pathway is an image patch, extracted by center cropping the large patch. By only utilizing a crop augmentation, image resolution stays the same as input image, but patch size is reduced by the crop dimension. The patch processing results in a much smaller patch of high resolution.

Second branch of DP architecture is considered a low-resolution one. Input to this pathway is an image patch, extracted by down sampling the large patch. By utilizing down sampling, image resolution is compromised and is therefore lesser than original image. In contrast to first DP pathway, this patch extraction provides a much greater field of view, with downside of bad resolution, defined by down sample parameter.

Both first and second branch are joint together at the end, where a fully connected block is used. Both branches provide unique aspects of input data and in conjunction with classification layer, they are performing the segmentation task. See **Fig. 1**

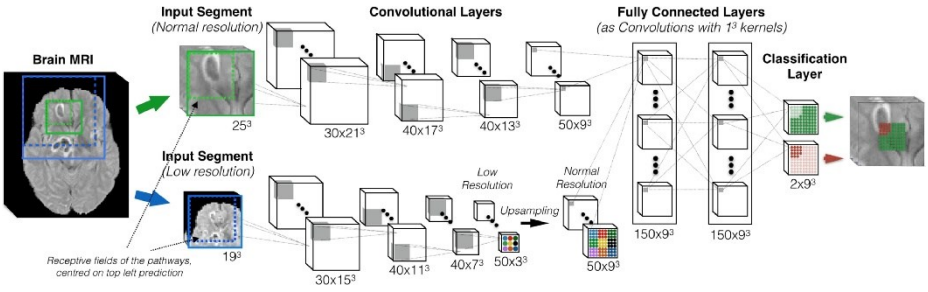

Fig. 1. Double pathway CNN architecture [6].

CNN architecture is predefined by total number of convolutional layers and their retrospective feature map count. Basic setup is upgraded through trial-and-error training and validation process.

2.2 Training overview

To achieve reliable and valid results, input data must have a proper preprocessing pipeline. For MSSEG-2 challenge task this conducted to image reorientation, atlas registration, image normalization and crop augmentation, described by **Table 1**.

Table 1. Data preprocessing steps.

1. Image re-orientation	To RAI orientation.
2. Atlas registration	To common space.
3. Image normalization	Rescale intensities to [0, 1].
4. Crop augmentation	Crop image to fit CNN architecture

Double pathway model requires proper patch extraction for input images and labels. Receptive field of CNN application depends on patch centers and its' size. It is an important aspect that determines final model performance. A task of new lesion detection requires a very detailed look on input image to detect small-scale changes, as well as robustness to not mark any pre-existing lesions. To achieve the most prominent result for specific task, patch centers were sampled subsequently with what we propose is a "reinforced centers" sampling.

Reinforced centers provide the CNN model with patches that meet certain condition. All other patches are discarded during training. With this approach, the consequence of small number of new lesions is leveled out and enables efficient and upgraded model training.

Extensive training protocols resolved in patch extraction and training parameters summed up in a **Table 2**.

Table 2. Double pathway CNN model parameters.

Parameter description	**Value**
Patch size – large	[97, 97, 97] (x, y, z)
High resolution - crop dimension	[33, 33, 33] (x, y, z)
Low resolution - down sample step	[3, 3, 3] (x, y, z)
Number of patches – training	288 + reinforced centers
Number of patches – inference	2304
High resolution – nr. convolutional layers	7
Low resolution – nr. convolutional layers	8
Fully connected – nr. convolutional layers	2
High resolution – feature maps	[30, 40, 50, 60, 70, 80, 90]
Low resolution – feature maps	[30, 40, 50, 60, 70, 80, 90, 100]
Fully connected – feature maps	[300, 300]

References

[1] C. Confavreux, D. A. Compston, O. R. Hommes, W. I. McDonald, and A. J. Thompson, 'EDMUS, a European database for multiple sclerosis', *J Neurol Neurosurg Psychiatry*, vol. 55, no. 8, pp. 671–676, Aug. 1992, doi: 10.1136/jnnp.55.8.671.

[2] S. Vukusic *et al.*, 'Observatoire Français de la Sclérose en Plaques (OFSEP): A unique multimodal nationwide MS registry in France', *Mult Scler*, vol. 26, no. 1, pp. 118–122, Jan. 2020, doi: 10.1177/1352458518815602.

[3] 'PubMed entry'. Accessed: Jul. 13, 2021. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/29320652

[4] T. Uher *et al.*, 'Combining clinical and magnetic resonance imaging markers enhances prediction of 12-year disability in multiple sclerosis', *Mult Scler*, vol. 23, no. 1, pp. 51–61, Jan. 2017, doi: 10.1177/1352458516642314.

[5] Ž. Lesjak, F. Pernuš, B. Likar, and Ž. Špiclin, 'Validation of White-Matter Lesion Change Detection Methods on a Novel Publicly Available MRI Image Database', *Neuroinform*, vol. 14, no. 4, pp. 403–420, Oct. 2016, doi: 10.1007/s12021-016-9301-1.

[6] K. Kamnitsas *et al.*, 'Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation', *Medical Image Analysis*, vol. 36, pp. 61–78, Feb. 2017, doi: 10.1016/j.media.2016.10.004.

Longitudinal Multiple Sclerosis Lesion Segmentation Using Pre-activation U-Net

Pooya Ashtari[1,2], Berardino Barile[1,2], Sabine Van Huffel[2], and Dominique Sappey-Marinier

[1] Department of Electrical Engineering (ESAT), STADIUS Centre for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Leuven, Belgium
{pooya.ashtari, berardino.barile, sabine.vanhuffel}@esat.kuleuven.be
[2] CREATIS (UMR 5220 CNRS & U1206 INSERM), Université Claude Bernard Lyon 1, Université de Lyon, Villeurbanne, France
{pooya.ashtari, berardino.barile, dominique.sappey-marinier}@creatis.insa-lyon.fr
[3] CERMEP - Imagerie du Vivant, Université de Lyon, Bron, France

Abstract. Automated segmentation of new multiple sclerosis (MS) lesions in 3D MRI data is an essential prerequisite for monitoring and quantifying MS progression. Manual delineation of such lesions is laborious, time-consuming, and expensive especially because raters need to deal with 3D images and several modalities. In this paper, we propose a 3D encoder-decoder architecture with residual blocks for the longitudinal segmentation of MS lesions. Due to a limited training set and the class imbalance problem, we apply intensive data augmentation and use deep supervision to train our model.

Keywords: Multiple sclerosis · Lesion segmentation · U-Net · Pre-activation block.

1 Introduction

Multiple sclerosis (MS) is a common chronic, autoimmune demyelinating disease of the central nervous system (CNS), which causes inflammatory lesions in the brain, particularly in white matter (WM). Multi-parametric MRI is widely used to diagnose and assess MS lesions in clinical practice. Particularly, FLuid Attenuated Inversion Recovery (FLAIR) images, distinguish white matter lesions appearing as high-intensity regions. It is highly relevant to monitor lesion activities, especially the appearance of new lesions and the enlargement of existing lesions, for various purposes, including prognosis and follow-up. Lesional changes between two longitudinal MRI scans from an MS patient are the most important markers for tracking disease progression and inflammatory changes. In particular, the accurate segmentation of new lesions is an essential prerequisite to quantify them and measure features, such as lesion volume and location. However, manual identification of such lesions is tedious, time-consuming, and expensive especially because experts need to deal with 3D images and several modalities; therefore, accurate computer-assisted methods are needed to automatically perform this task.

Longitudinal MS lesion segmentation is, however, very challenging since MS images often change subtly over time within a patient, and new lesions can be very small although they vary dramatically in shape, structure, and location across patients. MSSEG-2 [1,2] aims to develop

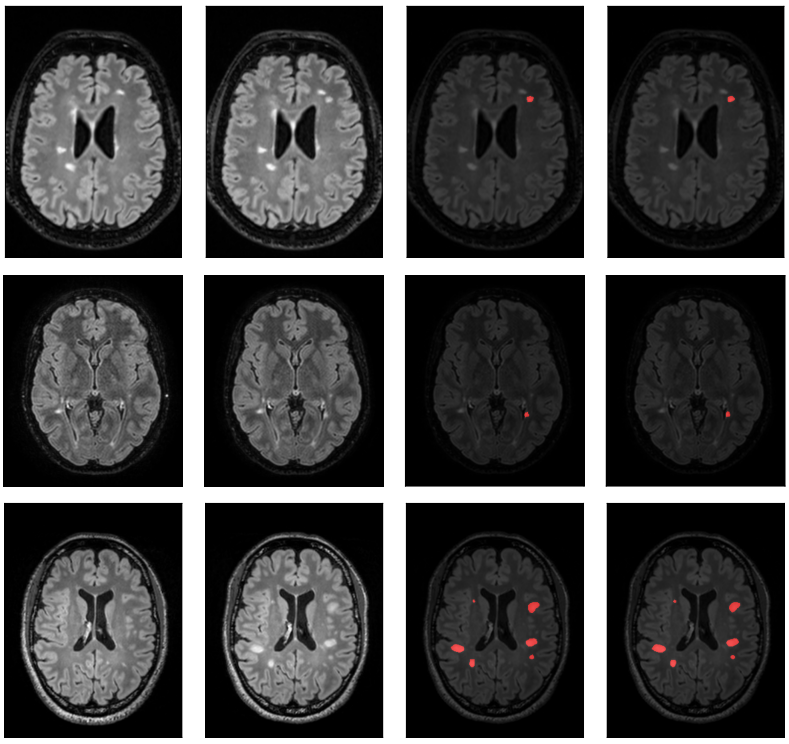

Fig. 1 FLAIR images along with the corresponding ground truth and predictions for typical MS cases.

effective data-driven algorithms for the segmentation of new MS lesions by providing a dataset of 40 3D FLAIR images acquired at two time points (with varying intervals) and registered in the intermediate space between the two time points. For each case, new lesions are manually annotated by multiple raters, and the consensus ground truths are obtained through a voxel-wise majority voting (see Figure 1).

Over the past decade, convolution neural networks (CNNs) with an encoder-decoder architecture, known as U-Net [3], have dominated medical image segmentation. In contrast to a hand-crafted approach, U-Net can automatically learn high-level task-specific features for MS lesion segmentation. This work proposes a 3D U-Net architecture with pre-activation ResNet blocks [4,5] for the segmentation of new lesions. We use deep supervision [6] and perform intensive data augmentation to effectively train our models.

The rest of this paper is organized as follows: Section 2 briefly reviews relevant semantic segmentation techniques. Section 3 presents our approach to longitudinal MS lesion segmentation. Experiments are presented in Section 4.2. We conclude this paper in Section 5.

2 Related work

Over the past few years, considerable efforts have been made in the development of fully convolutional neural networks for semantic segmentation. Encoder-decoder architectures, in particular U-Net [3] and its variants, are dominant in the segmentation of brain lesions. nnU-Net [7] makes minor modifications to the standard 3D U-Net [8], automatically configuring the key design choices. It has been successfully applied to many medical image segmentation tasks, including longitudinal MS lesion segmentation [9]. Aslani et al. [10] propose a deep architecture made up of multiple branches of convolutional encoder-decoder networks that perform slice-based MS lesion segmentation. La Rosa et al. [11] propose a U-Net-like model, to automatically segment cortical and white matter lesions based on 3D FLAIR and MP2RAGE images. Gessert et al. [12] address longitudinal segmentation of new and enlarged lesions using a two-path CNN that jointly processes two FLAIR images from two time points.

3 Method

In this section, we describe the proposed encoder-decoder architecture, called Pre-U-Net, and its building blocks.

3.1 Overall Architecture

The overall architecture, as shown in Figure 2, follows a U-Net-like style made up of encoder and decoder parts. A $3 \times 3 \times 3$ convolution is used as the stem layer. The network takes a 2-channel image of size $128 \times 128 \times 128$ and outputs a *probability map* with the same spatial size. The network has 4 levels, at each of which in the encoder (decoder), the input tensor is downsampled (upsampled) by a factor of two while the number of channels is doubled (halved). Downsampling and upsampling are performed via strided convolution and transposed convolution, respectively. The kernel size of all downsamplers and upsamplers is $3 \times 3 \times 3$. We use deep supervision at the three highest resolutions in the decoder, applying pointwise convolutions to get three auxiliary logit tensors.

3.2 Pre-activation ResNet Block

Pre-activation ResNet block [5] is the cornerstone of our U-Net model. This block is composed of two $3 \times 3 \times 3$ convolutions, each of which follows LeakyReLU activation and Group Normalization [13] (with a group size of 8). A $1 \times 1 \times 1$ convolution may be used in the shortcut connection to match the input dimension with the output dimension of the residual mapping. For future reference, we call the resulting model Pre-U-Net.

4 Experiments

All the models are implemented using PyTorch [14] and PyTorch Lighting [15] frameworks and trained on NVIDIA P100 GPUs. We evaluate the performance of Pre-U-Net for MS lesion

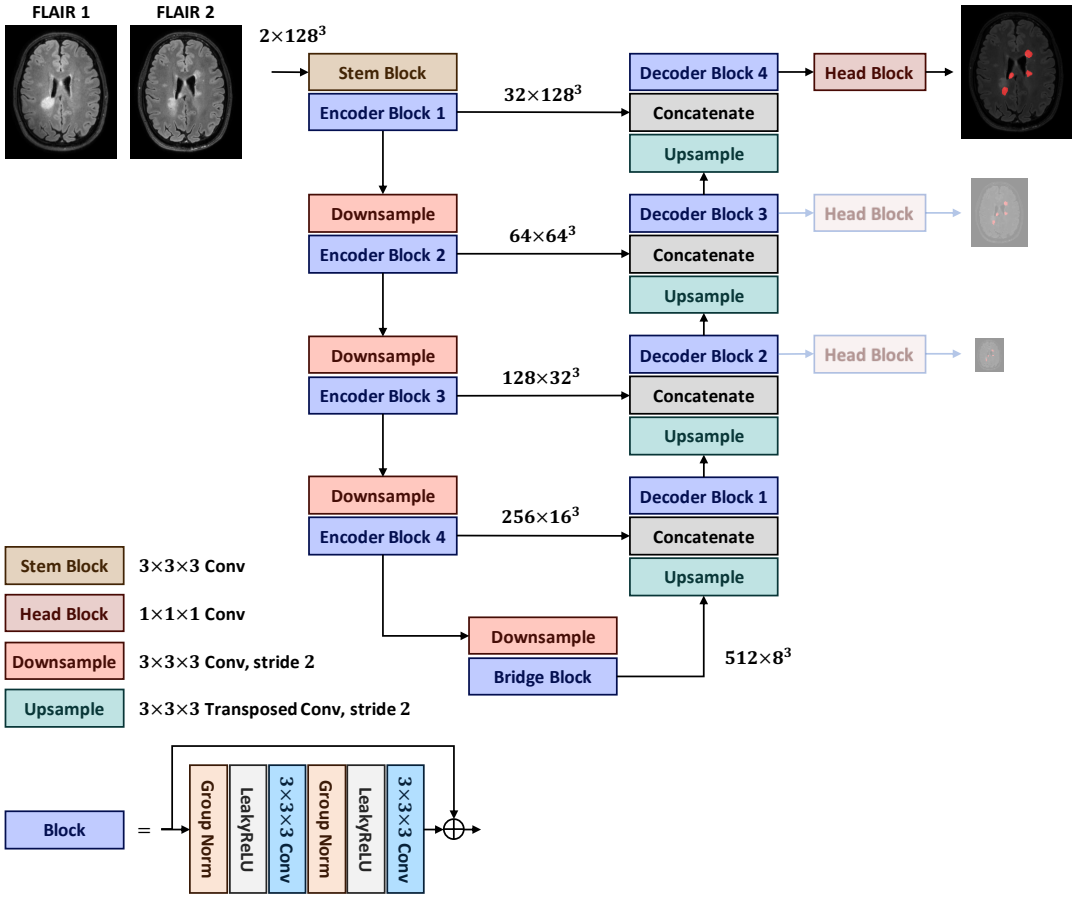

Fig. 2 The proposed encoder-decoder architecture. The two lower-resolution auxiliary maps are only used in the training phase as deep supervisions.

segmentation on the MSSEG-2 dataset. We follow the same training workflow in all the experiments. In the following, we first provide the details of this workflow, then present the evaluation protocol and the results.

4.1 Setup.

Preprocessing. For each patient, we first concatenate the two FLAIR images to form a 2-channel 3D image as the input. Each image and its ground truth are then cropped with a minimal box filtering out zero regions. MSSEG-2 data are heterogeneous in the sense the images may be acquired with different protocols using different scanners, making intensity values greatly vary across patients and even across time points within the same patient. Therefore, we normalize each image channel-wise using z-score to have intensities with zero mean and unit variance. Moreover, all the images and their ground truths are then resampled to the same voxel spacing

0.6 mm^3 using trilinear interpolation. Finally, we crop random $128 \times 128 \times 128$ patches and perform oversampling from lesion regions such that 50% of the patches contain some lesion to cope with the class imbalance problem.

Data Augmentation. To reduce overfitting caused by data insufficiency and heterogeneity, it is crucial to perform an effective data augmentation workflow before feeding the data into the network. More specifically, we apply spatial transforms, including random affine and random flip along all axes, and intensity transforms, including random Gaussian noise, random Gaussian smoothing, random intensity scaling and shifting, random bias field, and random contrast adjusting. All the transforms are computed on the fly on CPU using MONAI library [16].

Optimization. All networks are trained for 100k steps with batch size 2 (each batch is processed on one GPU) using AdamW optimizer with the initial learning rate of 1e−5, weight decay of 1e−2, and a cosine annealing scheduler. The loss function is the sum of soft Dice [17] and Focal loss [18]. The three deep supervision outputs and the corresponding downsampled ground truths are used for loss computation. The training objective is the weighted mean of the losses at all resolutions, with the weights being 1, 0.5, and 0.25 at resolutions 128^3, 64^3, and 32^3, respectively.

Inference. A test image in the inference is first subjected to z-score intensity normalization and resampled to a voxel spacing of 0.6 mm^3. The prediction is then made using a sliding window approach with a 50% overlap and a window size of $128 \times 128 \times 128$ (same as the patch size used in training). We resample the resulting logit tensor back to the original voxel spacing and finally threshold it to obtain a binary segmentation map.

4.2 Results

The proposed model, Pre-U-Net, has 26.3 million trainable parameters. We performed 5-fold cross-validation to estimate how capable Pre-U-Net is in generalizing to unseen data. The metric values are reported in Table 1. In particular, we could achieve a Dice score of 0.62 and a sensitivity of 0.58. The smaller value of sensitivity and higher value of specificity indicate a tendency to undersegmentation, which is attributed to the problem of class imbalance and insufficient data. Figure 1 shows the results of our model on representative MS cases.

5 Conclusion

We devised a U-Net-like architecture consisting of pre-activation blocks, called Pre-U-Net, for longitudinal MS lesion segmentation. We successfully trained our models by using data augmentation and deep supervision, alleviating the problem of data insufficiency and class imbalance. We evaluated the effectiveness of Pre-U-Net on the MSSEG-2 dataset, achieving a Dice score of ≈ 0.62 and sensitivity of ≈ 0.58 in segmenting new white matter lesions in 3D FLAIR images from an MS patient.

Table 1 Summary statistics of 5-Fold cross-validation scores for Pre-U-Net on the MSSEG-2 dataset.

	Dice (%)	Sensitivity (%)	Specificity (%)
Mean	62.46	57.67	99.99
SD	36.11	37.40	0.00
Median	72.38	63.78	99.99
25th quantile	39.05	24.92	99.99
75th quantile	100.00	100.00	100.00

Acknowledgements. The research leading to these results has received funding from EU H2020 MSCA-ITN-2018: INtegrating Magnetic Resonance SPectroscopy and Multimodal Imaging for Research and Education in MEDicine (INSPiRE-MED), funded by the European Commission under Grant Agreement #813120. This research also received funding from the Flemish Government (AI Research Program). Sabine Van Huffel and Pooya Ashtari are affiliated to Leuven.AI - KU Leuven institute for AI, B-3000, Leuven, Belgium.

References

1. Sandra Vukusic, Romain Casey, Fabien Rollot, Bruno Brochet, Jean Pelletier, David-Axel Laplaud, Jérôme De Sèze, François Cotton, Thibault Moreau, Bruno Stankoff, et al. Observatoire français de la sclérose en plaques (ofsep): A unique multimodal nationwide ms registry in france. *Multiple Sclerosis Journal*, 26(1):118–122, 2020.
2. C Confavreux, DA Compston, OR Hommes, WI McDonald, and AJ Thompson. Edmus, a european database for multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 55(8):671–676, 1992.
3. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
4. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
5. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016.
6. Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, and Zhuowen Tu. Deeply-supervised nets. In *Artificial intelligence and statistics*, pages 562–570. PMLR, 2015.
7. Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H Maier-Hein. No new-net. In *International MICCAI Brainlesion Workshop*, pages 234–244. Springer, 2018.
8. Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.
9. Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, pages 1–9, 2020.

10. Shahab Aslani, Michael Dayan, Loredana Storelli, Massimo Filippi, Vittorio Murino, Maria A Rocca, and Diego Sona. Multi-branch convolutional neural network for multiple sclerosis lesion segmentation. *NeuroImage*, 196:1–15, 2019.
11. Francesco La Rosa, Ahmed Abdulkadir, Mário João Fartaria, Reza Rahmanzadeh, Po-Jui Lu, Riccardo Galbusera, Muhamed Barakovic, Jean-Philippe Thiran, Cristina Granziera, and Merixtell Bach Cuadra. Multiple sclerosis cortical and wm lesion segmentation at 3t mri: a deep learning method based on flair and mp2rage. *NeuroImage: Clinical*, 27:102335, 2020.
12. Nils Gessert, Julia Krüger, Roland Opfer, Ann-Christin Ostwaldt, Praveena Manogaran, Hagen H Kitzler, Sven Schippling, and Alexander Schlaefer. Multiple sclerosis lesion activity segmentation with attention-guided two-path cnns. *Computerized Medical Imaging and Graphics*, 84:101772, 2020.
13. Yuxin Wu and Kaiming He. Group normalization. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.
14. Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*, pages 8026–8037, 2019.
15. WA Falcon. Pytorch lightning. *GitHub. Note: https://github.com/PyTorchLightning/pytorch-lightning Cited by*, 3, 2019.
16. Nic Ma, Wenqi Li, Richard Brown, Yiheng Wang, Behrooz, Benjamin Gorman, Hans Johnson, Isaac Yang, Eric Kerfoot, charliebudd, Yiwen Li, Mohammad Adil, Yuan-Ting Hsieh, Arpit Aggarwal, Cameron Trentz, adam aji, masadcv, Mark Graham, Ben Murray, Gagan Daroach, Petru-Daniel Tudosiu, myron, Matt McCormick, Ambros, Balamurali, Christian Baker, Jan Sellner, Lucas Fidon, and cgrain. Project-monai/monai: 0.5.3, June 2021.
17. Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016.
18. Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

A UNet Pipeline for Segmentation of New MS Lesions

Cory Efird, Dylan Miller, and Dana Cobzas

MacEwan University, Edmonton AB, Canada

Abstract. A pipeline for the second multiple sclerosis segmentation challenge (MSSEG-2) hosted by MICCAI is proposed. Two FLAIR images taken at different time-points are used as a multi-channel input to a 3D CNN to detect new lesions. Patch sampling strategies are adopted to keep the input volume shape manageable in terms of memory requirements. To further improve results, multiple models and patch orientations are ensembled. Performance is evaluated against nn-UNet.

Keywords: Multiple Sclerosis · Segmentation · Deep Learning

1 Method

The proposed method uses a 3D convolutional neural network (CNN) to detect new multiple sclerosis (MS) lesions in FLAIR images taken at two different time-points. We have chosen a relatively simple approach where both time-points are used as a multi-channel input to the CNN. This is viable due to the accurate co-registration that was performed between the two time-points. As a result, the network is able to produce spatio-temporal features early in the first few layers. The network is trained using patches that are randomly sampled from brain volumes. At inference time, predictions are generated for evenly spaced overlapping patches that cover the entire volume. Afterward, the patches are combined by interpolating the overlapping regions.

A survey of MS lesion segmentation with convolutional neural networks (CNNs) was recently published [1], which guided the development of our method. Furthermore, the success of the nn-UNet pipeline [3] influenced the pipeline design, and served as a baseline method for comparisons.

1.1 Data Processing

Pre-Processing First, the standard processing package that is provided by the contest organizers is applied. This includes brain extraction, bias field correction, and masking of non-brain image regions. The processing is extended with the following steps: All images are re-sampled to a target spacing that is close to $1\times1\times1\text{mm}^3$, with a tolerance of 0.11mm for each axis. The tolerance allows for integer multiples and divisions of the original spacing to be preferred (e.g. a spacing of 0.45mm is resampled to 0.9mm). The volumes are cropped to the

bounding box of the brain mask. At this point, augmentations are applied to training data. Finally, the intensity of the FLAIR images is adjusted by clipping values below the 0.05^{th} and above the 99.5^{th} percentiles, and then transforming the remaining values onto the range $[-1, 1]$.

Augmentation A number of spatial and intensity augmentations implemented in the TorchIO package [2] are applied to the training data before intensity normalization. First, a random volume orientation is chosen. The volume axes may be permuted or flipped so that 48 orientations are possible. If a spatial augmentation is applied (p=0.75), it is either an elastic deformation (p=0.2), or an affine transformation (p=0.8). Intensity augmentations include procedurally generated bias fields (p=0.5), modifying gamma by raising image values to a random power (p=0.8), a random blur kernel (p=0.2), and random high-frequency noise (p=0.35).

1.2 Implementation

Patch Extraction During training, patches with a size of 96×96×96 are sampled from random locations in the brain volume. A weighted volume is used to set the relative probability that a voxel will be chosen as the center of a patch. Voxels that are background, brain, and lesion have probabilities of 0, 1, and 100 respectively.

Model Our network architecture is a standard 3D UNet with 6 layers and residual blocks (Figure 1). The number of feature maps are linearly increased by 40 every 2 downsampling operations, as opposed to the common doubling approach. This saved many parameters without harming segmentation performance. Adding an anti-aliasing step when downsampling was shown to improve shift-invariance for classification tasks [5]. Anti-aliased convolutions with a stride of 2 are used for upsampling and downsampling to test this operation in a segmentation setting.

Loss Function A recent review of segmentation loss functions reported that a combination of dice and distribution-based loss functions is most reliable [4]. We adopt a commonly used hybrid of logistic and dice losses, which is formulated as follows:

$$\mathcal{L} = \frac{1}{C}\sum_{c=1}^{C}\left[\underbrace{-\frac{1}{N}\sum_{n=1}^{N} g_{n,c}\log p_{n,c}}_{\text{logistic term}} + \underbrace{1 - \frac{2\sum_{n=1}^{N} p_{n,c}g_{n,c}}{\sum_{n=1}^{N} p_{n,c}^2 + \sum_{n=1}^{N} g_{n,c}^2}}_{\text{dice term}}\right]$$

Training Five models are trained for cross validation. For each model 4 folds make up 32 participants in the training set, and the remaining fold has 8 participants for validation. The SGD optimizer is used with a batch size of 4 and a

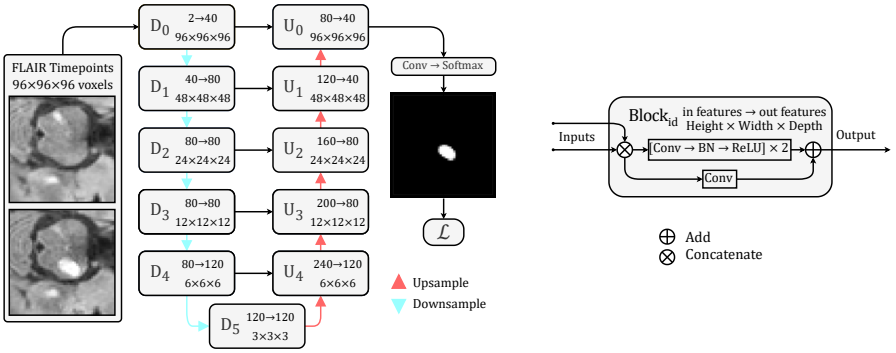

Fig. 1. Proposed model architecture. The convolutional block has a typical [Conv → BN → ReLU] × 2 local topology. A residual path has a single convolution to match the number of output channels. Feature maps are channel-wise concatenated when blocks have multiple input connections.

learning rate of 0.001. Training was halted if there was no improvement in mean dice score across the validation subjects for 2000 iterations, which resulted in an average of 7000 iterations per model. To train all models it took 330 hours on Tesla V100 GPUs with 32GB of RAM. The code was written using the PyTorch framework.

Inference To generate final predictions, a sliding window extracts patches with a 50% overlap across the entire volume. Due to the overlap, there are 8 patches which contribute to a voxel. Additionally, 8 patch flips are passed into the model and ensembled, and then all 5 models from the cross validation are ensembled at test time. In total there are $5 \cdot 8 \cdot 8 = 320$ forward passes through the model that contribute to every voxel.

2 Results

The commonly used Dice Similarity Coefficient (DSC) is reported, however, it has shortcomings when evaluating lesion segmentation. For subjects with no lesions, DSC is undefined. To handle this we let DSC=1 if the model correctly predicts no lesions, and DSC=0 if the model predicts even a single lesion voxel. A more robust test [6] was re-implemented to obtain precision, recall, and F1 measures for lesion detection. The performance on the test set has not been released at the time of writing this paper. Instead, single-model cross validation results are reported in Table 1, and 3D surface renderings of lesion segmentations are displayed in figure 2. In summary, our pipeline has a high detection recall, but lower precision when compared to nn-UNet.

Pipeline	Detection F1	Detection Precision	Detection Recall	Dice
nn-UNet	**0.77**	**0.70**	0.55	0.50
Ours	0.71	0.58	**0.88**	**0.54**

Table 1. Mean detection F1, precision, recall, and dice scores for all 40 participants. Lesion segmentations were produced by the model where the participant was in the validation set.

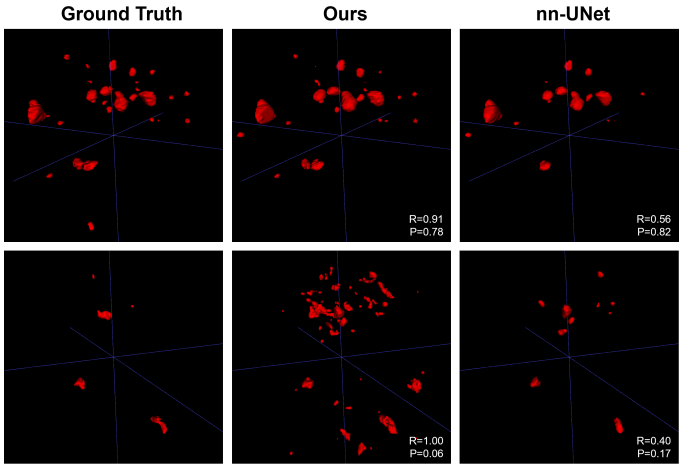

Fig. 2. 3D surface rendering of segmentation results for subjects 095, 057, 094 and 026 (from top to bottom row). Detection precision (P), and detection recall (R) is displayed for predictions from nn-UNet and our pipeline.

References

1. Zhang H., Oguz I. Multiple Sclerosis Lesion Segmentation - A Survey of Supervised CNN-Based Methods. In: Crimi A., Bakas S. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2020. Lecture Notes in Computer Science, vol 12658, 2021.
2. F. Pérez-García, R. Sparks, and S. Ourselin. TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. Computer Methods and Programs in Biomedicine (June 2021), p. 106236. ISSN: 0169-2607
3. Isensee, F., Jaeger, P.F., Kohl, S.A.A. et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18, 203–211 (2021). https://doi.org/10.1038/s41592-020-01008-z
4. J. Ma. Loss odyssey in medical image segmentation. Medical Image Analysis, vol. 71 2021.
5. Zhang, R. Making Convolutional Networks Shift-Invariant Again. ICML, 2019.
6. Commowick, O., Istace, A., Kain, M. et al. Objective Evaluation of Multiple Sclerosis Lesion Segmentation using a Data Management and Processing Infrastructure. Sci Rep 8, 13650 (2018). https://doi.org/10.1038/s41598-018-31911-7

Triplanar U-Net with Orientation Aggregation for New Lesions Segmentation

Tiziano Dalbis*, Thomas Fritz*, Joana Grilo*,
Sebastian Hitziger*, and Wen Xin Ling*

Mediaire GmbH, Möckernstraße 63, 10965 Berlin, Germany, https://mediaire.de/

Abstract. We present a method for new lesion segmentation in T2/FLAIR sequences which is based on the 2D U-Net. Our approach processes the 3D volumes slice-wise, across the coronal, axial, and sagittal planes. The final new lesion segmentation mask is then obtained by aggregating the 3D segmentation masks from all three orientations.

1 Introduction

The objective of the MSSEG-2 challenge is the detection and segmentation of new MS lesions by comparing one time point of a patient's T2/FLAIR sequence to another. The provided data consists of 100 pairs of 3D FLAIR volumes, each corresponding to two scans acquired at different time points (1-3 years apart) of the same patient. 40 of these datasets image pairs are provided as a training set, together with the new lesion ground truth mask. The remaining 60 volumes are kept for testing the submitted models.

The performance of the submitted models is evaluated in terms of (a) their ability to detect new lesions measured by an F1 score and (b) the segmentation accuracy of the new lesions, measured by the Dice score.

We here present our approach to the challenge, which uses as a base model a U-Net [1] architecture. It is trained on 2D slices from the coronal, axial, and sagittal planes and aggregates the predictions of each orientation to obtain the final segmentation.

We submit two segmentation pipelines, A and B, that use the same model architecture but make use of different data: while the model in pipeline A is trained only on the official training data, we use additional datasets for training the model in pipeline B, as described below. Besides this difference, the two pipelines are identical.

2 Data and Preprocessing

For training the model in **pipeline A**, we use only the official training data from the MSSEG-2 challenge, consisting of 3D FLAIR sequences from 40 patients acquired at two different points in time. The two time point volumes of each

* equal contribution

patient have already been transformed onto a common middle point through rigid registration. See `https://portal.fli-iam.irisa.fr/msseg-2/data/` for more information on the dataset.

For the model in **pipeline B**, we add our own datasets to the official training data, consisting of pairs of 3D FLAIR images from 46 patients.

We apply the following preprocessing steps: (1) affine registration of each pair of FLAIR images to the MNI template using the ANTs processing toolbox, (2) cropping the FOV close to the brain, (3) resampling the volume to 256 x 256 x 256 voxels, and (4) pixel normalization through mean subtraction and division by the standard deviation.

To increase the generalization ability of the model, data augmentation is performed on the preprocessed 3D volumes, including contrast augmentation, rotations, flipping across the three orthogonal planes, elastic deformations, and bias field augmentation.

3 Model

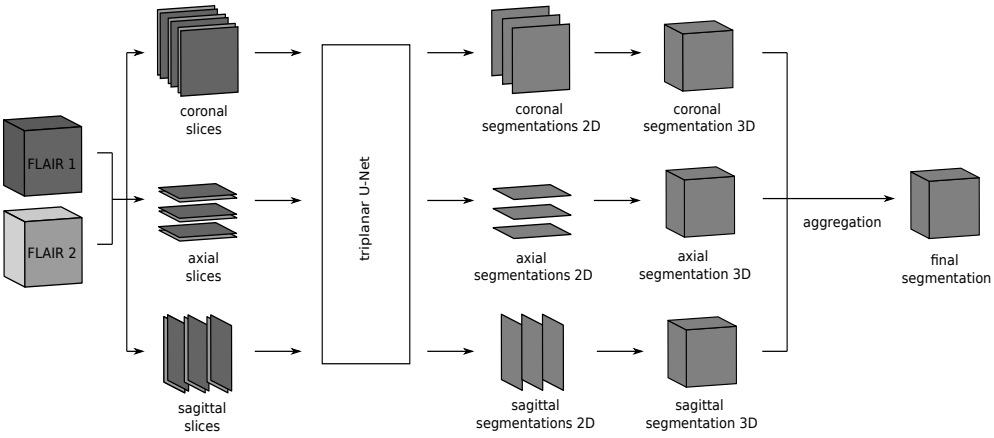

Fig. 1. Segmentation process using our triplanar U-Net, previously trained on slices from all three orthogonal planes. The 3D FLAIR input volumes are sliced along the coronal, axial, and sagittal planes and grouped together in pairs of corresponding slices. For each of the three orientations, the segmentation is now performed independently: (i) each slice pair is given as a two-channel 2D input to the triplanar U-Net and (ii) the resulting single-channel 2D segmentation masks are reassembled to 3D volumes. In the final step, the three 3D segmentations are aggregated to yield the final segmentation.

The basis of our fully convolutional neural network, which we refer to as triplanar U-Net, is a 2D U-Net [1]. It has two input channels with corresponding

slices, either axial, coronal or sagittal, from the two different time points of each patient. The output of the model is a single-channel 2D binary mask, representing the segmentation of the new lesions found in the corresponding slice.

We note that similar triplanar U-Net architectures have been suggested before, for instance, in [2] and [3]. However, a major difference of our approach is the training of a single U-Net on slices from all orientations (opposed to training three orientation-specific U-Nets).

3.1 Training

We train the triplanar U-Net on batches, each combining a total of 20 axial, coronal, and sagittal slices from different patient volumes for robustness. For the updates of the model weights, we use stochastic gradient descent with momentum and an initial learning rate of 0.0001, which is reduced when the validation loss plateaus. Training is performed until the validation loss stops decreasing, which was the case after around 100 epochs.

Loss function In addition to the Dice loss, we use a TopK cross-entropy loss [4] to focus on the hard cases. The combined loss function is the sum of the Dice loss and the TopK loss.

Cross validation We train the triplanar U-Net on five different folds, for each of which we hold out 20% of the training data for validation.

3.2 Inference

The segmentation process at inference is depicted in Figure 1. From the two 3D FLAIR images of each patient, three datasets are created, consisting of pairs of axial, coronal, and sagittal slices, respectively. For each such dataset, segmentation is performed slice-wise with the triplanar U-Net, resulting in a set of single-channel 2D masks, which are then reassembled to a 3D binary mask. The combined segmentation is obtained by aggregating the three individual masks through majority voting.

Ensemble average At test time, we perform inference with the models trained on the five folds. The predictions are averaged across the folds to yield the final prediction.

3.3 Implementation

The model is implemented and trained in Python using the PyTorch package.

References

1. Ronneberger, O., Fischer, P., and Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
2. Roy, A. G., Conjeti, S., Navab, N., and Wachinger, C. Alzheimer's Disease Neuroimaging Initiative. (2019). QuickNAT: A fully convolutional network for quick and accurate segmentation of neuroanatomy. NeuroImage, 186, 713-727.
3. Sundaresan, V., Zamboni, G., Rothwell, P. M., Jenkinson, M., and Griffanti, L. (2021). Triplanar ensemble U-Net model for white matter hyperintensities segmentation on MR images. Medical Image Analysis, 73, 102184.
4. Wu, Z., Shen, C., and Hengel, A. V. D. (2016). Bridging category-level and instance-level semantic image segmentation. arXiv preprint arXiv:1605.06885.

Segmentation of New MS Lesions with Tiramisu and 2.5D Stacked Slices

Huahong Zhang[1], Hao Li[1], and Ipek Oguz[1(✉)]

Vanderbilt University, Nashville, TN 37235, USA

Abstract. White matter lesion segmentation from magnetic resonance images (MRIs) is an important task in multiple sclerosis (MS) studies. To better understand the disease progression, quantitative analysis of lesions between different time-points is essential. In this paper, we propose a novel method for segmentation of new MS lesions. We use a fully convolutional neural network (CNN) to segment MS lesions from different time-points. We stack slices from 3 different orthogonal views to achieve a 2.5D method which provides both local and global information for segmentation. We also apply an augmentation strategy for domain generalization to increase the learning ability of model from the source domain to the target domain.

Keywords: Multiple sclerosis · Segmentation · Domain Generalization · Deep learning.

1 Introduction

Magnetic resonance images (MRIs) are widely used to quantitatively analyze white matter lesions in Multiple Sclerosis (MS) patients. However, the segmentation of MS lesions is a challenging task, both for human experts and for automated algorithms. This is mainly due to the large variability of lesion appearance, shape and location among different subjects as well as scanners and acquisition protocols. Accurate segmentation methods with good generalizability are thus desirable, and deep learning methods have dominated this field in recent years [2].

Quantifying new MS lesions is of particular interest for identifying recent disease activity. Based on our previous work [3], we propose to use 2.5D stacked slices and Tiramisu network for new lesion segmentation.

2 Materials and Method

2.1 Dataset and pre-processing

The MSSEG-2 challenge dataset[1] contains 100 MS patients, and each patient has FLAIR images at 2 different time-points. The gap between the two time-points is 1-3 years. The images are scanned by 15 different scanners (1.5T or 3T)

[1] https://portal.fli-iam.irisa.fr/msseg-2/

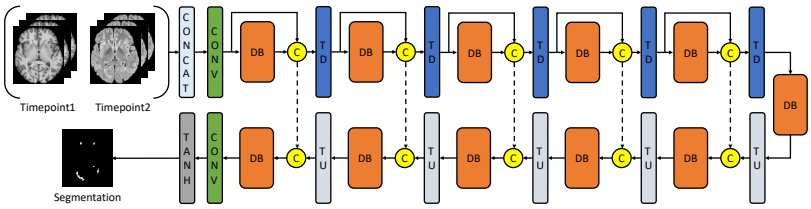

Fig. 1. Network architecture of the proposed framework.

and split into training/testing set as 40/60. The pre-processing is performed by Anima[2] as suggested by the challenge organizers. This includes brain extraction at each time point and masking the FLAIR images with the union of the masks from both time points, as well as bias correction.

2.2 2.5D stacked slices

Our 2.5D stacked slices [3] can be split into 2 parts: (1) stack of slices and (2) 2.5D. A stack of slices is defined as a center slice and its adjacent slices where the center slice must contain at least one voxel of lesion mask. The local context from the adjacent slices helps improve the segmentation accuracy. The term 2.5D is defined as stacked slices along three orthogonal planes, i.e., axial, sagittal and coronal, providing global context. Compared to 3D patches, our 2.5D stacked slices thus allow both local and global information during training with only modest computational requirements for training.

2.3 Network architecture

Fig. 1 shows the network architecture which is adopted from the Tiramisu model [1]. The model takes concatenation of MRIs from different time-points as input and outputs the segmentation of new lesions. Further details of how we adapt this architecture for MS segmentation can be found in [3].

Our network consists of a downsampling path, an upsampling path and skip connections between the paths. In the downsampling path, a convolutional layer (CONV) and 5 dense blocks (DB) are used to extract the feature maps. The goal of the transition down blocks (TD) is to make various resolutions of the feature maps. Upsampling path is symmetric to the downsampling path except for replacing TD blocks with TU blocks. Upsampling path fuses the information from skip connections and the downsampling path. Before outputting the final segmentation, a Tanh layer is applied.

We used L2 loss for training the networks, defined as $\mathcal{L}_{L_2} = |F(X) - Y|_2$ where X is the inputs of all available modalities and $F(\cdot)$ represents the function of the neural network. Y is the training lesion mask which is the 'ground truth'.

[2] https://anima.irisa.fr/

The L2 loss is able to provide a good trade-off between recall and precision, both voxel-wise and lesion-wise [3].

2.4 Data augmentation

We hypothesize that the data augmentation would help to close the gap between datasets from different sites, leading to more robust MS lesion segmentation. We thus randomly apply data transformations during the training process. These can be classified into three categories:

- **Image quality augmentation.** Image quality is associated with resolution and noise level. We randomly (1) blur, $p = 0.2$, (2) down-sample, $p = 0.1$, (3) sharpen, $p = 0.2$, (4) add Gaussian noise, $p = 0.2$.
- **Image intensity distribution augmentation.** To accommodate the different image appearance between MRIs from different sites, we randomly (1) apply Contrast Limited Adaptive Equalization (CLAHE), $p = 0.2$, (2) adjust brightness and contrast, $p = 0.2$, (3) apply radial gradient brightness, $p = 0.2$, (4) apply horizontal or vertical gradient brightness, $p = 0.2$.
- **Spatial augmentation.** We use spatial augmentations to deal with the high variability of lesion size and shape. We randomly (1) assign an aspect ratio in the range [0.5, 2] if the original ratio is near 1.0, with the probability p = 0.2, (2) reshape based on the ratio, p = 0.8, (3) rotate with angle in [1°, 45°], p = 0.3, (4) distort image, p = 0.3 (5) scale, p = 0.3, (6) crop the image into the training size (128×128), (7) flip, p = 0.3 and (8) transpose, p = 0.3. All MRIs and lesion masks for a given subject/time-point undergo the same spatial-based augmentations.

2.5 Implementation details

The Adam optimizer was used with a momentum term of 0.5 and an initial learning rate of 0.0002. The learning rate was left at the initial value for the first 100 epochs, and then dropped to 0 in 300 epochs. Furthermore, a 5-fold validation strategy was applied. We evaluated the model every 10 epochs and the one with the best Dice + F1 score was chosen as the best model. To get robust predictions, we ensembled all 5-fold models and used models from different settings. For example, we train models with or without augmentations, with different input image sizes, etc. The training process was conducted on NVIDIA GPUs and was implemented using PyTorch.

3 Conclusion

In this work, we propose to use the Tiramisu network with 2.5D stacked slices for new MS lesion detection and segmentation. In the preliminary tests, our method shows promising results.

References

1. Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., Bengio, Y.: The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 11–19 (2017)
2. Zhang, H., Oguz, I.: Multiple sclerosis lesion segmentation - a survey of supervised cnn-based methods. In: Crimi, A., Bakas, S. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 11–29. Springer International Publishing, Cham (2021)
3. Zhang, H., Valcarcel, A.M., Bakshi, R., Chu, R., Bagnato, F., Shinohara, R.T., Hett, K., Oguz, I.: Multiple sclerosis lesion segmentation with tiramisu and 2.5 d stacked slices. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 338–346. Springer (2019)

A subtraction image-based method to detect new appearing multiple sclerosis lesions on single-contrast FLAIR MRI

Francesco La Rosa[1,2,3], Jean-Philippe Thiran[1,3], and Meritxell Bach Cuadra[3,2,1]

[1] LTS5, Ecole Polytechnique Fédérale de Lausanne, Switzerland
[2] Medical Image Analysis Laboratory, CIBM, University of Lausanne, Switzerland
[3] Radiology Department, Lausanne University Hospital, Switzerland

Abstract. We propose an unsupervised method for the segmentation of new appearing MS lesions in FLAIR MRI. Starting with the subtraction image between two time-points, we enforce spatial constraints as well as radiomics features to improve the final segmentation. Results show that our method achieves a detection rate similar to the experts' one, while having, however, a higher false positive rate.

1 Introduction

Multiple sclerosis (MS) is an autoimmune demyelinating disease affecting the central nervous system. Magnetic resonance imaging (MRI) is a key imaging tool for assessing MS diagnosis, progression and therapy response, with a focus on cortical and white matter lesions (WMLs) dissemination in space and time [8]. Studies have shown that the appearance of new MS lesions might reflect inflammatory and degenerative processes and it is crucial for treatment's decisions [7,10]. This is the main motivation of the MSSEG-2 challenge, which evaluates methods for the detection and segmentation of new MS lesions in fluid attenuation inversion recovery (FLAIR) images. In this short paper, we present our submission to the challenge. Our method is unsupervised, fully-automated, and relies on the subtraction image with some spatial and radiomics constraints.

2 Methodology

2.1 Dataset

The MSSEG-2 challenge's dataset includes the MRIs of 100 patients, each with two time-points (acquired from 1 to 3 years apart). For each time-point a 3D FLAIR image is available. Forty patients' scans are available to the participants for training and the remaining 60 are kept for the testing phase. Overall, 15 MRI scanners (both at 1.5T and 3T) from three brands were used to acquire the data. The images provided are already co-registered to the mid-space between the two time-points. No additional pre-processing is performed. The annotations of new appearing lesions from 4 experts, as well as their consensus, are provided.

2.2 Pipeline

Our pipeline is inspired by a previous work proposing the subtraction image as starting point to segment new appearing lesions in multi-contrast MRI [4]. Significant changes, however, were made to adapt it to a single FLAIR contrast as input and to enhance its performance. An overall scheme of our approach is shown in Figure 1. First, we run the longitudinal version of SAMSEG [3]. This is a tool, based on a generative model, which is included in Freesurfer, and that robustly segments dozens of brain structures and WML. It can run also on multi-contrast data and it achieves an accurate segmentation even with a single T2-weighted sequence (in our case FLAIR) [3]. Second, each skull-stripped image is normalized with zero mean and unit standard deviation. Third, the subtraction image is obtained between the second and the first time-point, respectively. Fourth, the quality of the image co-registration is estimated by the mutual information metric computed with Ants [1]. Depending on this value, different thresholds are applied to the subtraction image, where we also discard all the connected components smaller than 5 voxels. Fifth, some spatial constraints are applied: we keep only the connected components which overlap for at least 1/10 of their volume with SAMSEG's WM segmentation, and discard those touching the cerebrospinal fluid (CSF), gray matter and other structures, including MS lesions of the first time-point. Sixth, connected components whose contrasts with the neighbour voxels in the second time-point is lower than 0.9 are discarded. Seventh, we discard lesions with a least axis length smaller than 1.2 and a sphericity lower than 0.2, computing these radiomics features with the Pyradiomics package [9]. The segmentation obtained in this way is our final output. The pipeline takes about 2 hours to run on a Ubuntu 18.04 machine with 64Gb of RAM.

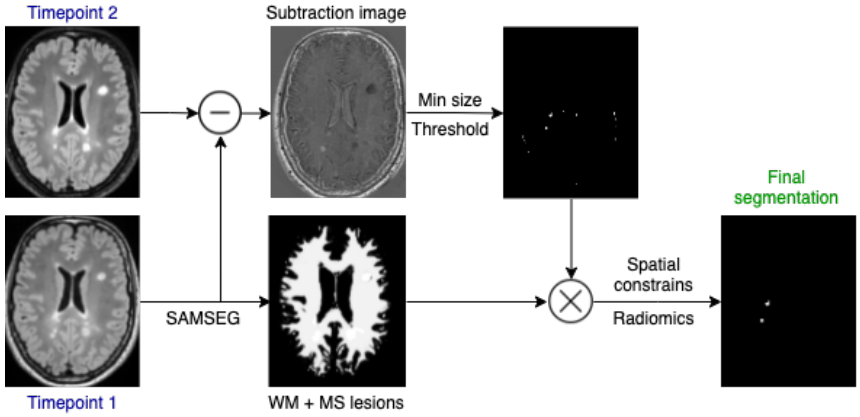

Fig. 1. Simplified scheme of the processing pipeline proposed.

3 Results

We evaluated our proposed method considering the lesion-wise true positive rate (LTPR), lesion-wise false positive rate (LFPR), and Dice coefficient. The experts' consensus was considered as ground truth. The results of our proposed method were compared with SAMSEG's lesion segmentation (considering lesions in the second time-point non-overlapping with lesions in the first one) and the four experts' annotations.

Table 1. Lesion-wise TPR, FPR and Dice coefficient for the training dataset. In bold the best value per each metric.

	LTPR	LFPR	Dice
Expert 1	**0.88**	0.11	**0.53**
Expert 2	0.68	**0.09**	0.45
Expert 3	0.74	**0.09**	0.47
Expert 4	0.69	**0.09**	0.46
SAMSEG-long [3]	0.17	0.71	0.12
Proposed method	0.75	0.38	0.36

4 Discussion

As of today, deep learning-based techniques represent the state-of-the-art methods for MS cortical and white matter lesion segmentation [5,6], as well as for the longitudinal lesion assessment [2]. However, given that the MSSEG-2 challenge dataset, includes very heterogeneous data (acquired with 15 different scanners at both 1.5T and 3T) and a relative small number of lesions (only 153 new appearing lesions), we propose an unsupervised method simply based on the subtraction image of the two time-points. We added spatial constraints which helped reducing the false positives in case of slight miss-alignment between the two time-points. Moreover, some radiomics features enforcing an ovoid shape for the new appearing lesions helped further improving the performance. The results obtained for the training set (see Table 1) show that in terms of detection rate our proposed method is close to the single experts' sensitivity, but it generates more false positives. Overall, it greatly outperforms the new lesions' segmentation generated by SAMSEG. In the testing set, we expect similar, but perhaps slightly worse results as a few parameters of the pipeline were tuned based on the training cases. Future work could include training a convolutional neural network to reduce the high number of false positives.

Acknowledgments. This project is supported by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie project TRABIT (agreement No 765148), the Novartis Research Foundation, the

Centre d′Imagerie BioMédicale of the University of Lausanne, the Swiss Federal Institute of Technology Lausanne, the University of Geneva, the Centre Hospitalier Universitaire Vaudois, and the Hôpitaux Universitaires de Genève.

References

1. Avants, B.B., Tustison, N., Song, G.: Advanced normalization tools (ants). Insight j **2**(365), 1–35 (2009)
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. Cerri, S., Hoopes, A., Greve, D.N., Mühlau, M., Van Leemput, K.: A longitudinal method for simultaneous whole-brain and lesion segmentation in multiple sclerosis. In: Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-oncology, pp. 119–128. Springer (2020)
4. Ganiler, O., Oliver, A., Diez, Y., Freixenet, J., Vilanova, J.C., Beltran, B., Ramió-Torrentà, L., Rovira, À., Lladó, X.: A subtraction pipeline for automatic detection of new appearing multiple sclerosis lesions in longitudinal studies. Neuroradiology **56**(5), 363–374 (2014)
5. La Rosa, F., Abdulkadir, A., Fartaria, M.J., Rahmanzadeh, R., Lu, P.J., Galbusera, R., Barakovic, M., Thiran, J.P., Granziera, C., Bach Cuadra, M.: Multiple sclerosis cortical and WM lesion segmentation at 3T MRI: a deep learning method based on FLAIR and MP2RAGE. NeuroImage: Clinical p. 102335 (Jun 2020). https://doi.org/10.1016/j.nicl.2020.102335, `https://linkinghub.elsevier.com/retrieve/pii/S2213158220301728`
6. La Rosa, F., Beck, E.S., Abdulkadir, A., Thiran, J.P., Reich, D.S., Sati, P., Cuadra, M.B.: Automated detection of cortical lesions in multiple sclerosis patients with 7t mri. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 584–593. Springer (2020)
7. Meier, D.S., Weiner, H., Guttmann, C.R.: Mr imaging intensity modeling of damage and repair in multiple sclerosis: relationship of short-term lesion recovery to progression and disability. American Journal of Neuroradiology **28**(10), 1956–1963 (2007)
8. Thompson, A.J., Banwell, B.L., Barkhof, F., Carroll, W.M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M.S., et al.: Diagnosis of multiple sclerosis: 2017 revisions of the mcdonald criteria. The Lancet Neurology **17**(2), 162–173 (2018)
9. Van Griethuysen, J.J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G., Fillion-Robin, J.C., Pieper, S., Aerts, H.J.: Computational radiomics system to decode the radiographic phenotype. Cancer research **77**(21), e104–e107 (2017)
10. Vrenken, H., Jenkinson, M., Horsfield, M., Battaglini, M., Van Schijndel, R., Rostrup, E., Geurts, J., Fisher, E., Zijdenbos, A., Ashburner, J., et al.: Recommendations to improve imaging and analysis of brain lesion load and atrophy in longitudinal studies of multiple sclerosis. Journal of neurology **260**(10), 2458–2471 (2013)

Team NeuroPoly: Description of the Pipelines for the MICCAI 2021 MS New Lesions Segmentation Challenge

Uzay Macar[1,2*], Enamundram Naga Karthik[1,2*], Charley Gros[1,2], Andréanne Lemay[1,2], Julien Cohen-Adad[1,2,3]

[1] NeuroPoly Lab, Institute of Biomedical Engineering, Polytechnique Montréal, Montréal, QC, Canada
[2] MILA - Québec AI Institute, Montréal, QC, Canada

[3] Functional Neuroimaging Unit, CRIUGM, Université de Montréal, Montréal, QC, Canada

Abstract. This paper gives a detailed description of the pipelines used for the challenge including the data preprocessing steps applied. Our pipelines have sufficient overlap in terms of the architecture and the parameters used, hence are all described within this paper. Our code for this work can be found at: https://github.com/ivadomed/ms-challenge-2021.

1. Introduction

Magnetic Resonance Imaging (MRI) is routinely used for the diagnosis of multiple sclerosis (MS), which is amongst the most prevalent neurological diseases. There are, however, urgent unmet needs for an earlier diagnosis and better monitoring of MS. In particular, segmenting a new lesion in a patient is extremely helpful for choosing the best treatment for this particular patient early on. Segmenting new lesions requires: (i) MRI data of the same participant across multiple time points and (ii) a radiologist who can identify new lesions by studying the MRI scans across multiple time points. This task is time-consuming and prone to error given the high dimensionality of the data. Moreover, in some cases, multiple contrasts are available, which further augments the difficulty of the visual task. Deep learning (DL) methods have proven useful for the segmentation of MS lesions in the brain [1] and spinal cord [2] MRI scans. However, some challenges remain, including the high class imbalance and out-of-distribution generalization.

In this paper, we propose three pipelines for the segmentation of new lesions from the brain and cervical spinal cord MRI scans. The 3D U-Net [3] architecture is fundamental to each of our three pipelines. Since this architecture was not originally developed for MRI, it required substantial modifications to be compatible with 3D kernels and a segmentation task with high class imbalance.

2. Related Work

Recent literature discusses some interesting methods for automatically segmenting MS lesions, thanks to the proliferation of DL techniques and the availability of structured data through open-source challenges. For instance, Kruger et al. [4] present an interesting method for automatically segmenting new MS lesions using a 3D convolutional encoder-decoder architecture. The encoder was pretrained with the task of finding lesions in single timepoint FLAIR scans and these pretrained weights were loaded during the actual training on longitudinal data containing baseline and follow-up FLAIR scans. The network was then tasked with producing a segmentation mask containing the locations of the new lesions. For ensuring stability and convergence during training, deep supervision [5] was introduced by adding segmentation layers in the expansive path of the network.

The winning team in the 2015 challenge on longitudinal white matter lesion segmentation of MS [6] used 3D convolutional neural networks (CNNs) with multi-channel 3D patches of MR volumes were used as the input. A separate CNN was trained for each GT and the final segmentation was obtained by combining the output voxel probabilities of these CNNs.

In most of the DL-based segmentation pipelines, the models are evaluated using binary ground truth (GT) data in the form of 0 and 1 voxel values, which typically fail to capture the rich surrounding information present in the experts' segmentations. This can manifest itself in terms of uncertainty in voxel values [1], inter-expert ambiguity [7], partial volume information (PVE) [8]. As a result, the outputs of these models are uncalibrated and do not generally represent the true confidence of the models' segmentation capability. In this regard, SoftSeg [9] presents the case of using "soft" GT masks instead of the conventional binary ("hard") GTs. An improvement in the Dice score for brain lesion and tumour segmentation tasks was reported when the models were trained with soft GT masks. The potential loss of "soft" (PVE) information by using sharp activation functions (such as sigmoid) as the final layers is also discussed and normalized ReLU is proposed as the better alternative. Therefore, our design choices for the pipelines used in this challenge are primarily inspired by [9].

3. Materials and Methods

3.1. Data

The dataset for the challenge consists of a total of 100 patients (40 for training, 60 for testing) each with two fluid-attenuated inversion recovery (FLAIR) images (baseline and follow-up), four ground-truth masks each annotated by an expert neuroradiologist, and a consensus ground-truth mask formed via a senior expert neuroradiologist confirming or declining the disputed lesions and fusing the final masks by majority voting. 15 different MRI scanners were used in the acquisition process, and the resulting dataset is highly variable with voxel dimensions ranging

from 0.6x0.5x0.5 mm^3 to 1.2x1.2x1.2 mm^3 (rounded to nearest decimal) and image sizes ranging from (144, 192, 192) to (320, 512, 512) across the two timepoints.

3.2. Preprocessing

We use the same preprocessing step described in this section in all of the proposed pipelines in order to increase the quality of the inputs to the pipelines described in the later sections. The preprocessing step is applied independently to each of the 40 patients provided to us in the training set, and it utilizes the following packages with versions provided next to them in parenthesis: FSL (5.0.11), Spinal Cord Toolbox (SCT, 5.3.0), ANTs (2.2.0), and ANIMA (4.0.1). The entire preprocessing stage is described below.

Initially, FLAIR images from each session and their associated GT masks were resampled to the isotropic resolution of 0.5x0.5x0.5 mm^3. The target resolution is isotropic because MS lesions have no preferred spatial orientation. The choice of voxel size was a compromise between the segmentation precision and the computational burden. For the next pre-processing steps, we split the analyses between the brain and the spinal cord (SC) given their different shape and co-registration constraints (i.e. quasi-rigid for the brain, non-linear for the spinal cord). We extracted the SC from both sessions and performed an initial co-registration step on the baseline FLAIR image using the segmentation masks of the SC and considering the follow-up FLAIR image as the reference. This first co-registration step was carried out with `sct_register_multimodal` from SCT and it used the slice-by-slice regularized registration algorithm [10] with the mean squares metric and a smoothing kernel of 3x3x3 mm^3. The brain was extracted using `bet2` from FSL from the follow-up FLAIR image, and the dilated versions of the brain and SC masks were summed. On observing that new lesions were annotated near the spinal cord region, we decided to incorporate the SC masks also. We then perform a finer co-registration step on the base FLAIR image using the binarized version of the joint brain-SC mask and taking the follow-up FLAIR image as the reference. This second co-registration step was carried out with `antsRegistration` from ANTs, using the neighborhood cross-correlation metric with symmetric normalization transform (SyN), shrink factor of 8x4x2 mm^3, and B-Spline interpolation method. The registered baseline and follow-up FLAIR images were then masked using the joint brain-SC mask, and the N4 bias field correction algorithm was applied to both images. The bias field correction step was carried out using `animaN4BiasCorrection` from ANIMA. Finally, the two FLAIR images were cropped using the joint brain-SC mask such that they only include the volume-of-interest.

After performing the above preprocessing step, the maximum image sizes (across all 40 subjects) along the sagittal, coronal, and axial axes in MNI space, were found to be 311, 391, and 487, respectively. In order to assess the performance of the final co-registration and the preprocessing step in general, we visualized the results in a before-after fashion for each of the 40 patients. The visualization is available at this link: https://bit.ly/3nMpXFM.

3.3. Data Augmentations

The preprocessed baseline and follow-up FLAIR images with their associated GT masks were initially center cropped according to the maximum image sizes mentioned in the previous subsection, and then divided into subvolumes of size S with a stride-size of R. In our pipelines, we have mostly experimented with $S \in \{32, 64, 128, 256\}$ and $R \in \{32, 64, 96, 128\}$. We applied the following data augmentations on the subvolumes while training all of our models: (i) random left-right (lateral) flipping, (ii) random affine transformations from `ivadomed` [11], (iii) random elastic transformation from `ivadomed`, and (iv) random MRI bias-field artifacts from `torchio` [12]. Subsequently, we normalized all images to zero mean and unit variance. Additionally, we use a naive data balancing strategy while training all of our models which duplicates each positive subvolume as many times as required to reach a balanced set (i.e. equivalent numbers) of positive and negative subvolumes. This is applied once before the first epoch in each training process. We observed that balancing the dataset helped with convergence, even though it increased the runtime of the training process. Duplicating subvolumes also poses the potential problem of overfitting, yet the data augmentations described above diversified the samples to a sufficient degree and effectively prevented this problem from occurring.

3.4. Challenge pipelines

This section describes the different pipelines we used in this challenge. Each pipeline features an adaptation of the 3D U-Net [3] architecture for the purpose of segmenting MS lesions. A detailed description of each pipeline is provided below.

Pipeline #1: Modified 3D U-Net

A U-Net-based architecture was chosen due to its exceptional generalization capabilities achieved by propagating contextual information from the contractive (downsampling) path to the higher resolution layers in the expansive (upsampling) path. The 3D U-Net [3] was proposed as the three-dimensional generalization of U-Net [13] for volumetric segmentation of the *Xenopus* Kidney from sparsely annotated maps. All the 2D operations in the standard U-Net architecture were replaced with their 3D counterparts. We propose to adapt this 3D U-Net model for the challenge by introducing some modifications to the network architecture described below.

The core of the architecture contains four downsampling and upsampling operations. In the contractive path, the blocks at each scale contain 2 sets of 3x3x3 convolutional and instance normalization (IN) layers with strides 2 and 1, respectively, to enable downsampling. A LeakyReLU activation layer is added between each convolutional and IN set. In the expansive path, instead of using transposed convolution layers, the nearest neighbour upsampling is used to avoid the checkerboard artifacts [14]. Each upsampling layer was sandwiched between blocks containing 3x3x3 convolutional of stride 1, IN, and leakyReLU layers. 1x1x1 convolutional layers were used at each scale to match the dimensions of the feature maps concatenated from the contractive

path. It is important to note that the original 3D U-Net paper used batch normalization and ReLU activation layers. IN was chosen in this work to account for small batch sizes. The normalized ReLU was used as the final activation layer as inspired by [9], where a ReLU activation is applied to the output and normalized by its maximum value to get an output between 0 and 1, thereby obtaining a full range of prediction values. While the sigmoid layer also squashes its input between 0 and 1, it tends to narrow the range of soft values that carry PVE information required for a good soft prediction. Dice Loss [15] was used as the loss function.

Pipeline #2: Ensemble of 3D U-Nets

An ensemble of 4 U-Net models was used in our second pipeline. 3 of these models were Modified 3D U-Net as described in the first pipeline, and the last model was an Attention 3D U-Net as described in more detail in the third pipeline. The final prediction was obtained by taking the soft average of 4 segmentations (one per model). Each U-Net model was trained and validated independently with a different data split, hence diversifying the final prediction in the hopes of achieving better generalization performance.

Pipeline #3: Attention 3D U-Net with Monte-Carlo Dropout

Building on top of the first pipeline, our third pipeline features a Modified 3D U-Net supplemented with attention gate (AG) blocks as proposed by [16]. Our motivation to use attention blocks arises from the fact that the challenge dataset contains varying shapes of lesions which constitute a small percentage of the subvolumes they belong to. In this scenario, attention can help the network learn lesion structures more efficiently and in a more informed way by highlighting salient regions and suppressing irrelevant regions.

Typically, using a single set of best weights from the model results in predicted segmentations with high variance. In order to mitigate this issue, we additionally used Monte-Carlo (MC) Dropout [17]. The idea is to use dropout during inference such that when each test input is stochastically forward-passed through the network T times (where T refers to the number of MC samples), we obtain a probability distribution over the output space. By taking the mean of these T samples, we get the averaged segmentation output.

3.5. Training & Evaluation

The dataset was split into three parts: a held-out test set of volumes, a validation set of subvolumes, and a training set of subvolumes. First, a random subset of 8 subjects out of 40 was chosen as the held-out test set. Then, the remaining 32 subjects had their subvolumes extracted and a second split was applied where 80% of them were randomly chosen as the training set, and the remaining 20% became the validation set.

In each pipeline, we trained our models with the soft Dice score and saved the model with the best soft Dice score computed on the validation set after 100 epochs. The

evaluation of our pipeline, mainly for purposes of model selection, then consisted of two stages: (i) soft and hard Dice score metrics computed on the subvolumes in the validation set, and (ii) Dice score, Jaccard score, positive predictive value (PPV), F1 score, and SurfaceDistance metrics computed on the volumes in the test set as implemented by `animaSegPerfAnalyzer` in ANIMA.

References

1. Nair T, Precup D, Arnold DL, Arbel T. Exploring uncertainty measures in deep networks for Multiple sclerosis lesion detection and segmentation. Med Image Anal. 2020;59: 101557.

2. Gros C, De Leener B, Badji A, Maranzano J, Eden D, Dupont SM, et al. Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. Neuroimage. 2019;184: 901–915.

3. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. arXiv [cs.CV]. 2016. Available: http://arxiv.org/abs/1606.06650

4. Krüger J, Opfer R, Gessert N, Ostwaldt A-C, Manogaran P, Kitzler HH, et al. Fully automated longitudinal segmentation of new or enlarged multiple sclerosis lesions using 3D convolutional neural networks. Neuroimage Clin. 2020;28: 102445.

5. Dou Q, Yu L, Chen H, Jin Y, Yang X, Qin J, et al. 3D deeply supervised network for automated segmentation of volumetric medical images. Med Image Anal. 2017;41: 40–54.

6. Carass A, Roy S, Jog A, Cuzzocreo JL, Magrath E, Gherman A, et al. Longitudinal multiple sclerosis lesion segmentation: Resource and challenge. Neuroimage. 2017;148: 77–102.

7. Carass A, Roy S, Gherman A, Reinhold JC, Jesson A, Arbel T, et al. Evaluating White Matter Lesion Segmentations with Refined Sørensen-Dice Analysis. Sci Rep. 2020;10: 8242.

8. Chaves H, Dorr F, Costa ME, Serra MM, Slezak DF, Farez MF, et al. Brain volumes quantification from MRI in healthy controls: Assessing correlation, agreement and robustness of a convolutional neural network-based software against FreeSurfer, CAT12 and FSL. J Neuroradiol. 2021;48: 147–156.

9. Gros C, Lemay A, Cohen-Adad J. SoftSeg: Advantages of soft versus binary training for image segmentation. Med Image Anal. 2021;71: 102038.

10. De Leener B, Lévy S, Dupont SM, Fonov VS, Stikov N, Louis Collins D, et al. SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. Neuroimage. 2017;145: 24–43.

11. Gros C, Lemay A, Vincent O, Rouhier L, Bourget M-H, Bucquet A, et al. Ivadomed: A medical imaging deep learning toolbox. J Open Source Softw. 2021;6: 2868.

12. Pérez-García F, Sparks R, Ourselin S. TorchIO: a Python library for efficient loading,

preprocessing, augmentation and patch-based sampling of medical images in deep learning. Comput Methods Programs Biomed. 2021; 106236.

13. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv [cs.CV]. 2015. Available: http://arxiv.org/abs/1505.04597

14. Odena A, Dumoulin V, Olah C. Deconvolution and Checkerboard Artifacts. Distill. 2016;1. doi:10.23915/distill.00003

15. Milletari F, Navab N, Ahmadi S-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. arXiv [cs.CV]. 2016. Available: http://arxiv.org/abs/1606.04797

16. Oktay O, Schlemper J, Folgoc LL, Lee MJ, Heinrich, M, Misawa K, Mori K, McDonagh SG, Hammerla N, Kainz B, Glocker B, Rueckert D. Attention U-Net: Learning Where to Look for the Pancreas. arXiv [cs.CV]. 2018. Available: https://arxiv.org/abs/1804.03999

17. Gal Y, Ghahramani Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. arXiv [stat.ML]. 2016. Available: https://arxiv.org/abs/1506.02142

MSSEG-2 Challenge, Team New Brain: Cascaded networks for new MS lesion detection

Berke Doga Basaran[1], Paul M. Matthews[2], and Wenjia Bai[1,2]

[1] Data Science Institute, Department of Computing, Imperial College London, London, UK
[2] Department of Brain Sciences, Imperial College London, London, UK

1 Introduction

Multiple sclerosis (MS) is an inflammatory and demyelinating neurological disease of the central nervous system [1]. Image-based biomarkers, such as lesions defined on magnetic resonance images (MRI), play an important role in MS diagnosis and patient monitoring. The detection of newly formed lesions provides crucial information for assessing disease activity and treatment outcome [2]. However, it is not an easy task to delineate these new lesions. Some of the new lesions are small and subtle. Manual segmentation is tedious and prone to human errors with inter-rater variabilities. In this paper, we propose a novel automated method for new MS lesion segmentation. The developed method is based on a cascade of two segmentation networks, the first generating a lesion probability map for both timepoints of a longitudinal scan and the second segmenting new lesions formed in the second timepoint. We present the results on the MSSEG-2 challenge dataset.

2 Methods

The proposed method consists of an image preprocessing stage and two segmentation stages, using the U-Net convolutional neural networks. The first network produces a probability map for all MS lesions, trained on the ISBI 2015 MS Lesion Segmentation Challenge dataset (ISBI2015) [3] and the MICCAI 2016 MS Lesion Segmentation Challenge dataset (MS2016) [4]. The second network detects only new lesions, trained on the provided dataset by the MSSEG-2 challenge.

2.1 Image Preprocessing

Each subject in the MSSEG-2 challenge dataset contains longitudinal brain MR scans at two timepoints, t_1 and t_2. We regard the image at t_2 as the reference frame and the image at t_1 is registered onto it using affine transformation, followed by free form deformation using MIRTK [5]. Following registration, a brain mask is extracted from the t_2 scan using the FSL toolbox [6] and then applied to both t_1 and t_2 scans to remove non-brain tissues. Afterwards, the images are resampled to the average voxel spacing in the dataset (0.5 x 0.98 x 0.98 mm^3) and intensity normalised.

2.2 Stage One

In the first stage, a U-Net is trained to detect all lesions on both the t_1 and t_2 scans. FLAIR images from the ISBI2015 and MS2016 datasets are used to train this segmentation network.

The U-Net implemented has five downsampling and five upsampling paths, no dropout and uses instance normalisation. The network is trained using the Adam optimizer and a cross-entropy loss for 5,000 epochs. Data augmentation is performed, including flipping, rotation, image rescaling, gamma augmentation, elastic deformation and intensity alteration. A 3D patch size of $128 \times 128 \times 128$ is used, due to GPU memory limits. Patches are randomly sampled from the lesions and background with 50% probability. At testing time, the images are passed through the network via a tiling strategy.

Rather than saving the binary segmentation, the 2-channel output of the network, corresponding to the background (healthy) and foreground (lesion) classes, is saved as a lesion probability map. The difference between the lesion probability maps at t_2 and t_1 is calculated, which will be fed into the second stage network as an input channel. Hyperintense regions in this difference map signify where there may be new lesions in the brain.

2.3 Stage Two

The second network takes in two input channels: the lesion probability difference map from the first network and the intensity image at t_2.

The Stage Two U-Net implemented is identical to the Stage One U-Net by hyperparameters. However, due to time constraints gamma augmentation was not performed. The best model is selected based on the average Dice accuracy of the validation set. Four models are trained using different folds of the training data. Test images are passed through each model and their outputs are ensembled, generating the new lesion segmentation result.

3 Experiments and Results

The objective of this challenge is to detect and accurately segment new lesions at the second timepoint of longitudinal FLAIR MR scans. Only new lesion masks at t_2 are provided by the challenge. Lesions at t_1 and repeat lesions at t_2 are not available, which forms a challenge for model training. To overcome this challenge, additional datasets are used, including ISBI2015 [3] and MS2016 [4]. The Stage One network, which segments all lesions, is trained on 17 training images (13 from MS2016, 4 from ISBI2015) and validated on 3 images (2 from MS2016, 1 from ISBI2015). It achieved a validation DICE accuracy of 74.3%.

The Stage Two network, which segments new lesions, is trained using the provided MSSEG-2 dataset. Unlike the ISBI2015 and MS2016 datasets, the MSSEG-2 dataset contains lesions located outside of the brain white matter, with lesions at the cerebellum or spinal cord in certain cases. Furthermore, due

to the small size and location of some of the new lesions, new lesions do not appear as hyperintense as repeat lesions. For model training, patients in the MSSEG-2 dataset with no new lesions (ids: 15, 19, 49, 51, 52, 68, 70, 84, 89, 90, 96) are removed from the dataset. The remaining 29 cases are split into four groups to train separate models, which then form an ensemble model.

To avoid overfitting to the ISBI2015 dataset, only the first timepoint of each case in the dataset is used during Stage One. The Stage One network is trained on 17 images (13 from MS2016, 4 from ISBI2015) and validated on 3 (2 from MS2016, 1 from ISBI2015). In Stage Two, all 29 patients with new lesions from the MSSEG-2 dataset are used in the training process.

The model used to create the probability maps in Stage One achieved a validation Dice accuracy of 74.3%. The average Dice accuracy of the four folds trained in the Stage Two network for detecting new lesions was initially 53.3%. Upon visual inspection, the model exhibited very few false positives, yet showed general undersegmentation. The low Dice score was due to false negatives, from less hyperintense lesions, or lesions not in the brain white matter. The threshold was for lesion classification was decreased from 50% to 10%, increasing the final Dice score up to 62.5%. Figure 1 provides an example pipeline segmentation.

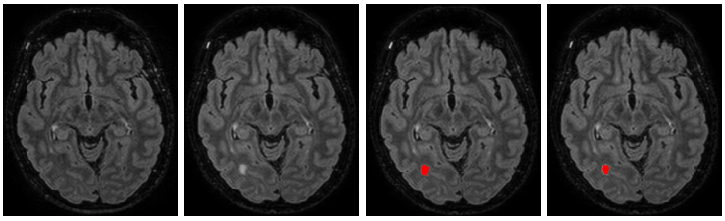

Fig. 1: **Case 69. (a)** Scan at t_1. **(b)** Scan at t_2. **(c)** Ground truth. **(d)** Segmentation of the proposed method.

4 Discussion

Unlike traditional single network architectures, the proposed cascaded networks utilise the rich information generated from a model trained to segment all lesions and create a probability map of where new lesions may be. While the method works well for lesions which appear hyperintense in the white matter region, it does not address lesions which have grown from t_1 to t_2. However, false positives are rarely observed.

Secondly, some cases exhibit new lesions in the spinal cord. These lesions are not detected by the proposed method due to the brain extraction in image pre-processing. The motivation of brain extraction is to remove non-brain tissues in lesion segmentation and improve the alignment for images between different timepoints. Future work may extend the current method to perform lesion segmentation in the spinal cord region.

Finally, it is worth noting that the proposed method is able to detect lesions missed by human analysts. For example, A new lesion in Case 13 is not labelled but segmented by the proposed method and verified by a second reader.

Acknowledgements This work was supported by the UKRI CDT in AI for Healthcare http://ai4health.io (Grant No. EP/S023283/1)

References

1. Filippi, M., Rocca, M. A., Barkhof, F., Brück, W., Chen, J. T., Comi, G., DeLuca, G., De Stefano, N., Erickson, B. J., Evangelou, N., Fazekas, F., Geurts, J. J., Lucchinetti, C., Miller, D. H., Pelletier, D., Popescu, B. F., Lassmann, H., & Attendees of the Correlation between Pathological MRI findings in MS workshop (2012). Association between pathological and MRI findings in multiple sclerosis. The Lancet. Neurology, 11(4), 349–360. https://doi.org/10.1016/S1474-4422(12)70003-0
2. C. Elliott, D. L. Arnold, D. L. Collins and T. Arbel, "Temporally Consistent Probabilistic Detection of New Multiple Sclerosis Lesions in Brain MRI," in IEEE Transactions on Medical Imaging, vol. 32, no. 8, pp. 1490-1503, Aug. 2013, doi: 10.1109/TMI.2013.2258403.
3. Carass, A., Roy, S., Jog A., Cuzzocreo J. L., Magrath E., et al. Longitudinal multiple sclerosis lesion segmentation: Resource and challenge. NeuroImage, 148(December 2016):77–102, 2017
4. Commowick, O., Istace, A., Kain, M., Laurent, B., Leray, F., Simon, M., Pop, S. C., Girard, P., Améli, R., Ferré, J. C., Kerbrat, A., Tourdias, T., Cervenansky, F., Glatard, T., Beaumont, J., Doyle, S., Forbes, F., Knight, J., Khademi, A., ... Barillot, C. (2018). Objective Evaluation of Multiple Sclerosis Lesion Segmentation using a Data Management and Processing Infrastructure. Scientific Reports, 8(1), 1–17. https://doi.org/10.1038/s41598-018-31911-7
5. Medical Image Registration ToolKit (MIRTK), https://mirtk.github.io/. Last accessed 24 Jun 2021
6. Jenkinson M., Beckmann C.F. , Behrens T.E.J., Woolrich M.W., SmithS.M., FSL, NeuroImage, Volume 62, Issue 2, 2012, Pages 782-790, ISSN 1053-8119, https://doi.org/10.1016/j.neuroimage.2011.09.015.

Robust 3D MRI Segmentation of Multiple Sclerosis Lesions

Md Mahfuzur Rahman Siddiquee and Andriy Myronenko

NVIDIA, Santa Clara, CA
{mdmahfuzurr,amyronenko}@nvidia.com

1 Introduction

Multiple sclerosis (MS) affects the central nervous system and is the most common immune-mediated disorder [2]. About 2.3 million people, with widely varying rates in different regions and among different populations, were affected globally in 2015 [8, 6]. About 18,900 people died from MS in the same year, increased from 12,000 in 1990 [8]. In this context, interest in automatic detection of MS has grown and MS segmentation challenge (MSSEG2) is playing a key-role in development of such methods. MSSEG2 provides a dataset containing 40 training subjects. Each of these subjects have 2 MRI volumes at 2 different time points. The objective of the challenge is to build models to segment MS lesions automatically. As a part of the effort, we have build a segmentation solution consisting an ensemble of 10 models. These models are built on a shallow and a deep version of SegResNet [5]. Our design choice of the proposed solution was mainly influenced by the tiny shape of the MS lesions. In the following sections, we will describe our methods in details.

2 Method

2.1 The Network

We implemented our approach with MONAI[1] [1]. We use the encoder-decoder backbone based on [5] with an asymmetrically larger encoder to extract image features and a smaller decoder to reconstruct the segmentation mask [7, 9, 10]. We have trained 2 different network architectures in our solution. They are depicted in Fig. 1.

Encoder part The encoder part uses ResNet [3] blocks. We have combined 10 models for final predictions. For the first set of 5 models, we have used 2 stages of down-sampling, stages having 4 and 4 blocks, respectively. In the second set of 5 models, we have used 4 stages of down-sampling, stages having 1, 2, 2, and 4 blocks, respectively. We have used batch normalization and ReLU. Each block's output is followed by additive identity skip connection. We follow a common

[1] https://github.com/Project-MONAI/MONAI

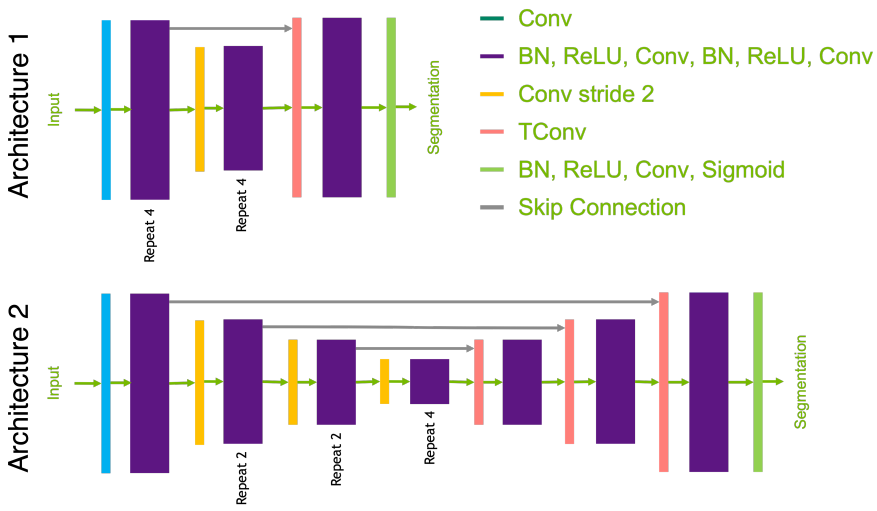

Fig. 1. Network architectures used in our solution.

CNN approach to progressively downsize image dimensions by 2 and simultaneously increase feature size by 2. For downsizing we use strided convolutions. All convolutions are 3x3x3 with initial number of filters equal to 64 for the first set of 5 models and 32 for the rests. The encoder is trained with different input regions—for the first set of 5 models, it is $16 \times 16 \times 16$ and $128 \times 128 \times 128$ for the rests.

Decoder part The decoder structure is similar to the encoder one, but with a single block per each spatial level. Each decoder level begins with upsizing with transposed convolution: reducing the number of features by a factor of 2 and doubling the spatial dimension, followed by an addition of encoder output of the equivalent spatial level. The end of the decoder has the same spatial size as the original image, and the number of features equal to the initial input feature size, followed by 1x1x1 convolution into 2 channels and a softmax.

2.2 Dataset

We have used the MSSEG2 dataset only for training the model. We have randomly split the entire dataset into 5-folds and trained a model for each.

2.3 Loss

We have used Dice loss for training [4].

2.4 Optimization

We use AdamW optimizer with initial learning rate of $2e^{-4}$ and decrease it to zero at the end of final epoch using Cosine annealing scheduler. The model is trained of 4 GPUs, each GPU optimizing for batch-size of 512 for the first set of 5 models and batch-size of 1 for other 5 models. However, we have calculated batch normalization across all the GPUs. We have ensembled 10 models for the submission. All the models were trained for 200 epochs.

2.5 Regularization

We use L2 norm regularization on the convolutional kernel parameters with a weight of $1e^{-5}$.

2.6 Data preprocessing and augmentation

We normalize all input images to have zero mean and unit std (based on nonzero voxels only). We have applied random flip on each axis, random contrast adjustment, and random brightness data augmentations.

3 Results on Cross-Validation

Our cross-validation results on the 5-folds can be found in Tab. 1.

Fold 1	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Average**
0.5635	0.5407	0.3813	0.5481	0.5073	0.5082

Table 1. Average DICE using 5-fold cross-validation.

4 The Team

Team Name: NVAUTO
Team Members: Md Mahfuzur Rahman Siddiquee, Andriy Myronenko

References

1. Project-monai/monai, https://doi.org/10.5281/zenodo.5083813
2. Berer, K., Krishnamoorthy, G.: Microbial view of central nervous system autoimmunity. FEBS letters **588**(22), 4207–4213 (2014)
3. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: European conference on computer vision. pp. 630–645. Springer (2016)

4. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. IEEE (2016)
5. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization. In: International MICCAI Brainlesion Workshop. pp. 311–320. Springer (2018)
6. Organization, W.H., et al.: Atlas: multiple sclerosis resources in the world 2008. World Health Organization (2008)
7. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
8. Vos, T., Allen, C., Arora, M., Barber, R.M., Bhutta, Z.A., Brown, A., Carter, A., Casey, D.C., Charlson, F.J., Chen, A.Z., et al.: Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the global burden of disease study 2015. The lancet **388**(10053), 1545–1602 (2016)
9. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp. 3–11. Springer (2018)
10. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging **39**(6), 1856–1867 (2019)

Consensus Learning with Multi-Rater Labels for Segmenting and Detecting New Lesions

Brennan Nichyporuk[1], Kirill Vasilevski[1], Anjun Hu[1], Chelsea Myers-Colet[1], Jillian Cardinell[1], Justin Szeto[1], Jean-Pierre Falet[1,2], Eric Zimmermann[1], Julien Schroeter[1], Douglas L. Arnold[2], and Tal Arbel[1]

[1] CIM & MILA, McGill University, Canada
[2] Department of Neurology and Neurosurgery, McGill University, Canada
brennann@cim.mcgill.ca

Abstract. We present a robust method to detect new MS lesions for the 2021 MSSEG-2 challenge. The proposed method consists of a set of networks whereby each neural network (NN) is trained using individual rater labels. The predictions from these individual rater NNs are then used to train another model to generate the final consensus label for a given input image. Further, each NN is trained with a lesion reweighting strategy to manage the tradeoff between lesion detection and segmentation, ensuring that small lesions are not overlooked by the network. Lastly, feature-wise linear modulation is used to condition each NN on the resolution of the image, maintaining the information content of each input image while being robust to the inconsistent resolutions in the dataset.

Keywords: multi-rater segmentation · lesion detection · lesion segmentation · ensemble · feature-wise linear modulation

1 Introduction

Identification of new and enlarging lesions in Multiple Sclerosis (MS) patients is an important task in medical image analysis as they reflect disease activity and can provide important information on treatment response [7]. We propose a multi-stage ensemble of NNs for new MS lesion segmentation and detection. Our approach leverages the information provided by each individual expert and the consensus rater to explicitly model the process used to generate the final consensus labeling. Specifically, we propose an ensemble of identical U-Net based networks [2], modified to include: (1) lesion size reweighting [8] and (2) feature-wise linear modulation (FiLM) [11]. The former is implemented in order to ensure accurate segmentation and detection of small lesions, while the latter accounts for different image acquisition parameters.

2 Method

2.1 Modeling Raters and Consensus

In the context of medical image segmentation problems, there can be significant variability between the predictions generated from different experts [5]. To

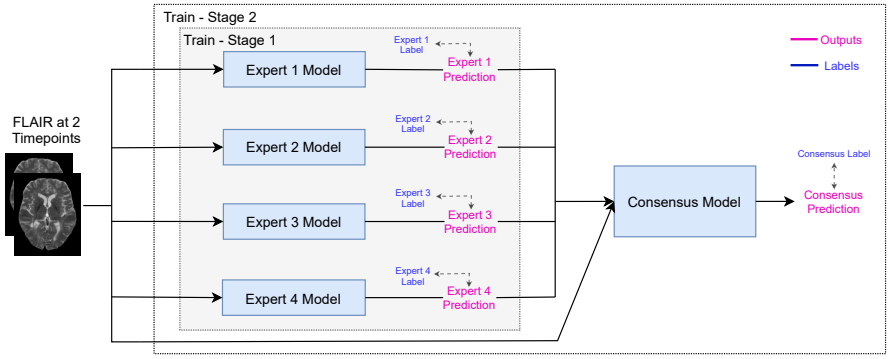

Fig. 1. System overview showing our two-phase training approach. Stage 1 (shown in the inner block) shows how the expert models are trained. Stage 2 (shown in the outer block) shows how the consensus model is trained with predictions from the expert models.

produce a final labeling that is as consistent as possible, it is common practice for a senior expert radiologist to form a consensus label by analyzing the labels produced by each expert [13]. This process may reduce the variability in the final labeling, but it also excludes information that may be present in the labels of individual experts. Indeed, previous work has shown that using only consensus labels in segmentation challenges results in overconfidence compared to utilizing labels from individual raters [3, 4].

We propose to explicitly model the consensus process. Specifically, our approach involves training a separate nnU-Net [2] to model each expert. The predictions from each of the four expert models are then cascaded into a fifth model that models the consensus process. Models are trained in a two-stage process, as shown in Figure 1. The inner box shows the Stage 1 models, where each independent network is trained based on the labels from the corresponding expert rater. During Stage 2 of training, each expert model is fixed, and then used to produce predictions in their expert's style to train the consensus nnU-Net. The consensus nnU-Net learns to produce a consensus-style label from the labels generated by each expert nnU-Net along with the two input FLAIR images. Effectively, the consensus nnU-Net is trained to model the combination of the simple voting procedure and the advanced expert corrections done during the generation of the consensus label.

2.2 Lesion Size Reweighting

The large size variability of new MS lesions often renders the traditional binary cross entropy (BCE) loss function inadequate for combined segmentation and detection tasks, regularly oversegmenting large lesions or missing small lesions entirely [8]. To train a model that is robust in detecting and segmenting lesions of all sizes, we employ the lesion reweighting strategy originally proposed by

Nichyporuk *et al.* [8]. This strategy reweights each voxel as a function of the size of its associated lesion, with a weight of 1 given to non-lesion voxels. By employing this strategy, we can balance the weight of small lesions such that: (1) they are given enough importance to be detected by the model and (2) they are given less importance than large lesions affording accurate segmentation performance.

2.3 Conditioning on Resolution

As medical imaging datasets can be composed of images acquired using a combination of different scanner models, sub-groups within a dataset can be biased by different scanner-specific contrast levels and resolutions, which can cause problems when training NNs [6]. While contrast variations can be accounted for more easily, resampling to a common resolution comes with a loss of information when downsampling and interpolation errors when upsampling [10]. To avoid these pitfalls, we use FiLM [11] to condition the network on the image resolution of each sample. FiLM is able to modulate the activations of the NN, tailoring predictions to a given resolution. Our FiLM layer is implemented as follows:

$$\text{FiLM(z)} = (\gamma + \gamma_r)\left(\frac{z - \mu(z)}{\sigma(z)}\right) + (\beta + \beta_r)$$

where γ and β are general learnable affine parameters. γ_r and β_r are a function of the image resolution and are modeled using an MLP.

2.4 Implementation Details

The base model underlying our approach is a nnU-Net [2]. Preprocessing includes N4 bias correction [12], Nyul normalization [9], and symmetric diffeomorphic image registration [1]. Final predictions are generated by using the cross-validated models (k=4) as an ensemble.

3 Conclusions

We present a two-stage NN consensus learning method designed to leverage label uncertainty across individual expert annotators to produce a more robust consensus model for new MS lesion labels. Additionally, we propose using a lesion-size reweighting strategy to ensure balanced lesion detection independent of lesion size. Finally, we use feature-wise linear modulation to condition each instance normalization layer on the image resolution, improving robustness to scanner associated resolution differences.

Acknowledgements This work was funded by the CIFAR AI Chairs Program, the Natural Sciences and Engineering Research Council of Canada (NSERC), and an award from the International Progressive MS Alliance (PA-1603-08175).

References

[1] B. B. Avants et al. "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain". In: *Medical image analysis* (Feb. 2008).

[2] Fabian Isensee et al. "No new-net". In: *International MICCAI Brainlesion Workshop*. 2018.

[3] Martin Holm Jensen et al. "Improving uncertainty estimation in convolutional neural networks using inter-rater agreement". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2019.

[4] Wei Ji et al. "Learning Calibrated Medical Image Segmentation via Multi-Rater Agreement Modeling". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

[5] Leo Joskowicz et al. "Inter-observer variability of manual contour delineation of structures in CT". In: *European radiology* (2019).

[6] Neerav Karani et al. "A lifelong learning approach to brain MR segmentation across scanners and protocols". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2018.

[7] Ulrike W. Kaunzner, Mais Al-Kawaz, and Susan A. Gauthier. "Defining Disease Activity and Response to Therapy in MS". In: *Current Treatment Options in Neurology* (Apr. 2017).

[8] Brennan Nichyporuk et al. "Optimizing Operating Points for High Performance Lesion Detection and Segmentation Using Lesion Size Reweighting". In: *MIDL*. 2021.

[9] L.G. Nyul, J.K. Udupa, and Xuan Zhang. "New variants of a method of MRI scale standardization". In: *IEEE Transactions on Medical Imaging* (2000).

[10] J. Anthony Parker, Robert V. Kenyon, and Donald E. Troxel. "Comparison of Interpolating Methods for Image Resampling". In: *IEEE Transactions on Medical Imaging* (Mar. 1983).

[11] Ethan Perez et al. "FiLM: Visual Reasoning with a General Conditioning Layer". In: *AAAI*. 2018.

[12] N. J. Tustison et al. "IEEE Trans Med ImagingN4ITK: improved N3 bias correction". In: *IEEE Transactions on Medical Imaging* (June 2010).

[13] Simon K Warfield, Kelly H Zou, and William M Wells. "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation". In: *IEEE transactions on medical imaging* (2004).

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

Marius Schmidt-Mengin[1,2], Arya Yazdan-Panah[1], Théodore Soulier[1], Mariem Hamzaoui[1], Bruno Stankoff[1], Nicholas Ayache[3], 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.

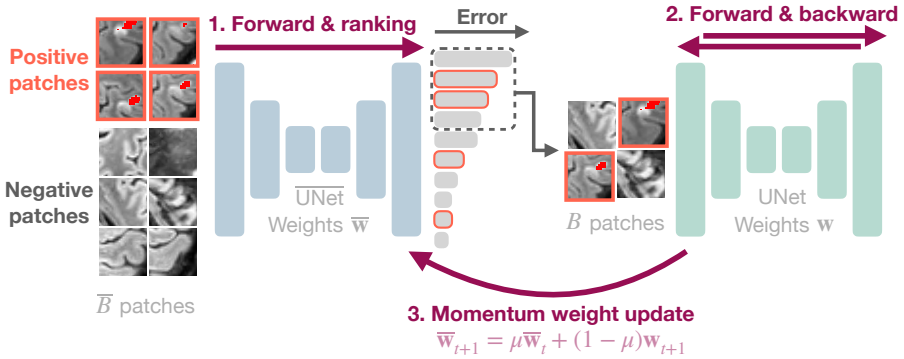

Fig. 1. Illustration of our training strategy. $\overline{B}$ patches are fed to a first 3D U-Net and the segmentation errors are computed for each patch. The patches are ranked according to their errors, and the top_B patches are selected to perform a training step with a second 3D U-Net. The weights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

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

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 $\overline{B} = 128$ 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 $\overline{\texttt{UNet}}$, 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

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

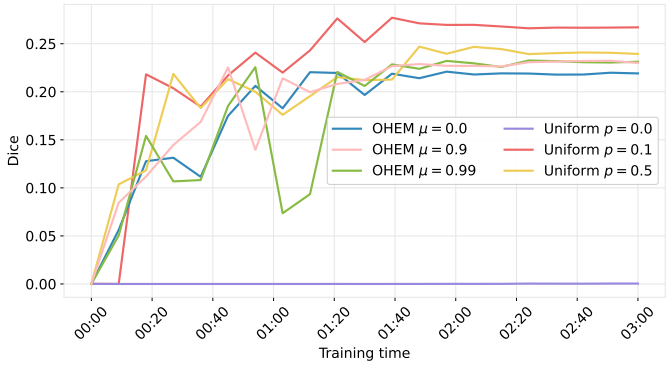

Fig. 2. Dice on the validation set during training. "Uniform" means that positive patches are sampled with probability p, and that with probability $1 - p$ any other patch (not necessarily negative) is selected. In the end, uniform sampling with $p > 0$ ended up performing best, and $\mu > 0$ seems to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

Intensity based Regions Of Interest (ROIs) preselection followed by Convolutional Neuronal Network (CNN) based segmentation for new lesions detection in Multiple Sclerosis

Mariem Hamzaoui [1], Théodore Soulier [1], Arya Yazdan-Panah [1], Marius Schmidt-Mengin [1,2], Olivier Colliot [1,2], Nicholas Ayache [3], Bruno Stankoff [1]

[1] Sorbonne Université, Institut du Cerveau, Inserm-CNRS, Paris, France
[2] Inria – Aramis project-team, Paris, France
[3] Inria – Epione project-team, Sophia-Antipolis, France
arya.yazdan-panah@icm-institute.org

Abstract. Detecting new lesions is a key aspect of the radiological follow-up of patients with Multiple Sclerosis (MS), leading to eventual changes in their therapeutics. Our pipeline for new lesion detection based on two consecutive FLAIR MRIs consists in two steps. We start by a detection of potential Region Of Interest (ROI) containing new lesions using a sensitive classical image processing procedure. It detects the connected voxels which intensity increases in the second visit compared to the first one. We then apply a filtering procedure to segment voxels corresponding to new lesions in those potential ROI by applying Convolutional Neuronal Networks (CNN).

Keywords: classical image processing, CNNs, hybrid approach.

Introduction

The reduction or absence of new lesion formation over time is a key radiological endpoint in clinical trials assessing disease modifying therapies in Multiple Sclerosis. Novel lesion identification and segmentation is generally done manually, or using semi-automated procedures, by radiologists or neurologists and is time consuming. The aim of MICCAI MSSEG-2 Worskhop was to develop a new automatic method to segment new lesions based on two FLAIR MRI of the same patient.

1 Step one: MRIs post-processing for potential ROI detection

The aim of the first step (Fig. 1) is to detect the ROIs that could potentially contain a new lesion by a sensitive classical image processing procedure, so that those ROIs could be used as patches for the next step. Images were corrected for bias field with the N4 algorithm using "animaN4BiasCorrection" [1]. To enhance the quality of the registration between the two time points [3], we applied a rigid registration, using ANTs [2]. We extracted a brain mask using 'bet' (FSL [4]) on both images and used their intersection to restrict the analysis. Finally an affine registration of the two FLAIRs was done using ANTs.

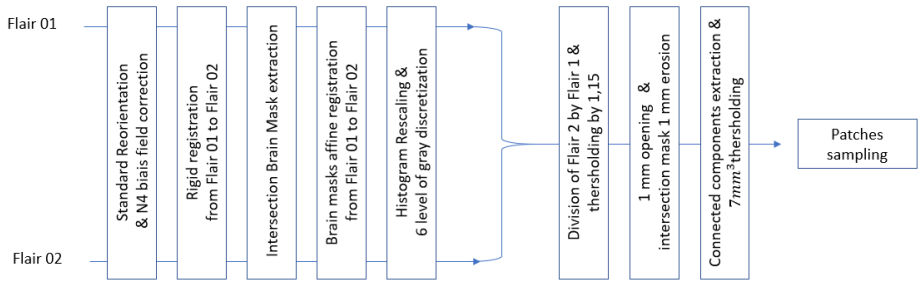

Fig. 1. Step one overview

We applied a two-step intensity matching procedure. First, we used a histogram matching from ANTs ("ImageMath"), based on landmarks intensity scaling between the two images, considering the second visit as the reference. To further reduce the small variations in the signal, intensities were discretized on 6 levels of grey.

New lesions are defined as parenchymal areas that are iso-intense at the first visit but hyperintense at the second. When considering the ratio of the two visits, as intensities have been matched, new objects should consist in areas with an intensity ratio greater than 1. By extracting ratio values in the ground truth, we chose a threshold of 1.15 to include 85% of ground truth voxels.

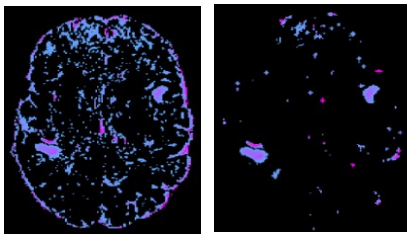

Fig. 2. map of all voxels which intensity ratio between the two visits is greater to 1.15, before (left) and after (right) connected components distinction and volume thresholding (>7mm³).

Due to remaining field bias, registration innacuracies at the interfaces, a high number of voxels external to the ground truth mask are still selected by this method (Fig. 2a). We mark them as False Positive (FP) voxels, as opposed to True Positive (TP) voxels, that are included in the ground truth mask. We added a geometrical constraint to those candidate voxels to eliminate FPs. We used an opening of one millimeter to individualize adjacent structures. Then, we proceeded to distinguish connected components (CC) ("animaConnectedComponents" [1]) and deleted those below 7 mm³, thus eliminating isolated voxels. We eroded by 1mm the brain mask to avoid ROIs in the pial surface.

Each resulting CC was used as a target to crop a patch with a padding of two voxels, defining our potential ROI (Fig. 2 right). In all patches we sampled the processed FLAIR from the first and second visit after discretization, the ground truth, and three distance maps: from pial surface to brain depth (DM), from vertex to neck (DMz) and from occiput to nose (DMx). Those distance maps will be inserted into convolutional layers in later steps to give anatomical references for segmentation.

2 Step two: Filtering of true positive voxels in previously found ROIs by CNN

2.1. Batch constitution: infratentorial imbalance

We used the torchIO [5] to generate batches. Each training batch included 32 patches as described above. We identified infratentorial patches based on a threshold in the DMz. The fewer number of infratentorial new lesions, smaller volume of this area and the high complexity of the brain anatomy in this region, lead us to imbalance our training toward infratentorial regions. In the training set, we selected with a uniform probability, as many FP patches as TP patches. To increase the specificity of the segmentation and decrease the number of FP voxels, notably in infratentorial areas, we added to the training set 30% of extra patches with a higher probability of being infratentorial and FP patches. In detail, those extra patches were constituted by 85% of infratentorial patches (33% TP and 67% FP infratentorial patches) and 15% of supratentorial FP patches. We used the data augmentation methods provided by the torchIO library, applying random noise and random bias field with, each, a 30% probability.

2.2. Network architecture

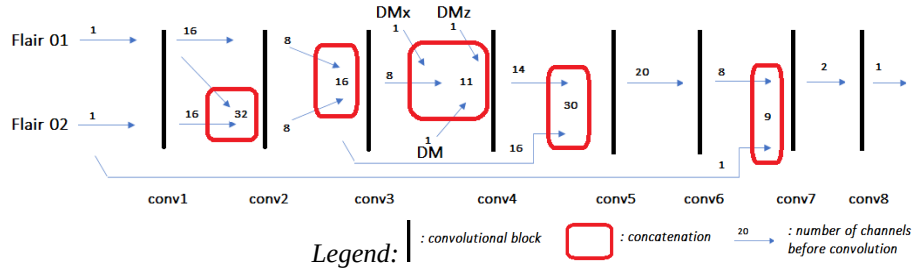

Fig. 3. CNN architecture for patch segmentation

The architecture shown in fig. 3, was implemented with Pytorch [7]. Each convolutional block was formed by a 3D convolutional layer (in and out channels described in Fig. 3) without striding and with a padding of one, followed by a 3D batch normalization, and finally a Rectified Linear Unifier activation function. Due to the variable patch sizes, no pooling layers were added [6]. Between conv3 and conv4, we added the three distance maps created in step one. There are two recurrent connections: at short distance between conv2 and conv5, and at long distance between Flair 02 input image and conv7. This second long-distance recurrent connection was given to help segment at the last steps by concatenating conv6's output to the input Flair 02, on which the final segmentation has to be based on.

2.3. Training

The absence of pooling layers in our architecture (fig. 3) is known to increase the risk of over-fitting [6]. In consequence, we limited our training to 20 epochs, with a

decreasing learning rate using an ADAM optimizer, beginning from a learning rate of 0.001 at the first epoch and decaying it by 0.1 every 4 epochs.

2.4. Inference

As patches could overlap, for each voxel of the global image, we kept only the value from the output patch where this voxel was the closest to the patch's center. Indeed, as each patch was cropped around its own potential ROI, peripheral voxels are more at risk of belonging to an adjacent ROI caught in the field of view. Moreover, as we observed that bigger patches contained more FP voxels, taking the distance from center as a discriminator is in favor of smaller patches in the final prediction. We applied a sigmoid to the prediction and binarized the final image with threshold of 0.75.

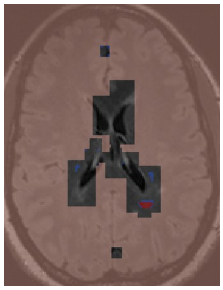

Fig. 4. Inference with all patches overlaid on Flair 02 image, before sigmoid, with low probabilities (in blue) and high probabilities (in red), the only ones kept in the final binarized image.

Acknowledgements

The research leading to these results was made possible by the combined efforts of several parties including: IHU-A-ICM, and the 3IA Côte d'Azur (ANR-19-P3IA-0002), that we would like to express our gratitude to. we also thank Mr DIAZ MELO (ARAMIS team, Paris Brain Institute) for his helps with the dockers.

References

1. tool provided by MICCAI MSSEG2 Workshop
2. Advanced Normalization Tools, http://stnava.github.io/ANTs/
3. Mehdi Hadj-Hamou, Marco Lorenzi, Nicholas Ayache, Xavier Pennec. 'Longitudinal Analysis of Image Time Series with Diffeomorphic Deformations: A Computational Framework Based on Stationary Velocity Fields' Front. Neurosciences. 2016 Jun 3;10:236
4. FMRIB Software Library, http://ftp.nmr.mgh.harvard.edu/pub/dist/freesurfer/tutorial_packages/OSX/fsl_501/doc/index.html
5. Torch IO 0.18.41, https://torchio.readthedocs.io/
6. L. Alzubaidi et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future Directions. J Big Data. 2021; 8(1): 53.
7. PyTorch, https://pytorch.org/

Detection of lesion change in multiple sclerosis using a cascade of 3D-to-2D networks

Richard McKinley, Franca Wagner, Roland Wiest

Support Centre for Advanced Neuroimaging, University Institute of Diagnostic and Interventional Neuroradiology, Inselspital, Bern University Hospital, Bern, Switzerland

Abstract. We present a method for detecting new FLAIR lesions in patients with Multiple Sclerosis, developed using the training dataset of the MSSEG2 dataset. The individual FLAIR images are skull-stripped, and then initially segmented using an adaptation of our previously published DeepSCAN method. The skull-stripped FLAIR images and soft lesion masks are then fed to a second network with similar structure which identifies the new lesions.

Introduction

Markers of lesion progression are routinely assessed by radiologists in clinical routine [3]. However, there is a considerable variation in lesion quantification between expert readers with different backgrounds (e.g. neuroradiologists vs. neuroimmunologists), and also substantial intra-rater variation [2]. This variation is only compounded in the context of longitudinal imaging.

In the 2016 MSSEG lesion segmentation challenge at MICCAI, CNN-based approaches won both arms of the challenge [8,9,1]. Subsequent papers have used the data arising from this challenge as a reference dataset, improving on the results achieved. Robustness of these methods when applied to datasets from centers distinct from the training set is vital: most methods submitted to the MSSEG challenge failed to generalize to data from an unseen center. One solution is to tune an existing method to new data, which can be achieved using surprisingly few cases. [10] This is not always practical: however, modern, fully-convolutional networks generalize well even when trained on cases from a single centre. [6]

Given a method for single-timepoint lesion segmentation, it is tempting to think that the problem of detecting new or growing lesions is a trivial step. However, naive methods based on lesion counting and lesion volumetry do a very poor job at identifying progressive timepoints. [7]. This problem is made even harder if the goal (as in the MSSEG2 challenge) is to identify only new lesions, and not growing lesions: morphology cannot help, since new lesions may be contiguous

with existing lesions. For this purpose, supervised learning seems ideal, and the MSSEG2 dataset provides a labeled dataset of serial images where labeled voxels correspond to new lesions.

Our approach is a *cascade* of two models: one to detect lesions in each timepoint, and one to find the new lesions. We leverage an existing lesion segmentation model to generate lesion probability masks. These, plus the FLAIR images, are fed to a second classifier, trained on the MSSEG2 dataset, which segments the new lesions.

1 Methods

1.1 Skull stripping

In principle high-performing deep learning methods should function without skull-stripping: however our base method was originally trained on skull-stripped images. In addition, in our experience skull-stripping increases robustness across different sites/scanners/protocols. For skull-stripping in this pipeline we utilize the HD-BET algorithm. [4]

1.2 3D-to-2D network for lesion and new lesion segmentation: DeepSCAN

For lesion segmentation in one timepoint, and detection of new lesions, in two timepoints, we use a DeepSCAN neural network. The DeepSCAN architecture is a '3D-to-2D' architecture, meaning that the network takes as input a 3D volume, and predicts labels for the central 2D slice of that volume. The 3D-to-2D structure of DeepSCAN's layers combines the benefits of a fully 3D network (3D context) and a 2D network (smaller memory footprint of the input, leading to more available memory for feature maps)

The DeepSCAN architecture for MS lesion and neuroanatomy segmentation is shown in Figure 1 . Details of the initial training can be found in the original paper. [6] This network was trained to segment lesions from T1-weighted, T2-weighted, and T2-FLAIR imaging with 1mm isovoxel resolution. To adapt the network for the MSSEG2 challenge, we resumed training while randomly dropping one or two channels as data augmentation. This yielded a network able to segment lesions and other brain structures from T2-FLAIR imaging alone.

We then trained a network with identical structure to Figure 1, but with four input channels (the two skull-stripped FLAIR sequences, plus the lesion probability maps obtained from our lesion segmentation model, at 1mm isotropic resolution), and one output channel, the new lesion class as derived from the MSSEG2 training set. Nonzero FLAIR voxels were standardized on a per-volume

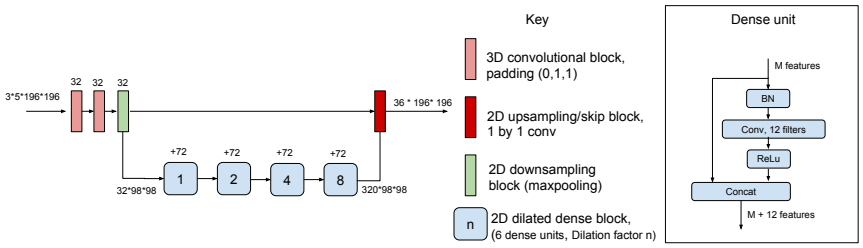

Fig. 1. The DeepSCAN architecture for lesion and brain-structure segmentation, from [6]

basis by subtracting the mean nonzero intensity and dividing by the standard deviation of the voxels of nonzero intensity. No standardization or postprocessing was applied to the lesion segmentation masks. The loss function was focal loss.[5] The network was trained with a Cosine annealing with restarts learning rate schedule, with a maximum learning rate of 1e-5, a minimum learning rate of 1e-7, and twenty epoch of 5000 batches.

When applied to a new case, the classifier is applied in the axial, coronal and sagittal direction. Minimal test-time augmentation is performed by also segmenting the volumes mirrored in the sagittal direction. An ensembling of these six outputs is then performed by averaging raw logit outputs. Before applying the model, the skull-stripped timepoint 1 is rigidly registered to the skull-stripped timepoint 2: this is mostly unnecessary when using the MSSEG2 dataset, but may correct registration problems caused by tissue outside the brain.

1.3 Results

We trained the new lesion detection model in cross-validation across all training cases in the MSSEG dataset, where it achieved a mean Dice coefficient of 0.56.

References

1. Commowick, O., Istace, A., Kain, M., Laurent, B., Leray, F., Simon, M., Pop, S., Girard, P., Améli, R., Ferré, J., Kerbrat, A., Tourdias, T., Cervenansky, F., Glatard, T., Beaumont, J., Doyle, S., Forbes, F., Knight, J., Khademi, A., Mahbod, A., Wang, C., McKinley, R., Wagner, F., Muschelli, J., Sweeney, E., Roura, E., Lladó, X., Santos, M., Santos, W., Silva-Filho, A., Tomas-Fernandez, X., Urien, H., Bloch, I., Valverde, S., Cabezas, M., Vera-Olmos, F., Malpica, N., Guttmann, C., Vukusic, S., Edan, G., Dojat, M., Styner, M., Warfield, S., Cotton, F., Barillot, C.: Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. Scientific Reports 8(1) (12 2018)

2. Erbayat Altay, E., Fisher, E., Jones, S., Hara-Cleaver, C., Lee, J.C., Rudick, R.: Reliability of classifying multiple sclerosis disease activity using magnetic resonance imaging in a multiple sclerosis clinic. JAMA Neurol 70, 338–44 (03 2013)
3. Graber, J.J., Dhib-Jalbut, S.: Biomarkers of disease activity in multiple sclerosis. Journal of the neurological sciences 305 1-2, 1–10 (2011)
4. Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H., Heiland, S., Wick, W., et al.: Automated brain extraction of multisequence mri using artificial neural networks. Human Brain Mapping 40(17), 4952–4964 (Aug 2019), http://dx.doi.org/10.1002/hbm.24750
5. Lin, T.Y., Goyal, P., Girshick, R.B., He, K., Dollár, P.: Focal loss for dense object detection. 2017 IEEE International Conference on Computer Vision (ICCV) pp. 2999–3007 (2017)
6. McKinley, R., Wepfer, R., Aschwanden, F., Grunder, L., Muri, R., Rummel, C., Verma, R., Weisstanner, C., Reyes, M., Salmen, A., Chan, A., Wagner, F., Wiest, R.: Simultaneous lesion and brain segmentation in multiple sclerosis using deep neural networks. Scientific Reports 11 (2021)
7. McKinley, R., Wepfer, R., Grunder, L., Aschwanden, F., Fischer, T., Friedli, C., Muri, R., Rummel, C., Verma, R., Weisstanner, C., Wiestler, B., Berger, C., Eichinger, P., Muhlau, M., Reyes, M., Salmen, A., Chan, A., Wiest, R., Wagner, F.: Automatic detection of lesion load change in multiple sclerosis using convolutional neural networks with segmentation confidence. NeuroImage: Clinical 25, 102104 (2020), http://www.sciencedirect.com/science/article/pii/S2213158219304516
8. McKinley, R., Wepfer, R., Gundersen, T., Wagner, F., Chan, A., Wiest, R., Reyes, M.: Nabla-net: A deep dag-like convolutional architecture for biomedical image segmentation. In: Crimi, A., Menze, B., Maier, O., Reyes, M., Winzeck, S., Handels, H. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 119–128. Springer International Publishing, Cham (2016)
9. Valverde, S., Cabezas, M., Roura, E., Gonzalez-Villa, S., Pareto, D., Vilanova, J.C., Ramio-Torrenta, L., Rovira, A., Oliver, A., Llado, X.: Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. Neuroimage 155, 159–168 (07 2017)
10. Valverde, S., Salem, M., Cabezas, M., Pareto, D., Vilanova, J.C., Ramió-Torrentà, L., Rovira, A., Salvi, J., Oliver, A., Lladó, X.: One-shot domain adaptation in multiple sclerosis lesion segmentation using convolutional neural networks. NeuroImage. Clinical p. 101638 (2018)

Convolutional Neural Network for MS new lesions segmentation challenge: using a data management and processing infrastructure (MSSEG-2)

Isabella Medeiros de Sousa[1][0000-0002-9577-7458] and Marcela de Oliveira[1,2][0000-0003-4144-7629]

[1] INESC TEC and Faculty of Engineering, University of Porto, Porto 4200-465, Portugal
[2] School of Sciences - São Paulo State University (UNESP), Bauru SP 17033-360, Brazil
up201900023@edu.fe.up.pt
marcela.oliveira@unesp.br

Abstract. Magnetic resonance imaging (MRI) is the most used exam for diagnosis and follow-up of multiple sclerosis (MS). MS is a neuroinflammatory and neurodegenerative disease characterized by demyelination of neuron axon, which causes lesions in white matter that can be observed in vivo by MRI. Such lesions may provide quantitative assessments of the inflammatory activity of the disease. The visualization of new lesions in longitudinal MRIs, is used as marker of disease activity and an outcome measure in clinical trials. Although manual segmentations are considered as the gold standard, this process is time consuming and error prone. Therefore, automated new lesion identification and quantification of the MRI are active areas and challenging in MS research. The purpose of this study was to perform the new lesions segmentation from follow-up scans of the same subject with MS, via a convolutional neural network (CNN) model. This method uses a subtraction of the MRI scans as network input and a segmentation mask of the new lesion as an output.

Keywords: Convolutional Neural Network, New Lesion, Multiple Sclerosis.

1 Introduction

Multiple Sclerosis (MS) is a demyelinating disease of the central nervous system, which shows inflammatory white matter pathology (lesions) [1]. These brain lesions can be observed in vivo by magnetic resonance imaging (MRI), and may provide quantitative assessments of the inflammatory activity of the disease. In addition, longitudinal quantitative characterization of lesion load and relevant tissue volumetry have a central importance to clinical evaluations aimed at monitoring disease progression and evaluating the efficacy of the disease modifying therapy [2]. The visualization of new lesions in longitudinal MRIs, is used as marker of disease activity and an outcome measure in clinical trials [3]. However, to detect new lesions from MRIs can be very challenging, because they are very small and their appearance can be subtle [4].

Several studies have shown the detection of new lesions in T2 sequences, but Fluid attenuated inversion recovery (FLAIR) sequences have begun to complement or even replace standard T2-weighted imaging [5]. McDonald guidelines emphasize the importance of lesion location rather than the number of lesions, and white lesions are best visualized with FLAIR sequences [6]. Manual segmentation of lesion from follow-up MRI scans is still considered the gold standard, but it is time consuming and inter-rater

variability is high [7]. In recent decades, several studies have applied machine learning to segmentation of lesions in MRI of patients with MS [2], [8]–[11]. However, most of these studies perform the lesion detection in just one moment, without considering different times for the same patient. In 2017, Carass et al. in conjunction with the ISBI 2015 conference, organized a longitudinal lesion segmentation challenge providing training and test data to registered participants [12]. The participating teams had different segmentation techniques and good results, but a disadvantage of this challenge was the use of a small database (five subjects for training and fourteen for testing).

A particular deep learning technique, Convolutional Neural networks (CNNs), has gained popularity as a promising method for detecting MRI lesions in MS patients [5]. The most used networks follow a U-Net like architecture with a down layer and up layer structure [13]. In this case, the network uses an MRI scans as input and a segmentation mask of the lesion as an output [5], [14]. Thus, this paper presents a convolutional neural network framework for detecting new FLAIR lesions from multi-modal reference and follow-up scans of the same subject for MS patients.

Our proposed segmentation network, CNNNewLesion, requires the availability of one modality (FLAIR MRI) and follow up timepoints and FLAIR lesion labels at reference. This method was developed and applied for the Longitudinal Multiple Sclerosis Lesion Segmentation Challenge of MICCAI 2021: MSSEG-2 challenge on MS new lesions segmentation challenge using a data management and processing infrastructure. Our proposed segmentation network (CNNNewLesion) can be found in different versions. Older versions of this same method can be found available on the VIP platform (CNNNewLesion and CNNNewLesion2) or as a Docker image (cnnmsseg:1.5 and cnnmsseg:2.5). But in these older versions the input images from the network must already be preprocessed. The final version of our method is CNNNewLesionFinal (VIP platform) or cnnmsseg:4.2 (Docker image), and in this case the method performs both preprocessing, including skull-stripper, and the new lesion segmentation.

2 Materials and Methods

2.1 Patient Sample and Image Data

This study included one dataset composed of MR neuroimaging data of 100 patients selected from the HD Cohort (http://www.ofsep.org/en/hd-cohort) of the French Registry on Multiple Sclerosis(http://www.ofsep.org/en/). For each patient, 3D FLAIR images have been acquired at two time points, with varying interval of time between the two scans. Those images were then registered in the intermediate space between the two time points. First, the transformation to go from the space of the first time point to the space of the second time point was computed. Then, half of this transformation was applied on the first time point, and half of the opposite transformation was applied on the second time point. In this way, both images have similar interpolation artifacts. 4 experts manually segmented new Multiple Sclerosis lesions (lesions appearing on the second time point, but not the first). All lesions delineated by three of four experts were selected, and another expert reviewed the remaining lesions (those delineated by two or less experts) to validate or reject them. A voxel wise majority voting was applied to create the final consensus masks. The dataset was split in one training set and one testing set, in such way that 60% of the lesions belong to the testing set.

2.2 Imaging Preprocessing

MRIs from the training group were preprocessed in five steps: 1. skull-stripped and bias corrected; 2. resliced to 1mm^3; 3. anisotropically diffused; 4. normalized and standardized; and 5. resized for input. In the first step the MRIs of subjects in the training were skull-stripped with the Anima library (Anima scripts: RRID:SCR_017072 - https://anima.irisa.fr), which consisted in a brain extraction followed by a bias field correction. In the second step, the MRIs were resliced at isotropic resolution using cubic spline interpolation to the axial 1 mm^3. In the third step, an anisotropic diffusion filter was used for noise reduction [15], [16]. In the fourth step, the image intensities were normalized. This normalization of a particular slice of the modality is performed by subtracting the mean value of that slice and dividing by its standard deviation. And finally, in the fifth step, resizing the slice to the shape (256,256), dimension required for the input image of the network that will segment the new lesions. In this preprocessing task, it is important to mention that the ground truth images also had to go through the isotropic resolution and resize steps, ensuring that the new segmented lesions automatically (network output) would be in the same dimension as the ground truth images.

2.3 New Lesion Segmentation

The new lesion segmentation was performed using the convolutional neural network training model with the dataset from training group. To obtain the training model, we use the subtraction between the two images at different times (longitudinal MRIs) as input images (both images had already gone through the preprocessing described previously in 2.2). The U-Net network architecture (adapted from Ghosal *et al*. 2019 [17]) consists of a contracting path (down-layer) and other symmetric expanding path (up-layer) used to feature extraction and precise localization, respectively, see Fig. 1.

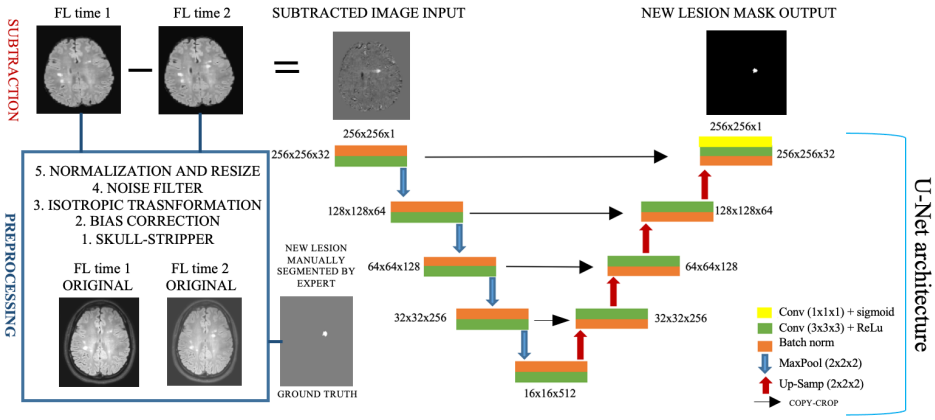

Fig. 1. Flowchart of the proposed method for new lesion volume segmentation.

For the convolution in the down-layer from feature extraction and deconvolution in the up-layer from expanding path, 3 × 3 kernels were used. Rectified linear unit (ReLU) was used as the activation function [18]. A 2 × 2 Max pooling operation (stride 2) was used to reduce the image size in the down-layer [19]. We used the concatenations with the respective cropped feature map from the down-layer into the corresponding input

of the convolutional layer in the up-layer. Thus, the CNN layers learn the transformations from the intensities to the feature maps to obtain the final probabilistic new lesion mask. The output layer was a convolutional layer that detects whether a new brain lesion is or is not present [20]. The CNN were trained with a batch size of 32, the Adam optimizer, and a learning rate of 0.001. Cross-entropy was used as the loss function, and we trained the models for 60 epochs. To test model, we employed a cross-validation strategy of 80:20 split. The Dice coefficient, the accuracy, the sensitivity, the precision, and the specificity were used as performance measures to evaluate the final model. At this point we use the network output images (new segmented lesions) compared to the ground truth (already preprocessed). After that, the training model will be applied to the subtraction MRIs of the test group as input and the new lesions mask will be obtain as output. Our CNNNewLesionFinal method, performs all the preprocessing described in item 2.2 and the segmentation of new lesions by model trained in the test images. Finally, volume quantification can be performed with a count of the segmented voxels and present in mm^3. A schematic overview of the proposed method is in Fig. 1.

2.4 Implementation

We implement the CNN framework in Python with deep learning tools, TensorFlow and Keras. Our final proposed method, CNNNewLesionFinal, starts with two longitudinal images and ends with the segmentation of new sclerosis lesions in brain MRI images, as described above. This final method can be found available on the VIP platform (CNNNewLesionFinal) or as a Docker image (cnnmsseg:4.2). In this latest version the image goes through the resampling process, so the segmented image will be isometric, so for comparison between automatic new segmentation and manual segmentation, the ground truth also must be isometric. Older versions of this same method can be found available on the VIP platform (CNNNewLesion and CNNNewLesion2) or as a Docker image (cnnmsseg:1.5 and cnnmsseg:2.5).

References

[1] M. Styner *et al.*, "3D segmentation in the clinic: A grand challeng," *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, vol. 10, pp. 7–15, 2007.

[2] Z. Ding, J. Preiningerova, C. J. Cannistraci, T. L. Vollmer, J. C. Gore, and A. W. Anderson, "Quantification of multiple sclerosis lesion load and brain tissue volumetry using multiparameter MRI: Methodology and reproducibility," *Magn. Reson. Imaging*, vol. 23, no. 3, pp. 445–452, 2005, doi: 10.1016/j.mri.2004.12.005.

[3] U. W. Kaunzner, M. Al-Kawaz, and S. A. Gauthier, "Defining Disease Activity and Response to Therapy in MS," *Curr. Treat. Options Neurol.*, vol. 19, no. 5, p. 20, May 2017, doi: 10.1007/s11940-017-0454-5.

[4] N. M. Sepahvand, D. L. Arnold, and T. Arbel, "CNN Detection of New and Enlarging Multiple Sclerosis Lesions from Longitudinal Mri Using Subtraction Images," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2020, vol. 2020-April, no. c, pp. 127–130, doi: 10.1109/ISBI45749.2020.9098554.

[5] J. Krüger *et al.*, "Fully automated longitudinal segmentation of new or enlarged multiple sclerosis lesions using 3D convolutional neural networks," *NeuroImage Clin.*, vol. 28, p. 102445, 2020, doi: 10.1016/j.nicl.2020.102445.

[6] U. W. Kaunzner and S. A. Gauthier, “MRI in the assessment and monitoring of multiple sclerosis: an update on best practice,” *Ther. Adv. Neurol. Disord.*, vol. 10, no. 6, pp. 247–261, Jun. 2017, doi: 10.1177/1756285617708911.

[7] C. Egger *et al.*, “MRI FLAIR lesion segmentation in multiple sclerosis: Does automated segmentation hold up with manual annotation?,” *NeuroImage Clin.*, vol. 13, pp. 264–270, 2017, doi: 10.1016/j.nicl.2016.11.020.

[8] S. Roy, J. A. Butman, D. S. Reich, P. A. Calabresi, and D. L. Pham, “Multiple sclerosis lesion segmentation from brain MRI via fully convolutional neural networks,” *arXiv*, no. March, 2018.

[9] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, “Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions,” *J. Digit. Imaging*, vol. 30, no. 4, pp. 449–459, Aug. 2017, doi: 10.1007/s10278-017-9983-4.

[10] R. Khayati, M. Vafadust, F. Towhidkhah, and M. Nabavi, “Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and markov random field model,” *Comput. Biol. Med.*, vol. 38, no. 3, pp. 379–390, Mar. 2008, doi: 10.1016/j.compbiomed.2007.12.005.

[11] I. Dimitrov *et al.*, “Brain and Lesion Volumes Correlate With Edss in Relapsing-Remitting Multiple Sclerosis,” *J. IMAB - Annu. Proceeding (Scientific Pap.*, vol. 21, no. 4, pp. 1015–1018, 2015, doi: 10.5272/jimab.2015214.1015.

[12] A. Carass *et al.*, “Longitudinal multiple sclerosis lesion segmentation: Resource and challenge,” *Neuroimage*, vol. 148, pp. 77–102, 2017, doi: 10.1016/j.neuroimage.2016.12.064.

[13] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015*, 2015, pp. 234–241.

[14] A. Danelakis, T. Theoharis, and D. A. Verganelakis, “Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging,” *Comput. Med. Imaging Graph.*, vol. 70, pp. 83–100, 2018, doi: https://doi.org/10.1016/j.compmedimag.2018.10.002.

[15] G. Gerig, O. Kbler, R. Kikinis, and F. A. Jolesz, “Nonlinear Anisotropic Filtering of MRI Data,” *IEEE Trans. Med. Imaging*, vol. 11, no. 2, pp. 221–232, 1992, doi: 10.1109/42.141646.

[16] P. Perona and J. Malik, “Scale-Space and Edge Detection Using Anisotropic Diffusion,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, 1990, doi: 10.1109/34.56205.

[17] P. Ghosal, P. K. C. Prasad, and D. Nandi, “A Light Weighted Deep Learning Framework for Multiple Sclerosis Lesion Segmentation,” in *2019 Fifth International Conference on Image Information Processing (ICIIP)*, Nov. 2019, vol. 2019-Novem, pp. 526–531, doi: 10.1109/ICIIP47207.2019.8985674.

[18] V. Nair and G. E. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines,” 2010, doi: dl.acm.org/doi/10.5555/3104322.3104425.

[19] H. Hwang, H. Z. U. Rehman, and S. Lee, “3D U-Net for Skull Stripping in Brain MRI,” *Appl. Sci.*, vol. 9, no. 3, p. 569, Feb. 2019, doi: 10.3390/app9030569.

[20] M. de Oliveira *et al.*, “Quantification of Brain Lesions in Multiple Sclerosis Patients using Segmentation by Convolutional Neural Networks,” in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Dec. 2020, pp. 2045–2048, doi: 10.1109/BIBM49941.2020.9313244.

Estimating lesion activity through feature similarity: A dual path Unet approach for the MSSEG2 MICCAI challenge

Mariano Cabezas[1][0000−0002−4417−1704], Yuling Luo[2], Kain Kyle[1,2], Linda Ly[2], Chenyu Wang[1,2][0000−0001−7135−7662], and Michael Barnett[1,2]

[1] Brain and Mind Centre, University of Sydney, Sydney NSW 2050, Australia
{mariano.cabezas, kain.kyle, chenyu.wang, michael.barnett}@sydney.edu.au
[2] Sydney Neuroimaging Analysis Centre, University of Sydney, Sydney NSW 2050, Australia
{yuling.luo, kain, linda, tim, michael}@snac.com.au

Keywords: Brain · lesion activity · attention · deep learning

1 Introduction

Magnetic resonance (MR) imaging is a core paraclinical tool for predicting long-term disability and treatment response in multiple sclerosis (MS) patients. Several criteria related to the detection of disease activity on follow-up brain MR imaging studies have been proposed for prompt identification of suboptimal response in individual patients during the first 6 to 12 months after treatment initiation [4,1,6].

However, automatic lesion activity detection is a heavily imbalanced problem: the vast majority of voxels on the brain retain the same tissue appearance over time. Furthermore, changes in the brain can also include the effect of atrophy. On the other hand, it is increasingly difficult to acquire imaging from patients with lesion progression, especially those patients that are under treatment. As a consequence, the amount of training examples is usually small in terms of activity voxels hampering the training of deep neural networks. To that end, in this work we present a novel approach combining self-supervised pre-training on a private dataset without labels to exploit stable voxels, the inclusion of prior knowledge in the form of an attention gate between timepoints and a dual headed Unet architecture. In the following sections, we will provide a detailed explanation of each of these components, including preprocessing.

2 Preprocessing

Errors in the brain boundary are one of the most common sources of false positive detections, either due to inconsistencies on the noise and intensity homogeneity between temporal acquisitions, tissue changes and atrophy or misregistration errors. To reduce these errors, skull stripping is commonly used to focus the

segmentation on a common brain region. Furthermore, since FLAIR is only sensitive to white matter lesions, a rough brain mask is usually enough (even if part of the cortex is missing). Therefore, to speed up inference we used a simple strategy involving Otsu thresholding [3] and morphological operations. In brief, the threshold of both images is used to obtain an initial mask based on the intersection between baseline and follow-up. Afterwards, erosion is used to remove regions connected to the brain (skin and fat around the skull) and select the brain mask as the largest connected component, followed by dilation and operations to fill any holes inside the brain (usually related to the ventricles).

Longitudinal inconsistencies due to intensity inhomogeneity are another common source of false positives. To reduce that effect, we applied the N4 algorithm [7] as implemented in the SimpleITK python package. We used the default parameters, except for the number of iterations (400 in our case) and the number of resolution levels (3 in our case). We set up these values based on our empirical experience with previous experiments on other brain MRI segmentation approaches.

3 Self-supervised pre-training

To boost the performance of our approach and counter data imbalance we decided to train a Unet architecture with a private dataset comprising 113 subjects with lesion activity visually confirmed by a trained neuroanalyst. The images of this private dataset were acquired in different scanning machines from the 3 main vendors (Siemens, Philips an General Electrics) with a time lapse ranging from 12 months to 5 years and served as a sample of the normal distribution of image variability in a multi-centre clinical scenario.

For this step, a Unet network was implemented using residual blocks, ReLU activations and group normalisation. The number of filters was set to 16, 32, 64, 128, 256 and 512 with normalisation groups of size 4. Finally, to account for differences between the acquisitions, two extra residual blocks were used to reconstruct the final two images after the output of the decoder (one for each timepoint). In order to force the network to learn better image representations for longitudinal studies, we used a single autoencoder architecture for both timepoints to remove image style differences due to different acquisition parameters, noise and other intensity differences. Therefore, we wanted to enforce similar features for stable regions and dissimilar ones for changes in lesions. To train that in an unsupervised manner without labels, we used two different loss functions representing these two constraints: a consistency loss between the features of both images and a reconstruction loss for each image.

The consistency loss was computed as the sum of the mean squared error between the features of each timepoint for all the encoder blocks (including the bottleneck features). Since most of the brain should remain the same (stable disease), maximising the similarity between features should help get rid of noisy predictions due to acquisition differences. The reconstruction loss was computed, once again, using the mean squared error. However, in that case we compared

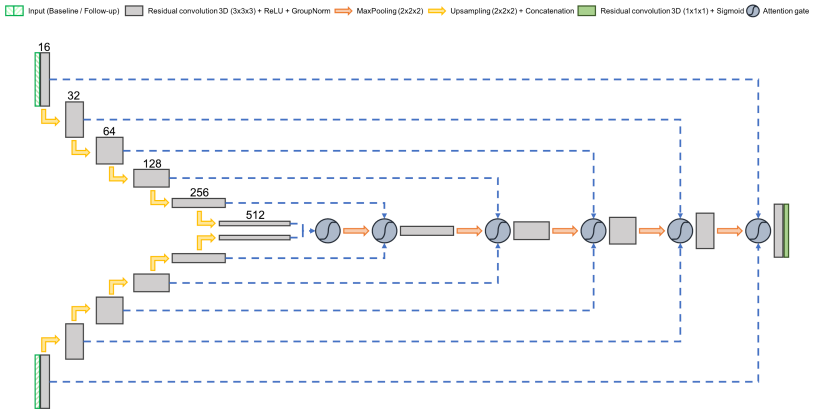

Fig. 1. Scheme of the dual path Unet architecture with its hyperparameters.

the reconstructed images to the original ones. By minimising this loss, we could take true image differences (due to lesion activity) into account.

4 Activity network

Due to its overall performance, we decided to also use a Unet network for the final activity prediction (see Figure 1 for details). However, instead of merging both images at the input level, we kept the pretrained encoder with shared weights in a dual encoding path (similarly to Gessert et al. [2]). Furthermore, in order to fuse the features from each image we implemented a novel attention block loosely based on the attention gate proposal by Schlemper et al. [5] and the self-attention mechanism proposed by Vaswani et al. [8]. Specifically, we used the cross-product between the normalised features of each timepoint (z-scores) to determine the Pearson's correlation coefficient of each pixel. The absolute value of that coefficient was then used as a weighting mechanism to focus the network on pixels with dissimilar features. That weighting mechanism was applied to an embedding of the subtraction between images. By using subtraction and a similarity measure per pixel we wanted to introduce the expert knowledge used in classical subtraction-based approaches to guide the network. Mathematically, if $nx_0 \in \mathbb{R}^{f \times H \times W \times D}$ represents the normalised feature vector (with f features) of the baseline image and $nx_1 \in \mathbb{R}^{f \times H \times W \times D}$ represents the normalised feature vector of the follow-up (both computed using an instance normalisation layer), the output of the novel attention gate can be defined as:

$$AG(nx_0, nx_1) = \alpha \cdot W_v^T \bullet (nx_1 - nx_0), \tag{1}$$

where $\alpha = |nx_0^T \bullet nx_1|$ and represents the feature correlation map between timepoints and W_v represents the weight of a 3D convolution operation. Finally, the decoder of the activity network combines the upsampled features of the

previous level, with the skip connection values coming from the attention gate as proposed on the original attention Unet network [5]. To train the whole process, we used the cross-entropy loss with a balanced sample of patches (50% from regions without activity and 50% with at least a voxel with lesion activity). This sampling was applied randomly at each epoch to crop patches of size $32 \times 32 \times 32$ from a set of overlapping patches (with a window of $16 \times 16 \times 16$) inside the brain region. We limited the repetition of negative samples by forcing the network to sample all of them before starting a new sequence. Due to GPU memory constraints, we used a batch size of 32 during training.

This training process was carried out using a 5-fold cross-validation approach. For inference, the 5 networks are used and the final prediction is obtained from the average of their results. Finally, to reduce any remaining false positives, lesions in the brain boundary and lesions smaller than 4 voxels are removed from the network prediction.

5 Implementation details

The whole framework was implemented using python 3.6.9, the pytorch package (version 1.7.1), the SciPy package (version 1.4.1) and the scikit-image package (version 0.16.2). The model training process was run on a private server with 64 CPU cores, 252 GB of RAM and a Quadro RTX 6000 GPU card.

References

1. Freedman, M.S., Selchen, D., Arnold, D.L., Prat, A., Banwell, B., Yeung, M., Morgenthau, D., Lapierre, Y., Group, C.M.S.W., et al.: Treatment optimization in MS: Canadian MS working group updated recommendations. Canadian J. Neuro. Sci. **40**(3), 307 – 323 (2013)
2. Gessert, N., Bengs, M., Krüger, J., Opfer, R., Ostwaldt, A.C., Manogaran, P., Schippling, S., Schlaefer, A.: 4D deep learning for multiple sclerosis lesion activity segmentation. In: Medical Imaging with Deep Learning (MIDL) (2020)
3. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics **9**(1), 62–66 (1979)
4. Prosperini, L., Mancinelli, C.R., De Giglio, L., De Angelis, F., Barletta, V., Pozzilli, C.: Interferon beta failure predicted by EMA criteria or isolated mri activity in multiple sclerosis. Mult. Scler. J. **20**(5), 566–576 (2014)
5. Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., Rueckert, D.: Attention gated networks: Learning to leverage salient regions in medical images. Med. Image Anal. **53**, 197–207 (2019)
6. Stangel, M., Penner, I.K., Kallmann, B.A., Lukas, C., Kieseier, B.C.: Towards the implementation of 'no evidence of disease activity' in multiple sclerosis treatment: the multiple sclerosis decision model. Ther. Adv. Neurol. Disord. **8**(1), 3 – 13 (2015)
7. Tustison, N., Avants, B., Cook, P., Zheng, Y., Egan, A., Yushkevich, P., Gee, J.: N4ITK: Improved N3 bias correction. IEEE Trans. Med. Imag. **29**(6), 1310 –1320 (2010)
8. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: 31st Conference on Neural Information Processing Systems (NIPS 2017) (2017)

New Multiple Sclerosis Lesion Detection with Convolutional Neural Registration Networks

Julia Andresen[1], Hristina Uzunova[2], Jan Ehrhardt[1], and Heinz Handels[1,2]

[1] Institute of Medical Informatics, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany
j.andresen@uni-luebeck.de
[2] German Research Center for Artificial Intelligence, Lübeck, Germany

Abstract. The delineation of new multiple sclerosis lesions is an important yet difficult and error-prone task. We present an automatic framework for new lesion segmentation composed of two convolutional neural networks (CNNs). The first CNN registers images from different timepoints and simultaneously segments non-corresponding regions. The second CNN refines the non-correspondence maps into segmentations of new lesions. Segmentation performance is shown to profit from the registration results of the first network.

Keywords: Multiple sclerosis · Convolutional neural network · Non-correspondence detection · Image registration.

1 Introduction

Newly appearing lesions are an important biomarker and outcome measure in monitoring multiple sclerosis (MS) indicating non-response to treatment with anti-inflammatory drugs [1]. Their detection and delineation however is a difficult and time-consuming task as many new lesions are rather small and subtle, resulting in high intra- and inter-rater variability [2]. We consider the detection of new lesions in a follow-up image with respect to a baseline image as a non-correspondence detection problem. Consequently, our deep learning-based method consists of two steps: (1) co-registration of baseline and follow-up with simultaneous non-correspondence estimation, and (2) lesion segmentation based on the detected non-correspondences and image information. Step 1 is trained in an unsupervised manner using a newly developed convolutional neural network (CNN), called NCR-Net. The images and the computed non-correspondence map are then used in a classical U-net architecture to generate the lesion segmentations. Our results show that the upstream non-correspondence detection significantly increases the segmentation quality.

2 Method

New MS lesions appearing between examinations show as hyperintense areas on the follow-up image that are not present in the baseline image. The first

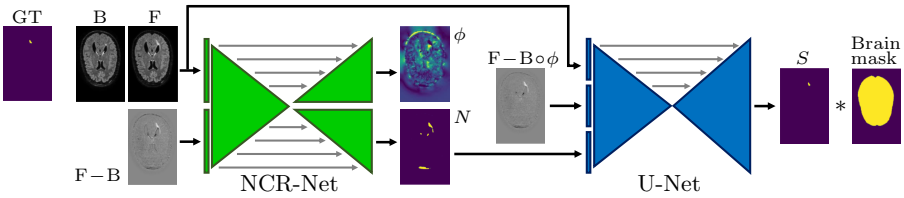

Fig. 1. New lesion segmentation framework. The first CNN registers the baseline onto the followup-image and generates non-correspondence maps. The second network uses the results of the first CNN to segment newly appeared MS lesions.

part of our new lesion segmentation framework is a CNN designed to simultaneously segment such non-corresponding image regions and to register baseline and follow-up images (NCR-Net). NCR-Net is composed of one encoder that takes the baseline B, the follow-up F and their subtraction image as input and two separate decoders that generate a registration deformation field ϕ and a non-correspondence map N. The network is trained unsupervised with a loss function composed of the normalized cross-correlation as a distance measure D and regularization terms for ϕ and N to optimize the network parameters θ:

$$\mathcal{L}(\theta; \mathrm{B}, \mathrm{F}) = \sum_{x \in \Omega} \underbrace{(1-N) \cdot \mathrm{D}\big[\mathrm{F}, \mathrm{B} \circ \phi\big]}_{\substack{\text{Masked image distance} \\ \text{favouring segmentation of} \\ \text{regions with high distances}}} + \overbrace{\alpha \mathrm{R}(\phi)}^{\text{Regularization of deformation}} + \underbrace{\beta N + \gamma \tanh\big(\|\nabla N\|_2\big)}_{\substack{\text{Regularization of} \\ \text{non-correspondence map}}} \quad (1)$$

The image distance D is masked by N and the regularization terms ensure smooth deformations as well as small and contiguous non-correspondence regions. The hyperparameters α, β and γ are empirically chosen.

Note that the network is able to compensate for non-correspondences due to grown or shrunk lesions by deforming the baseline image accordingly. However, distortions for new lesions are not desired and the deformation field is therefore set to zero in the final non-corresponding image regions given by N. New lesions can then be detected as hyperintense regions in the subtraction image $\mathrm{F} - \mathrm{B} \circ \phi$.

The registered subtraction image, the non-correspondence map and the original MRIs are inputs of the second part of our framework. The architecture of the second CNN is equivalent to a 3D U-Net [3], except that the different inputs are first passed to separate convolutional blocks to allow the network to treat inputs differently [4]. The U-Net is trained to generate segmentations S of new lesions based on ground-truth (GT) segmentations minimizing the Dice loss function.

Due to memory constraints, three successive horizontal slices are input to the CNNs rather than the whole 3D volumes. The full resolution is necessary to enable the detection of the mostly small lesions. To compensate for the class imbalance problem inherent in this segmentation problem slices containing new lesions in the GT segmentations are passed twice to the network during training, once with the original orientation and once flipped horizontally. Out of the slices

containing no new lesions only a small predefined percentage (1 % in our experiments) is used for training. To get segmentations for the whole volumes, we iterate through the horizontal slices keeping the segmentation result for the central slice in each iteration step. Both networks are trained for 400 epochs on batches of size 5 with Adam optimization and an initial learning rate of $1e^{-4}$. Finally, in a post-processing step, voxel segmentations outside the brain are eliminated using brain masks generated with the default pre-processing provided with the MSSEG-2 challenge [5, 6]. The whole pipeline is depicted in Figure 1.

3 Results

Five-fold cross-validation is performed for the results reported in this section splitting the data into training (32 patients) and test (8 patients) datasets. Dice score, mean surface distance and lesion sensitivity (proportion of detected GT lesions [7]) are reported for the images containing new lesions in the GT to evaluate lesion segmentation and detection performance. For the images containing no new lesions, the number and volume of erroneously detected lesions are considered. The results are averaged over all patients.

For comparison a U-Net without preceding registration is trained using the original MRIs and their difference image as input. We further report results for using the non-correspondence maps given by NCR-Net directly as new lesion segmentations. Although the performance drops significantly compared to the U-Net-based segmentations, this method has the advantage of being fully unsupervised. Furthermore, NCR-Net is trained in a supervised manner using the GT segmentations and adding the Dice loss to the loss function (1). Results show that the segmentation performance profits from the registration results of NCR-Net that amplify hyperintense areas in the subtraction images. Results are further improved passing the non-correspondence maps to the U-Net leading to significantly higher Dice scores. For challenge submission we use the last mentioned pipeline retraining NCR- and U-Net with all 40 training images.

Table 1. Results of the different CNNs considered for new MS lesion segmentation. In brackets the inputs passed to the U-Net are given. Dice Score (DSC), mean surface distance (MSD) and lesion sensitivity are reported for images containing new lesions in the ground truth (GT) while number and volume of erroneously segmented lesions are shown for the remaining images. Best results are presented in bold font and results that significantly outperform the U-Net-only approach are marked with *.

	Lesion(s) in GT			No lesion in GT	
Method	DSC	MSD	Sens.	Nr.	Vol.
U-Net(B, F, F$-$B)	0.4293	6.4152	0.7100	**1.9091**	14.8683
NCR-Net (unsupervised)	0.1420	16.7017	0.4910	45.3634	1851.8707
NCR-Net (supervised)	0.3289	10.7011	**0.7336**	11.7273	522.6306
NCR- and U-Net(B, F, F$-$B $\circ\, \phi$)	0.4317	9.2413	0.6822	2.8182	**13.6250**
NCR- and U-Net(B, F, F$-$B $\circ\, \phi$, N)	**0.4508***	**5.7948**	0.7177	2.3636	20.0467

References

1. Moraal, B., van den Elskamp, I. J., Knol, D. Let al.: "Long-Interval T2-Weighted Subtraction Magnetic Resonance Imaging: A Powerful New Outcome Measure in Multiple Sclerosis Trials". In: Annals of neurology, vol. 67(5), pp. 667–675 (2010)
2. Elliott, C., Arnold, D. L., Collins, L., Arbel, T.: "Temporally Consistent Probabilistic Detection of New Multiple Sclerosis Lesions in Brain MRI". In: IEEE Transactions on Medical Imaging, vol. 32(8), pp. 1490–1503 (2013)
3. Ronneberger, O., Fischer, P. Brox, T.: "U-net: Convolutional Networks for Biomedical Image Segmentation. In: International Conference on Medical image computing and computer-assisted intervention", pp. 234–241, Springer (2015)
4. Hering, A., Kuckertz, S., Heldmann, S., Heinrich, M. P.: "Enhancing Label-Driven Deep Deformable Image Registration with Local Distance Metrics for State-of-the-Art Cardiac Motion Tracking". In: Bildverarbeitung für die Medizin 2019, pp. 309–314, Springer (2019)
5. Manjón, J. V., Coupé, P. "volBrain: An Online MRI Brain Volumetry System". In: Frontiers in Neuroinformatics, vol. 10(30) (2016)
6. Masson, A.: "MICCAI 2021 - Longitudinal Multiple Sclerosis Lesion Segmentation Challenge". https://github.com/Inria-Empenn/lesion-segmentation-challenge-miccai21 (2021)
7. Commowick, O., Istace, A., Kain, M., et al.: "Objective Evaluation of Multiple Sclerosis Lesion Segmentation using a Data Managementand Processing Infrastructure". In: Scientific Reports, vol. 8(1):13650 (2018)

MSDetector: A fully convolutional neural network for the detection of new T2-w lesion in multiple sclerosis

Mostafa Salem[1,2][0000−0001−9915−8390], Arnau Oliver[1][0000−0002−0115−0647], Joaquim Salvi[1][0000−0002−9482−7126], and Xavier Lladó[1][0000−0003−2777−3479]

[1] Research Institute of Computer Vision and Robotics, University of Girona, Spain.
[2] Computer Science Department, Faculty of Computers and Information, Assiut University, Egypt.

Abstract. Longitudinal magnetic resonance imaging (MRI) has an important role in multiple sclerosis (MS) diagnosis and follow-up. Specifically, the presence of new T2-w lesions on brain MRI scans is considered a predictive biomarker for the disease. The MSSEG-2: MS new lesions segmentation challenge using a data management and processing infrastructure is a new MICCAI sponsored online challenge. The challenge is performing the evaluation task on a large database (100 patients, each with two time points) compiled from the OFSEP cohort, the French MS registry, with 3D FLAIR images from different centers and scanners. This paper describes the method of the VICOROB team at the university of Girona, Spain (MSDetector). A fully convolutional neural network (FCNN) that is used to perform the detection of new T2-w lesions in longitudinal brain MRI images.

1 Introduction

Multiple sclerosis (MS) is an inflammatory disease of the central nervous system, which is characterized by the presence of lesions in the brain and the spinal cord. Magnetic resonance imaging (MRI) has become a core paraclinical tool for diagnosing and predicting long-term disability and treatment response in MS patients. Follow-up brain MRI is required in patients who have not been diagnosed yet as MS patients but they show clinical and radiological findings suggestive to MS [1]. In this paper, we are extending our fully CNN approach to detect new T2-w lesions in longitudinal brain MRI images [2]. The proposed model combines intensity-based and deformation-based features within an end-to-end deep learning approach. The model is trained end-to-end to learn simultaneously the deformation field (DF) and the new T2-w lesions using a combined loss function.

2 Materials and Methods

2.1 Network architecture

Figure 1 shows the new T2-w MS lesion segmentation architecture. We have modified our previous T2-w MS lesion segmentation architecture [2] to the dataset

provided by the challenge. The proposed network is a fully CNN that takes the FLAIR image modality in both baseline and follow-up as inputs and outputs the new T2-w lesion segmentation. The network consists of two parts. The first part is a U-Net block that automatically learns the DF that non-linearly register the FLAIR baseline image to the follow-up space. The learned DF and the baseline and follow-up images are then fed to a second part of the network, another U-Net that performs the final detection and segments the new T2-w lesions. The network is trained end-to-end with gradient descent and simultaneously learns both DF and new T2-w lesion segments.

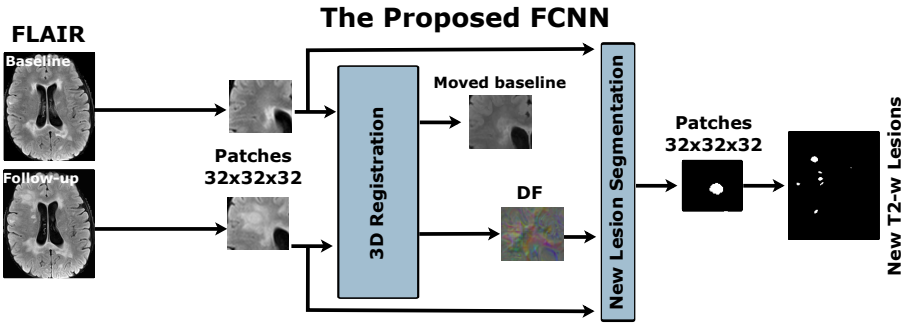

Fig. 1. Scheme of the new T2-w MS lesion segmentation network.

3D registration architecture: A 3D registration block is built for the FLAIR modality following the architecture explained in [2]. This block is inspired by the work of (VoxelMorph), which is a learning framework for deformable medical image registration [3]. The registration block learns the DF, which nonlinearly registers the FLAIR baseline image to the follow-up space.

3D segmentation architecture: A 3D segmentation CNN is also used for segmenting the new T2-w lesions. It is a two-branch network where each branch is a U-Net following the architecture explained in [2]. The inputs of the first branch are the FLAIR image modality in both baseline and follow-up, while the second branch input is the DF learned from the first registration blocks. The outputs of the two branches are concatenated before the classification step.

2.2 Loss functions

The loss function is the summation of two loss functions. One function is an unsupervised loss function that controls the registration part of the network [3]. It consists of two components: a similarity part that penalizes differences in appearance between the moved baseline and follow-up images and a regularization part that enforces a spatially smooth deformation and is often modeled as a linear operator on the spatial gradients of DF as stated in [3]. Therefore,

the registration block is trained in an unsupervised manner using the spatial transform block which is used to warp the baseline image to the follow-up image using the learned DF. The block learns the deformation field by minimizing the MSE between the warped baseline and the follow-up image during training. The other function is a supervised loss function $L_{CrossEntropy}$ (CrossEntropy) that controls the segmentation part of the network and penalizes differences between the segmentation and GT. The loss function L_{Total} is as follows:

$$L_{Total} = \underbrace{L_{CrossEntropy}(Seg, GT)}_{Segmentation\ loss\ function} + \underbrace{\left(\overbrace{\frac{1}{N}\sum_{i=1}^{N}(F_i - B(DF)_i)^2}^{Similarity\ part} + \overbrace{\lambda \sum_{p \in DF} \parallel \bigtriangledown DF(p) \parallel^2}^{Regularization\ part} \right)}_{Registration\ loss\ function}$$

where F, $B(DF)$, and DF are follow-up image, baseline image warped by DF (moved baseline), and DF for the FLAIR modality, respectively. Seg and GT are the automatic segmentation and the GT, respectively. N is the number of voxels in a patch and λ is a regularization parameter.

2.3 Datasets and preprocessing

The database used in this paper is the MSSEG-2 challenge dataset. A total of 100 MS patients were gathered. Only a 3D FLAIR sequence at a first timepoint and a 3D FLAIR sequence at a second timepoint (1 to 3 years after the first one) is available. A total of 15 different MRI scanners are represented (three GE scanners, six Philips scanners, and six Siemens scanners). The dataset is divided into 40 patients for training and 60 patients for testing.

For each patient, the same preprocessing steps were performed on both baseline and follow-up images. First, a brain mask was identified and delineated using the ROBEX Tool [4]. Second, the FLAIR images underwent a bias field correction step using the N4 algorithm from the ITK library. Finally, the baseline and the follow-up intensity values from all the training set were normalized using a histogram matching approach based on [5].

2.4 Training and implementation details

For training the network, 3D 32x32x32 patches with a step size of 16x16x16 were extracted from the basal and follow-up images of the FLAIR input modality. The extracted patches were divided into training and validation sets (70% for training and 30% for validation). The model was trained using Adam [6] with default parameters and regularization parameter $\lambda = 0.01$. The extracted patches were passed to the network for training in minibatches of size 4, and the network was set to train for 30 epochs. To prevent overfitting, the training process was automatically terminated when the validation accuracy did not increase after 5 epochs.

References

1. A. Rovira, M. P. Wattjes, M. Tintore, C. Tur, T. A. Yousry, M. P. Sormani, N. De Stefano, M. Filippi, C. Auger, M. A. Rocca, *et al.*, "Magnims consensus guidelines on the use of mri in multiple sclerosis-clinical implementation in the diagnostic process (vol 11, pg 471, 2015)," *NATURE REVIEWS NEUROLOGY*, vol. 11, no. 8, 2015.
2. M. Salem, S. Valverde, M. Cabezas, D. Pareto, A. Oliver, J. Salvi, l. Rovira, and X. Lladó, "A fully convolutional neural network for new t2-w lesion detection in multiple sclerosis," *NeuroImage: Clinical*, vol. 25, p. 102149, 2020.
3. G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "Voxelmorph: A learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2019.
4. J. E. Iglesias, C.-Y. Liu, P. M. Thompson, and Z. Tu, "Robust brain extraction across datasets and comparison with publicly available methods," *IEEE Transactions on Medical Imaging*, vol. 30, no. 9, pp. 1617–1634, 2011.
5. 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.
6. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.