Computerized Medical Imaging and Graphics 103 (2023) 102157

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Minimizing the effect of white matter lesions on deep learning based tissue segmentation for brain volumetry

Albert Clèrigues [a,*], Sergi Valverde [b], Joaquim Salvi [a], Arnau Oliver [a], Xavier Lladó [a]

[a] Institute of Computer Vision and Robotics, University of Girona, Spain
[b] Tensor Medical, Girona, Spain

ARTICLE INFO

Keywords:
Magnetic resonance imaging
Brain tissue segmentation
White matter lesions
Lesion inpainting
Deep learning
Brain volumetry

ABSTRACT

Automated methods for segmentation-based brain volumetry may be confounded by the presence of white matter (WM) lesions, which introduce abnormal intensities that can alter the classification of not only neighboring but also distant brain tissue. These lesions are common in pathologies where brain volumetry is also an important prognostic marker, such as in multiple sclerosis (MS), and thus reducing their effects is critical for improving volumetric accuracy and reliability. In this work, we analyze the effect of WM lesions on deep learning based brain tissue segmentation methods for brain volumetry and introduce techniques to reduce the error these lesions produce on the measured volumes. We propose a 3D patch-based deep learning framework for brain tissue segmentation which is trained on the outputs of a reference classical method. To deal more robustly with pathological cases having WM lesions, we use a combination of small patches and a percentile-based input normalization. To minimize the effect of WM lesions, we also propose a multi-task double U-Net architecture performing end-to-end inpainting and segmentation, along with a training data generation procedure. In the evaluation, we first analyze the error introduced by artificial WM lesions on our framework as well as in the reference segmentation method without the use of lesion inpainting techniques. To the best of our knowledge, this is the first analysis of WM lesion effect on a deep learning based tissue segmentation approach for brain volumetry. The proposed framework shows a significantly smaller and more localized error introduced by WM lesions than the reference segmentation method, that displays much larger global differences. We also evaluated the proposed lesion effect minimization technique by comparing the measured volumes before and after introducing artificial WM lesions to healthy images. The proposed approach performing end-to-end inpainting and segmentation effectively reduces the error introduced by small and large WM lesions in the resulting volumetry, obtaining absolute volume differences of 0.01 ± 0.03% for GM and 0.02 ± 0.04% for WM. Increasing the accuracy and reliability of automated brain volumetry methods will reduce the sample size needed to establish meaningful correlations in clinical studies and allow its use in individualized assessments as a diagnostic and prognostic marker for neurodegenerative pathologies.

1. Introduction

Global and regional volumetry of the brain parenchyma is a promising biomarker that can improve prognosis for multiple sclerosis (MS) patients (Bendfeldt et al., 2009; Lansley et al., 2013; Pérez-Miralles et al., 2013). Brain volume loss has been shown to be a predictor of disease progression and disability status in MS patients (Di Filippo et al., 2010; Ghione et al., 2020). Moreover, the rate of brain volume loss is also used to evaluate the effectiveness of disease-modifying treatments in clinical studies as well as for individualized treatment response assessment (Sotirchos et al., 2020; Cortese et al., 2022). Magnetic resonance (MR) imaging offers a noninvasive way to perform indirect volume measurements on the brain parenchyma and its distinct cerebrospinal fluid (CSF), gray matter (GM) and white matter (WM) components. In non-uniformity corrected T1-w MR images, these tissues are characterized by normally distributed intensity profiles with different means and variances. However, a characteristic of brain scans from MS patients is the presence of WM lesions appearing as a fourth intensity distribution that intersects with the brain tissue intensities to be measured. The presence of WM lesions can bias the characterization of

* Correspondence to: University of Girona, Ed. P-IV, Campus Montilivi, 17003 Girona, Spain.
E-mail address: albert.clerigues@udg.edu (A. Clèrigues).

https://doi.org/10.1016/j.compmedimag.2022.102157
Received 27 May 2022; Received in revised form 2 December 2022; Accepted 2 December 2022
Available online 13 December 2022

normal-appearing tissue intensities and interfere with brain tissue quantification methods (González-Villà et al., 2017). The error that WM lesions introduce is highly dependent on their aspect and size (Battaglini et al., 2012), which change over time, introducing varying levels of error in images taken at different timepoints. These lesions can especially affect the estimation of partial volumes found in the interfaces between brain tissues and have been observed to produce a boundary shifting effect (Magon et al., 2014). Reducing the error introduced by WM lesions in brain tissue segmentation is critical for improving the reliability and accuracy of cross-sectional and longitudinal brain volumetry methods.

Techniques that minimize the effect of WM lesions in brain tissue segmentation usually involve lesion inpainting as a preliminary step before segmentation. These techniques fill the lesioned voxels with intensities resembling the normal-appearing WM (NAWM) of that image. Chard et al. (2010) proposed the use of a Gaussian mixture model to characterize and sample the intensity distribution of NAWM in the whole image to fill lesioned voxels while also emulating scanner noise and nonuniformity. Battaglini et al. (2012) and Magon et al. (2014) both proposed similar local inpainting methods for preliminary brain tissue segmentation to then fill the lesion voxels with intensities similar to those of the NAWM adjacent to the lesioned voxels. Similarly, Valverde et al. (2014) also used preliminary tissue segmentation to characterize and sample the NAWM intensity distribution but did so on a slice-by-slice basis. Prados et al. (2016) proposed a non-local mean patch-based inpainting method that can work with longitudinal data and for any MR modality. More recently, data-driven methods using deep learning have been proposed based on the use of convolutional neural networks (CNNs) for lesion inpainting. Armanious et al. (2019) used a 2D conditional generative adversarial network (cGAN) to synthesize realistic looking intensities for a square patch removed from the input slice. Xiong et al. (2020) used 2D U-Net with a nonlesion attention module to inpaint lesioned voxels, while Zhang et al. (2020) proposed the use of 2D U-Net with edge priors as additional input to improve the inpainting quality. Manjón et al. (2020) proposed a 3D blind inpainting method that automatically inpaints any abnormal-looking voxels without requiring prior lesion segmentation, unlike the methods described previously that require a preliminary WM lesion mask. Tang et al. (2021) proposed an inpainting approach for MS lesions using dynamic learnable gate masks to improve the morphological and textural consistency of inpainted regions and reduce their effect on subsequent brain tissue segmentation. Most works cited above have been shown to improve the results for brain tissue segmentation methods by reducing segmentation differences between healthy and artificially lesioned image pairs. However, to the best of our knowledge, the recent deep learning-based brain tissue segmentation approaches (Rajchl et al., 2018; Guha Roy et al., 2019; Henschel et al., 2020) have not evaluated the effect of WM lesions.

In this work, we propose a 3D patch-based deep learning tissue segmentation framework for brain volumetry which learns from the outputs of a reference classical brain tissue segmentation method. In our approach, we improve the robustness on pathological cases having WM lesions by using small patches and a percentile-based input normalization. To further minimize the effect of WM lesions, we also propose the use of a multi-task double U-Net architecture performing end-to-end inpainting and segmentation. To train the proposed method as well as to evaluate the WM lesion effect, we use pairs of lesioned and non-lesioned versions of the same brain image. Since these pairs of images cannot be naturally obtained, artificial lesions are introduced into a set of scans from healthy subjects to obtain both versions of the same image. Our goal is to learn a segmentation model that can minimize the effect of WM lesions on the rest of the normal-appearing tissue in the image. During training, we use the artificially lesioned brain images as input and target the brain tissue probabilities of their originally healthy counterpart image as output. In this way, the system is trained to minimize the impact of WM lesion voxels on the segmentation of neighboring healthy tissue. In the proposed method, a preliminary WM lesion mask is used to occlude the lesioned voxels of the input patch by masking it with zeros. Then, a double chained U-Net architecture is used, where the first network inpaints the occluded lesion voxels and the second performs brain tissue segmentation from the inpainted patch. Both networks are trained end-to-end so that the inpainter network is also trained to aid in the segmentation task.

We evaluate the effect of WM lesions on our deep learning framework as well as on FAST (Zhang et al., 2001), the brain tissue segmentation method used to generate the training targets, which is implemented in the FSL package of analysis tools for structural MR brain imaging data. In the evaluation, we quantify the tissue volume differences between healthy and artificially lesioned versions of the same image for each of the considered tissue segmentation methods. Without performing lesion inpainting, our deep learning framework already shows significantly smaller and more localized volume differences due to the presence of WM lesions than the reference method. We then evaluate the extent to which the lesion effect minimization techniques reduce the error introduced on the measured tissue volumes. The FSL package also provides a WM lesion inpainting method (Battaglini et al., 2012), which is typically used along with FAST (Zhang et al., 2001). The FSL pipeline doing WM lesion inpainting and brain tissue segmentation is used as a baseline to compare against our deep learning approach. Additionally, we also compare against the case where we first inpaint the WM lesions with the FSL method and then perform the brain tissue segmentation with our deep learning approach. The proposed method doing end-to-end inpainting and tissue segmentation is faster and obtains significantly lower volume differences, especially when considering larger WM lesions. Even when the FSL_inpainting method is used to preprocess the image, our deep learning based tissue segmentation model still achieves significantly lower error and better performance on large WM lesions than the FSL pipeline. Thanks to the use of data-driven techniques, we are able to learn from a reference method while minimizing the WM lesion effect on the measured tissue volumes to almost negligible levels. The development framework is available to the research community at https://github.com/NIC-VICOROB/LITS.

2. Materials

Two different kinds of image datasets are used to train and evaluate the proposed method, healthy brain scans and lesioned brain scans with manually delineated WM lesion masks comprising small and large lesions from patients with multiple sclerosis (MS) and other pathologies. These brain images are used to generate artificially lesioned and healthy image pairs for training and evaluation. The location and morphology of artificial WM lesions introduced in the T1-w healthy images are taken from the WM lesion masks of lesioned brain scans, while their appearance is simulated by sampling intensities between the means of GM and WM tissue, similar to the work of Battaglini et al. (2012).

2.1. Healthy brain dataset

Calgary-Campinas Public Brain MR Dataset (Souza et al., 2018). This dataset is composed of 359 T1-weighted brain scans from 359 healthy adults with an average age of 53.5 $\pm$ 7.8 years, ranging between 29 and 80 years. Images were acquired on scanners from three vendors (GE, Philips, and Siemens) at two different magnetic field strengths of 1.5 T and 3 T, approximately 60 scans were obtained per vendor. Most scans in this dataset have a voxel size of $1.0 \times 1.0 \times 1.0$ mm^3 except for sixty scans acquired at $0.89 \times 0.89 \times 0.89$ mm^3 and another sixty acquired at $1.33 \times 1.0 \times 1.0$ mm^3. The dataset also includes silver standard brain masks generated through a consensus of several state-of-the-art automatic skull stripping methods.

2.2. Lesioned brain datasets

MSSEG Challenge (Commowick et al., 2018). The MSSEG Challenge hosted at the MICCAI 2016 international conference provided a multicentric database for training consisting of 15 multimodal (T1-w, T1-w gadolinium, T2-w, FLAIR and PD) MR images obtained from MS patients with an average lesion load of 20.8 ± 19.9 ml. Images were acquired on three different scanners at different voxel sizes: five images from a Philips Ingenia 3 T scanner at $0.7 \times 0.74 \times 0.74$ mm^3, five images from a Siemens Verio 3 T scanner at $1.1 \times 0.5 \times 0.5$ mm^3 and the remaining five images from a Siemens Aera 1.5 T scanner at $1.25 \times 1.03 \times 1.03$ mm^3. The MR images were rigidly coregistered to the FLAIR scan, which was manually annotated by 7 independent experts, and a consensus gold standard WM lesion segmentation approach was built.

ISBI 2015 Longitudinal MS Lesion Segmentation Challenge (Carass et al., 2017). This challenge provided a multimodal (T1-w, T2-w, FLAIR and PD) training dataset with 21 longitudinal scans from five MS patients with an average lesion load of 11.6 ± 10.5 ml. Images were acquired on a 3 T MRI Philips scanner with a voxel size of $0.82 \times 0.82 \times 1.17$ mm^3. Manual delineations were made by two experts identifying and segmenting white matter lesions on the MR images. The MR images from each subject as well as the expert WM lesion delineations were rigidly coregistered to the T1-w scan.

WMH Challenge 2017 (Kuijf et al., 2019). The training set provided 60 sets of brain MR images (3D T1 and 2D multislice FLAIR) from 60 subjects of two memory clinics showing cognitive impairment of presumed vascular origin with an average lesion load of 17.5 ± 17.1 ml. Images were taken with five different 3 T MR scanners from three different vendors (Siemens, Philips and GE) with voxel sizes of $1.0 \times 1.0 \times 1.0$ mm^3 and $0.94 \times 0.94 \times 1.0$ mm^3. The FLAIR scans from each subject were resampled and coregistered to the 3D T1 scan via an affine transform. The provided gold standard was made with manual annotations of white matter hyperintensities (WMHs) made by experts in accordance with the STandards for ReportIng Vascular changes on nEuroimaging (STRIVE) criteria (Wardlaw et al., 2013).

2.3. Preprocessing

In the image preprocessing stage, we generate the healthy and lesioned image pairs that are used for training and evaluation. The location and morphology of artificial WM lesions are obtained from WM lesion masks of the three lesioned brain datasets that are registered to the healthy dataset scans. In practice, all the available WM lesion masks from lesioned datasets are registered to each of the T1-w healthy images, allowing the generation of several artificially lesioned scans from a single healthy scan. The registered WM masks are then used to generate artificial lesions in the healthy T1-w brain scans with the intensities located within the GM/WM interface. The preprocessing steps are explained in detail in the following sections.

2.3.1. Skull-stripping

The healthy scans belonging to the Calgary–Campinas dataset images need to be skull-stripped before segmenting with FAST to consider only the intensities corresponding to the intracraneal cavity. For this, we use the provided silver brain masks, which are applied to generate the skull-stripped images. For the lesioned brain datasets, two of them (MICCAI 2016 MS lesion segmentation challenge and ISBI 2015 Longitudinal MS Lesion Segmentation Challenge) were already skull-stripped, while the WMH Challenge 2017 dataset is processed using ROBEX (Iglesias et al., 2011) on the T1-w images.

2.3.2. Lesion mask registration

In this step, all the available lesioned scans are linearly registered to each of the healthy images, obtaining several artificial WM lesion mask instances in the space of each healthy scan. This process is performed independently for the training and evaluation image sets. To avoid performing a large number of registrations, we first register all the healthy and lesioned images to a common space and then combine these transforms to obtain the desired transforms. Linear affine registration is performed with the skull-stripped T1-w images from both healthy and pathological datasets to the MNI ICBM 152 nonlinear 6th Generation Symmetric Average Brain template using FSL FLIRT (Jenkinson and Smith, 2001; Jenkinson et al., 2002) with default parameters. This results in a linear transform matrix T(*I,MNI*) for each image *I*, which can also be inverted to obtain T(*MNI,I*). Then, for any pair of healthy *H* and lesioned *L* images, we can compute T(*L,H*) using the previously computed transforms to the MNI as follows:

$$\mathrm{T}(L,H) = \mathrm{T}(L,MNI) \circ \mathrm{T}(MNI,H) \quad (1)$$

T(*L,H*) is computed for each lesioned and healthy control image pair and then applied to the binary WM lesion mask using nearest neighbor interpolation. Finally, we ensure that the registered lesions are introduced only to the WM of healthy images. For this, we use FAST (Zhang et al., 2001) to obtain a binary WM mask for each healthy image and keep only the voxels from registered lesion masks that are also classified as WM in the healthy image.

2.3.3. Artificial WM lesions

The registered WM lesion masks are then used to generate several artificially lesioned images from each healthy image. The artificial lesion intensities are filled as in the work of Battaglini et al (Battaglini et al., 2012)., which presented and evaluated the lesion inpainting method we use as a baseline. In their approach, a preliminary FAST (Zhang et al., 2001) tissue segmentation is used to estimate the mean intensities of GM and WM and is then used to generate the intensity distribution for artificial lesions. These are then filled with intensities between the normally appearing GM and WM, with a mean equal to the average of the GM and WM means and a standard deviation equal to a fourth of the interval between the GM and WM means (Battaglini et al., 2012).

During training and inference of the proposed methodology, the artificial lesion intensities are effectively ignored as they are occluded by filling them with zeros. Hence, the intensities of artificial lesions are only useful for evaluating the WM lesion effect of tissue segmentation methods when no inpainting is used.

3. Methods

The proposed deep learning brain tissue segmentation framework consists of a 3D patch-based approach which learns from the outputs of FAST (Zhang et al., 2001), an automatic brain tissue segmentation method implemented in the FSL package. The backbone of our framework consists of a 3D network, depicted in Fig. 1, which is derived from the U-Net architecture (Ronneberger et al., 2015) and uses residual convolution blocks and skip connections. The convolutional layers use $3 \times 3 \times 3$ kernels and are always preceded, except for the input and output nodes, by a batch normalization (BN) layer (Ioffe and Szegedy, 2015) and a parametric rectified linear unit (PReLu) activation (Nair and Hinton, 2010). The parameter distribution of the model is asymmetrical with respect to the residual blocks of the encoder using two convolutional layers, while a single layer is used in the decoder. The network has 4 resolution levels where the feature maps are downsampled by $2 \times 2 \times 2$ in each level of the encoder and upsampled by the same factor in the decoder. Downsampling is performed by concatenating the result of a max pooling operation and strided convolution as proposed by Szegedy et al (Szegedy et al., 2016)., while upsampling is performed with a transposed convolution that learns the upsampling operator for each feature map.

Within our patch-based deep learning framework, the introduction of a WM lesion in a healthy brain scan produces segmentation differences at both global and local levels. Global differences appear when the modification of a small part of the input has an effect on the output

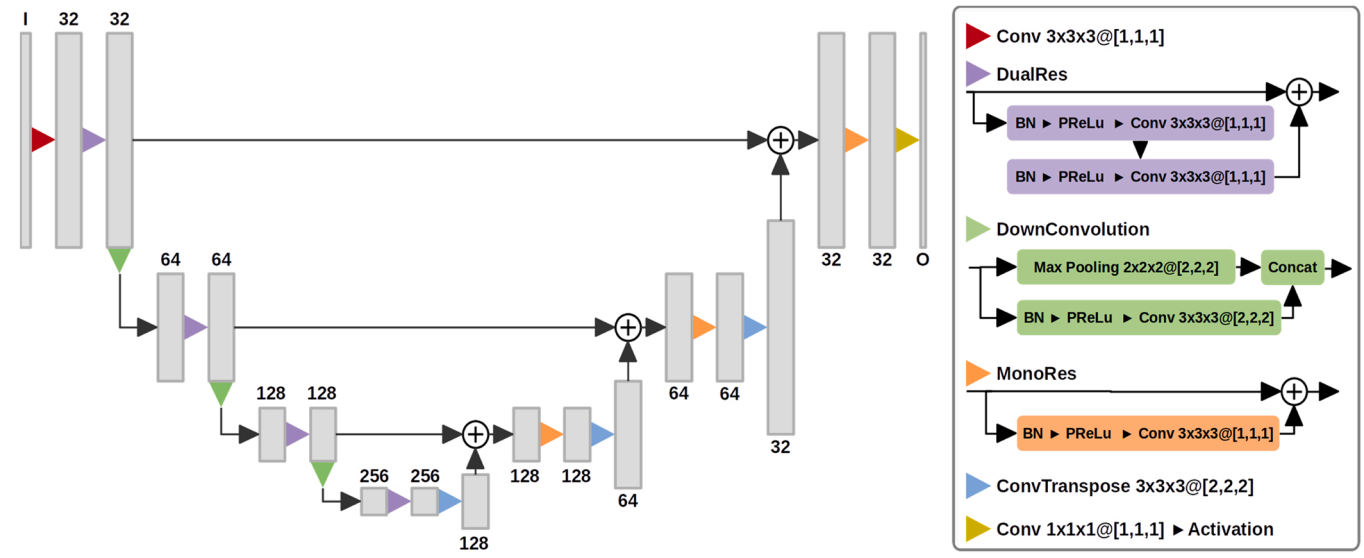


Fig. 1. Diagram of the U-Net derived model used as the backbone of our deep learning brain tissue segmentation framework. The network consists of a 3D U-Net model using residual convolution blocks and skip connections. The parameter distribution is asymmetrical, with the residual blocks of the encoder using two convolutional blocks while a single block is used in the decoder. In the convolutional layers (Conv), K_x K_y K_z@[S_x,S_y,S_z] indicates the kernel and stride dimensions in each axis. The gray boxes represent the feature maps with the number of channels indicated above or under it. The numbers of input and output feature maps are denoted I and O, respectively.

segmentation of the whole image. On the other hand, local differences are those where only the altered region and its neighborhood are affected. The main source of global segmentation differences in the proposed method is caused by input normalization which is applied to the T1-w scan. Input normalization is a technique used to homogenize the range and statistics of neural network inputs so that the variation is reduced and the model can be more finely tuned to the expected input values. Since our aim is to correctly segment the healthy tissue regardless of any intensity changes caused by the development and evolution of WM lesions, we want an input normalization procedure that is invariant to these appearance changes. Due to the combinatorial nature of neural networks, small perturbations in the input values can cause large output differences; thus, a small shift in the normalization parameters could have a measurable effect on the segmented tissue volumes. To minimize this, we propose the use of a minmax input normalization operation for T1-w MR images that maps intensities between the 0.05% and 99.95% percentiles to the $[-1,1]$ interval and then clamps to that same interval to clip any outliers within the desired range. This kind of normalization has much less variability between the healthy and artificially lesioned images than other tested techniques, such as z score normalization (zero mean and unit standard deviation) or intensity rescaling to the 0–1 range. Local segmentation differences due to the appearance of WM lesions are introduced not only to the lesioned voxels and their neighborhood but also to the whole patch where a lesioned voxel appears. Within the proposed patch-based approach, the introduction of artificial lesions in part of a patch will affect the output probabilities of the whole patch. During inference, the input image is sliced into patches to be segmented and then recombined for whole image segmentation. A larger patch size means that a larger proportion of patches contain lesioned voxels which introduce segmentation differences further from the lesioned voxels. Consequently, patch size is an important parameter for mitigating the local effect of WM lesions in patch-based brain tissue segmentation. To select the patch size, we empirically tested 5 isotropic patch sizes between $8 \times 8 \times 8$ and $40 \times 40 \times 40$. The best compromise between tissue segmentation performance and reducing the aforementioned differences is achieved by using a patch size of $16 \times 16 \times 16$.

To minimize the WM lesion effect within the proposed deep learning segmentation framework, we propose a multi-task double U-Net architecture, depicted in Fig. 2, where the first network performs inpainting and the second network segments the brain tissues. The aim is to obtain a segmentation model that can minimize the effect that a WM lesion has on its healthy neighborhood so that it can be correctly segmented despite the adjacent abnormal intensities. The proposed method takes a skull-stripped brain scan along with its binary WM lesion segmentation and outputs a probability distribution of brain tissue (CSF, GM and WM) for each input voxel. The lesioned area is occluded with zero-valued voxels before input to the network. First, the inpainter network inpaints any occluded lesion voxels in the input patch and tries to reconstruct the originally healthy intensities. The inpainted patch is then masked before tissue segmentation, keeping only the inpainted voxels from the first network and taking the original intensities for the rest of nonlesioned voxels. Finally, the second U-Net performs brain tissue segmentation from the inpainted masked patch and outputs a brain tissue probability distribution for each input voxel. During training, we input artificially lesioned images and target the tissue segmentation of the originally healthy image as output both networks are trained simultaneously in an end-to-end manner to allow the segmentation loss gradients from the second model to also backpropagate through the inpainter. This regularizes the inpainter toward inpainting in a way that should also help the tissue segmenter to more accurately approximate the healthy tissue probabilities. In this way, the goal of the inpainter is not to faithfully and accurately approximate the healthy T1 intensities, but rather, we want the tissue segmentation model to better approximate the healthy tissue probabilities regardless of any occluded zero-valued regions.

In the double chained U-Net configuration, the input of the first U-Net is a T1-w patch with WM lesions occluded and the binary WM lesion mask. The output of the inpainter is activated by a hyperbolic tangent function (tanh) to map the range of output intensities within the same $[-1,1]$ interval of input normalization. The input of the second network, the segmenter U-Net, is an inpainted T1-w patch and its output is activated using the Softmax function to obtain a tissue probability distribution for each input voxel.

3.1. Training

The double U-Net system is trained end-to-end using both the healthy

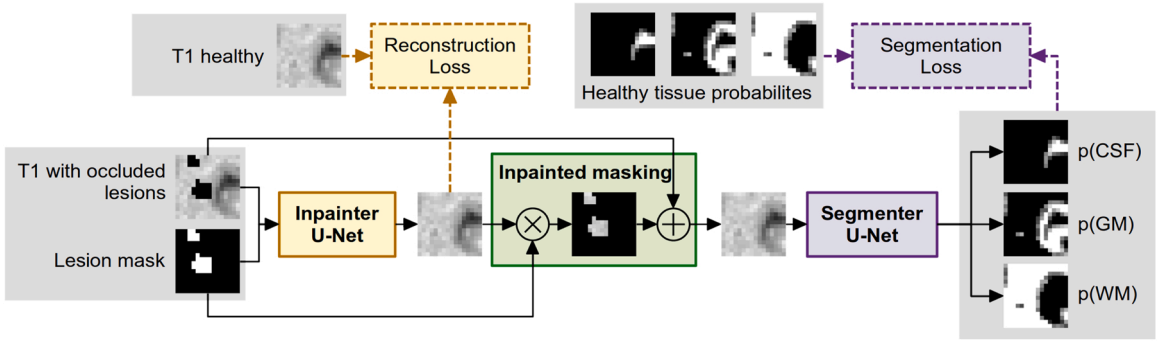


Fig. 2. Overview of the proposed patch-based double chained 3D U-Net architecture performing end-to-end inpainting and brain tissue segmentation. The binary WM lesion mask is used to occlude the lesion from the input patch with zero-valued voxels and is also input to the inpainter. Inpainted masking takes the inpainter output only for lesioned voxels and uses the original intensities for the rest of nonlesioned voxels. The tissue segmenter receives the inpainted masked patch and outputs a probability distribution among the background, CSF, GM and WM classes.

and artificially lesioned images as inputs, targeting the healthy image tissue probabilities as output in both cases. For the artificially lesioned images, the parameters for input normalization are computed only from the nonlesioned brain voxels. To train the proposed patch-based method, we first generate patch training and validation sets. From the training image set, we use 90% of the subjects to build the training patch set and the remaining 10% for the validation patch set. In total, we extract 1 million patches, 900,000 for training and 100,000 for validation. These patches are extracted centered on a set of voxels sampled using a deliberate strategy to balance the representation of segmentation classes as well as the representation of patches with and without lesions half of the patches are extracted evenly from the healthy images without artificial lesions. For each image, the patch centers are sampled using a preliminary FAST (Zhang et al., 2001) tissue segmentation as a guide to obtain 10% centered on the background class and 30% each from the CSF, GM and WM classes. The other half are taken evenly from all the available artificially lesioned images, centered on occluded artificially lesioned voxels. A random 3D offset of up to half the patch size is applied to the healthy and lesion sampled centers to increase the representation of boundaries. The model is then trained end-to-end using the Adadelta optimizer (Zeiler, 2012) with a learning rate of 0.2 and a batch size of 32 patches. To prevent overfitting, early stopping is performed when the loss on the validation set does not improve for 8 consecutive epochs. The loss function used for training and validation is composed of a reconstruction loss, used to train the inpainter, and a segmentation loss that is used to train both the tissue segmenter and inpainter networks. The reconstruction loss uses the mean squared error (MSE) between the original healthy patch and the patch reconstructed by the inpainter. For the segmentation loss, we use a version of the crossentropy loss using probabilistic targets, the probabilistic crossentropy (PCE) loss. Given an output voxel classification y over C classes and a target probability distribution t, the PCE loss is defined as follows:

$$\mathrm{PCE}(y,t) = \sum_{i=1}^{C} -y_i \cdot t_i \cdot \ln\left(\sum_{j=1}^{C} \exp(y_j)\right) \tag{2}$$

By using probabilities as targets instead of categorical labels, we encourage output segmentation that approximates the partial volume probabilities between tissues instead of trying to maximize the probability of the most likely tissue class. Finally, the loss function L is defined as follows:

$$L(\widehat{I_L}, \widehat{S_L}, I_H, S_H) = \mathrm{MSE}(\widehat{I_L}, I_H) + \mathrm{PCE}(\widehat{S_L}, S_H) \tag{3}$$

where $\widehat{I_L}$ is the patch reconstructed by the inpainter, $\widehat{S_L}$ is the tissue probability predicted from the inpainted masked patch and I_H and S_H are the originally healthy image intensities and brain tissue probabilities, respectively.

3.2. Inference

Once the network weights are trained, inference is performed from the T1-w MR image and its WM lesion mask by extracting overlapping patches sampled uniformly with a step size of $5 \times 5 \times 5$ and the same patch size of $16 \times 16 \times 16$ used during training. Performing inference on highly overlapping patches helps reduce block boundary artifacts and improve the spatial coherence of the output probabilities. These patches are then passed through the trained network, obtaining a probability distribution of brain tissue type for each voxel in each input patch. The probability distributions of overlapping patches are then combined through averaging into a common output segmentation space. Finally, the output is normalized to ensure that the tissue probability distributions of each voxel add up to one.

3.3. Implementation details

The proposed method is implemented with Python, using the Torch scientific computing framework (Paszke et al., 2017). All experiments are done on a GNU/Linux machine running Ubuntu 18.04 with 128 GB of RAM memory and an Intel® Core ™ i7–7800X CPU. Network training and inference are performed with an NVIDIA 1080 Ti GPU (NVIDIA Corp., United States) with 12 GB of G5X memory. Within our method, each U-Net model has approximately 7.03 million trainable parameters, which add up to a total of 14.06 million in the multi-task double U-Net configuration. In our system, the total training time of the proposed method is 22.25 h with an average inference time of 55 s per image in all tests performed. The development framework is available to the research community at https://github.com/NIC-VICOROB/LITS.

4. Evaluation and results

In this section, we evaluate the segmentation performance of the proposed methodology as well as the influence of WM lesions with and without lesion effect minimization. First, the healthy and pathological datasets are randomly split into a training and validation image set to train the proposed methodology and a testing set exclusively used for evaluation of the reported experimental results. From the Calgary–Campinas dataset, 45 scans are used for testing, 15 from each scanner, and the remaining 312 are used for training. The 15 WM lesion masks of the MSSEG Challenge dataset are split into 12 for training and 3 for testing. From the ISBI 2015 Longitudinal MS Lesion Segmentation Challenge dataset, 13 WM lesion masks are taken from 3 subjects for training and 8 masks are taken from the other 2 subjects for testing. Finally, we split the WMH Challenge 2017 dataset into 54 masks for training and 6 for testing. In total, the training set contains 312 healthy brain scans, each with 79 registered WM lesion masks, which amounts to 24,648 healthy and artificially lesioned training image pairs. The testing

set is composed of 45 healthy brain scans, each with 17 registered WM lesion masks, making a total of 720 healthy and artificially lesioned testing image pairs. In this way, we ensure that none of the healthy T1-w images or registered lesion masks used during training are used for the evaluation.

To quantitatively evaluate the segmentation differences between healthy and artificially lesioned images, we use the absolute volume difference metric defined in Eq. (4), which is computed separately for the volumes of segmented GM and WM tissues.

$$\text{Abs. volume difference}\,(\%) = 100 \cdot \frac{\left|V_{\text{lesioned}} - V_{\text{healthy}}\right|}{V_{\text{healthy}}} \tag{4}$$

Within the proposed methodology, we also evaluate the differences that WM lesions introduce at local and global scales. In the proposed patch-based method, the introduction of artificially lesioned voxels has a local effect by altering the output probabilities of the whole patch in which they appear. At the global level, the artificially lesioned voxels can alter the input normalization parameters and shift the input values for the whole image, which leads to global output segmentation differences. To evaluate these two effects separately, we also compute the evaluation metrics in two regions of interest (ROIs) related to the lesion neighborhood and patch size. To study the WM lesion effect at a local scale, we define the *within lesion neighborhood* ROI as all the voxels that might appear along the artificial WM lesion in an input patch. More specifically, we include all normally appearing tissue within a patch side length, 16 voxels, of an artificially lesioned voxel in any of the three dimensions. To study the global WM lesion effect, we define the *outside lesion neighborhood* ROI that encompasses all normally appearing voxels at a distance of a patch side length, 16 voxels, or more from an artificially lesioned voxel in all three dimensions.

To assess the statistical significance of differences between the segmentation differences of the baseline and proposed approaches we consider the paired t-test for related samples.

4.1. Tissue segmentation

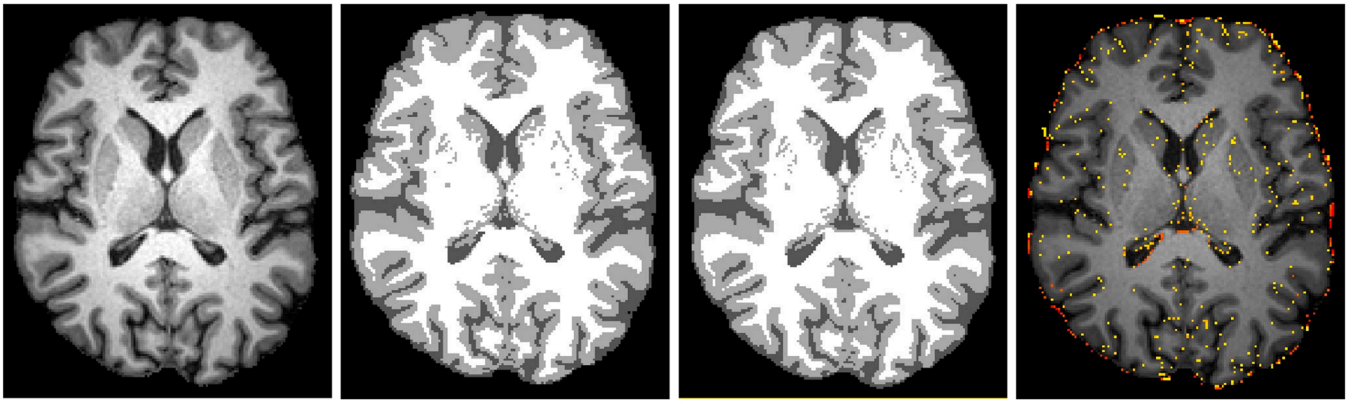
We evaluate the learned tissue segmentation model of the proposed approach by comparing it to FAST (Zhang et al., 2001), the reference method used during training. For this, we segment the 45 testing set images of the healthy dataset without artificially added lesions and compute the Dice similarity coefficient (DSC) with respect to the reference segmentations for the same images. The proposed approach obtains a DSC of 94.6 ± 2.5% in whole brain tissue (CSF + GM + WM) segmentation and a DSC of 99.0 ± 0.1% in parenchyma (GM + WM) segmentation. When individual tissues are considered, the DSCs are 94.6 ± 3.4% and 96.9 ± 1.6% for the GM and WM classes respectively. These results are in line with those of similar deep learning methods also using FAST segmentations as training targets (Rajchl et al., 2018).

Fig. 3 shows qualitative results of segmentation from FAST and the proposed approach as well as the differences between them, which are mainly located within tissue interfaces and in the brain mask edges the large segmentation differences located in the outer brain border appear because FAST assumes every nonzero voxel within the given brain mask has to be segmented as one of the tissues, which in this case is CSF. In contrast, the proposed approach does not make this assumption and mostly classifies voxels in the outer brain border as background instead of CSF. Although the interfaces between tissues with a strong partial volume effect are also a source of segmentation differences, Fig. 3d shows that the changes in classified tissues are due to quite small probability shifts that bias the most likely tissue class one way or the other. The probability differences are larger in the interfaces between WM and CSF, such as the ventricle border and, especially, in its lower left part. In these regions, the partial volumes between GM and CSF take an intensity value similar to that of the GM class and are mostly classified by FAST as GM, while the proposed deep learning approach tends to classify them as mostly WM.

4.2. Lesion effect

We evaluate the effect of WM lesions on tissue segmentation when no WM lesion effect minimization techniques are used. For this, we segment the healthy and artificially lesioned testing image pairs and compute the volume differences between each pair for GM and WM tissues. In this experiment, the inpainting network of the proposed method is essentially turned off, as empty WM masks are used for inference and artificial lesions are not occluded in the input images. We also evaluate the WM lesion effect on the FAST (Zhang et al., 2001) tissue segmentation method from the FSL package. Table 1 shows the absolute volume differences of GM and WM volumes for FAST and our deep learning segmentation method without inpainting. Overall, the proposed method is significantly less influenced by the presence of WM lesions at both local and global scales than FAST ($p < 0.01$). The proposed segmentation method shows an almost exclusively local influence, as nearly all the differences are located within the lesion neighborhood ROI. In contrast, the FAST segmentation method has a mostly global lesion influence, with high volume differences both within and outside the lesion neighborhood ROI.

Fig. 4 shows the tissue probability differences from a representative example of the lesion effect experiment for both tissue segmentation methods. In both cases, artificially lesioned voxels display large



(a) Healthy T1-w (b) FAST segmentation (c) Proposed segmentation (d) Segmentation differences

Fig. 3. Comparison of segmentation results of FAST (b) and the proposed approach (c). (d) Absolute probability differences of voxels changing their most likely tissue class overlaid with a yellow to red colormap, where yellow corresponds to a difference of 0.02 and red corresponds to a difference of 1.0 or higher in the voxelwise sum of absolute probability differences. Differences between both methods are mainly located within tissue interfaces and in the outer brain border.

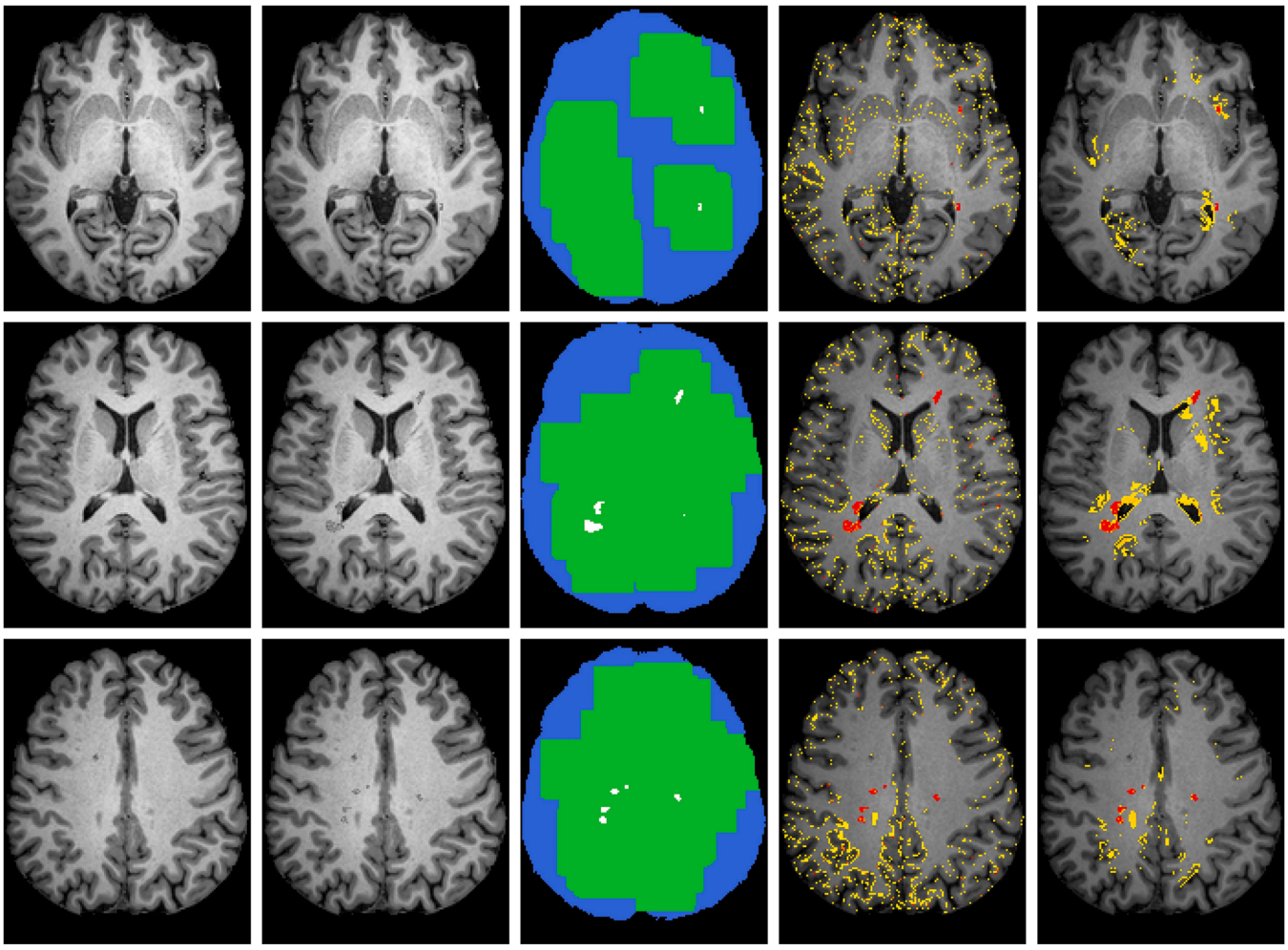
Table 1
Abs. volume differences (%) of the GM and WM of the healthy and artificially lesioned testing image pairs. The absolute volume differences of the proposed approach are all significantly lower ($p < 0.01$) than those of the baseline FAST method.

	FAST		Proposed (without inpainting)	
Tissue	mean ± std	median	mean ± std	median
(i) Whole brain				
GM	0.89 ± 1.14	0.27	0.07 ± 0.09	0.05
WM	1.22 ± 1.58	0.35	0.10 ± 0.11	0.07
(ii) Within lesion neighborhood				
GM	0.96 ± 1.23	0.34	0.13 ± 0.11	0.10
WM	1.10 ± 1.50	0.26	0.13 ± 0.13	0.11
(iii) Outside lesion neighborhood				
GM	0.70 ± 0.89	0.24	0.01 ± 0.03	0.00
WM	1.91 ± 2.66	0.54	0.01 ± 0.06	0.00

probability shifts caused by their newer darker intensities. The effect is not limited to these voxels and spreads to their neighborhood and even to the rest of the image. The FAST tissue segmentation method shows a large number of sparse small and medium probability differences mainly located in the interfaces between GM and WM tissue throughout the whole image. In contrast, the proposed patch-based deep learning approach displays groups of small probability shifts located around the artificially lesioned voxels and nearby structures. The differences are exclusively located in the *within lesion neighborhood* ROI, with no differences in the rest of the image. In contrast, the segmentation differences of FAST appearing within the whole image add up to a larger volume shift.

4.3. Lesion effect minimization

In this experiment, we evaluate how well the WM lesion effect minimization techniques reduce the GM and WM volume differences between segmentations of healthy and artificially lesioned images. In



(a) Healthy (b) Artificially lesioned (c) Lesion mask and ROIs (d) FAST differences (e) Proposed differences (no inpainting)

Fig. 4. Representative example of the absolute segmentation differences between healthy and artificially lesioned brain tissue segmentations without WM lesion effect minimization. Columns 4a and 4b show three axial slices from the healthy and artificially lesioned images that were segmented. In 4c, the artificial lesion mask is shown in white, the *within lesion neighborhood* ROI is shown in green and the *outside lesion neighborhood* ROI is shown in blue. In 4d and 4e, the absolute probability differences are shown overlaid in a yellow to red colormap, where yellow corresponds to a difference of 0.02 and red corresponds to a 1.0 or greater difference in the voxelwise sum of both GM and WM absolute probability differences. While the proposed approach shows large clusters of small differences close to the artificially lesioned voxels, FAST is affected by a larger number of sparsely distributed differences over the whole image which, overall, add up to a larger shift in measured tissue volumes.

the proposed approach, we perform end-to-end inpainting and segmentation by occluding the artificial lesions with zero-valued voxels and providing the WM lesion mask as an additional input to our network. As a baseline comparison, we evaluate the WM lesion inpainting algorithm (FSL_inpainting) provided in the FSL package (Battaglini et al., 2012) to fill the lesion intensities before segmenting the brain tissues with FAST (Zhang et al., 2001). We also evaluate the use of FSL_inpainting to inpaint the WM lesions prior to performing brain tissue segmentation with our deep learning model which, in this case, is provided with empty WM lesion masks to avoid using the inpainter network.

The absolute volume differences of the segmentations of healthy and artificially lesioned versions of the same image are summarized in Table 2. Compared with the results in Table 1, the use of FSL_inpainting and our proposed method significantly reduce the volume differences for all methods ($p < 0.01$). Compared with the FSL_inpainting + FAST pipeline, the FSL_inpainting + Proposed method obtains significantly lower volume differences in all the considered ROIs ($p < 0.001$). This shows that the proposed deep learn- ing brain tissue segmentation framework is more robust to the error introduced by WM lesions even when using classical inpainting methods. Compared with FSL_inpainting + FAST, the proposed approach obtains significantly lower volume differences in all ROIs ($p < 0.001$). However, when comparing with the FSL_inpainting + Proposed approach, the proposed method obtains significantly lower *Whole brain* and *within lesion neighborhood* volume differences ($p < 0.001$), but significantly higher *outside lesion neighborhood* differences ($p < 0.01$). In this case, the input normalization parameters are much less affected by the intensities inpainted by FSL_inpainting than by the zeroes that are used to occlude those same voxels within the proposed approach. The *outside lesion neighborhood* ROI differences of the proposed approach increase significantly compared to those without performing inpainting in Table 1 ($p < 10^{-8}$). This is due to the occlusion with zeroes that we perform to the artificially lesioned voxels in the proposed approach, which slightly change the value of input normalization parameters and increase the segmentation differences for the whole image.

Fig. 5 shows the correlation between the artificial lesion volume and absolute GM and WM volume differences for the evaluated methods. Larger lesion loads tend to increase the segmentation differences for all methods, however, the ones using our deep learning based brain tissue segmentation model show a much lower error when larger lesion volumes are considered. This shows that the poor performance of the FSL pipeline on big lesions is not related to FSL_inpainting, since the proposed deep learning based tissue segmentation framework also takes in images preprocessed with FSL_inpainting and performs much better on larger lesion loads.

In terms of execution time, a brain tissue segmentation done with FAST within our system takes an average of 3.25 min per scan, while the FSL_inpainting part takes less than a second to complete. In total, the FSL pipeline doing WM lesion inpainting and tissue segmentation takes 7 min to process a scan since it requires two separate FAST executions, one to obtain the white matter segmentation mask required by FSL_inpainting and another to obtain the actual brain tissue segmentation from the inpainted image. In contrast, the proposed method doing end-to-end inpainting and tissue segmentation takes an average of 1 min to process a single scan.

Table 2
Abs. volume differences (%) of the GM and WM of the segmentations of healthy and artificially lesioned testing image pairs when using lesion effect minimization techniques. Compared with the FSL inpainting + FAST method, both the Proposed and FSL_inpainting + Proposed approaches obtain significantly lower volume differences in all ROIs than the FSL_inpainting + FAST pipeline ($p < 0.001$). When comparing with the FSL_inpainting + Proposed approach, the proposed method obtains significantly lower *whole brain* and *within lesion neighborhood* volume differences ($p < 0.01$).

	FSL_inpainting + FAST		FSL_inpainting + Proposed		Proposed	
Tissue	mean ± std	median	mean ± std	median	mean ± std	median
(i) Whole brain						
GM	0.05 ± 0.09	0.014	0.02 ± 0.03	0.009	0.01 ± 0.03	0.004
WM	0.08 ± 0.14	0.020	0.03 ± 0.04	0.012	0.02 ± 0.04	0.005
(ii) Within lesion neighborhood						
GM	0.06 ± 0.10	0.019	0.04 ± 0.04	0.021	0.02 ± 0.03	0.008
WM	0.08 ± 0.14	0.018	0.04 ± 0.04	0.020	0.02 ± 0.03	0.007
(iii) Outside lesion neighborhood						
GM	0.04 ± 0.07	0.011	0.01 ± 0.02	0.000	0.01 ± 0.03	0.000
WM	0.13 ± 0.23	0.032	0.01 ± 0.04	0.000	0.03 ± 0.07	0.000

5. Discussion

In this work, we focused on deep learning methods for brain tissue segmentation and performed the first study on the effect of WM lesions in this kind of approaches. We have proposed a deep learning based framework for brain volumetry which learns from a reference classical method and incorporates techniques to deal better with pathological cases having WM lesions. We have also proposed a multi-task double U-Net architecture, along with a training data generation procedure, to embed the WM lesion effect reduction within the brain tissue segmentation method itself. In our approach, instead of performing lesion inpainting in a previous separate step, we perform end-to-end WM lesion inpainting and brain tissue segmentation. By jointly optimizing both tasks, the inpainter is also trained to aid in the segmentation task through the gradient updates coming from the segmentation loss. In this sense, the actual quality or accuracy of inpainting in our framework is not important as long as the output segmentation more faithfully approximates the healthy tissue probabilities.

Without any kind of lesion inpainting, the tissue volumes provided by the proposed deep learning based framework are much less affected by the presence of WM lesions compared to the reference method used for training. Since the introduced artificial lesions affect the tissue probabilities of the patches where they appear, the use of a small patch size constrains the local effect to a smaller area around the lesion. Artificial lesions also change the estimated input normalization parameters which are calculated using all the image intensities. However, the proposed input normalization based on image percentiles is quite robust against these intensity changes and avoids any global segmentation differences in most cases. In comparison, FAST is affected by a larger number of sparse segmentation differences spread out over the whole image which, overall, add up to a larger shift in measured tissue volumes. This is most likely due to the initial k-means clustering step that FAST performs over the entire image to estimate the mean intensity of each tissue, which is later used during the estimation of partial volume probabilities. The introduction of artificial lesions biases the estimated mean intensity of each tissue which in turn biases the estimation of partial volume distributions, producing the observed segmentation differences in the interfaces between tissues.

In terms of WM lesion effect minimization, both the FSL_inpainting and our proposed approach significantly reduce their effect on the measured tissue volumes. However, we obtain significantly lower volume differences than the baseline FSL pipeline, especially when considering larger lesion loads. The results in Fig. 5 show that our deep learning tissue segmentation framework provides significant improvement even when using FSL_inpainting to preprocess the images. Furthermore, our proposed deep learning framework is faster, taking just under a minute to segment a whole brain scan while the baseline FSL pipeline takes an average of 7 min.

The main limitation of this study is that we cannot assess or evaluate

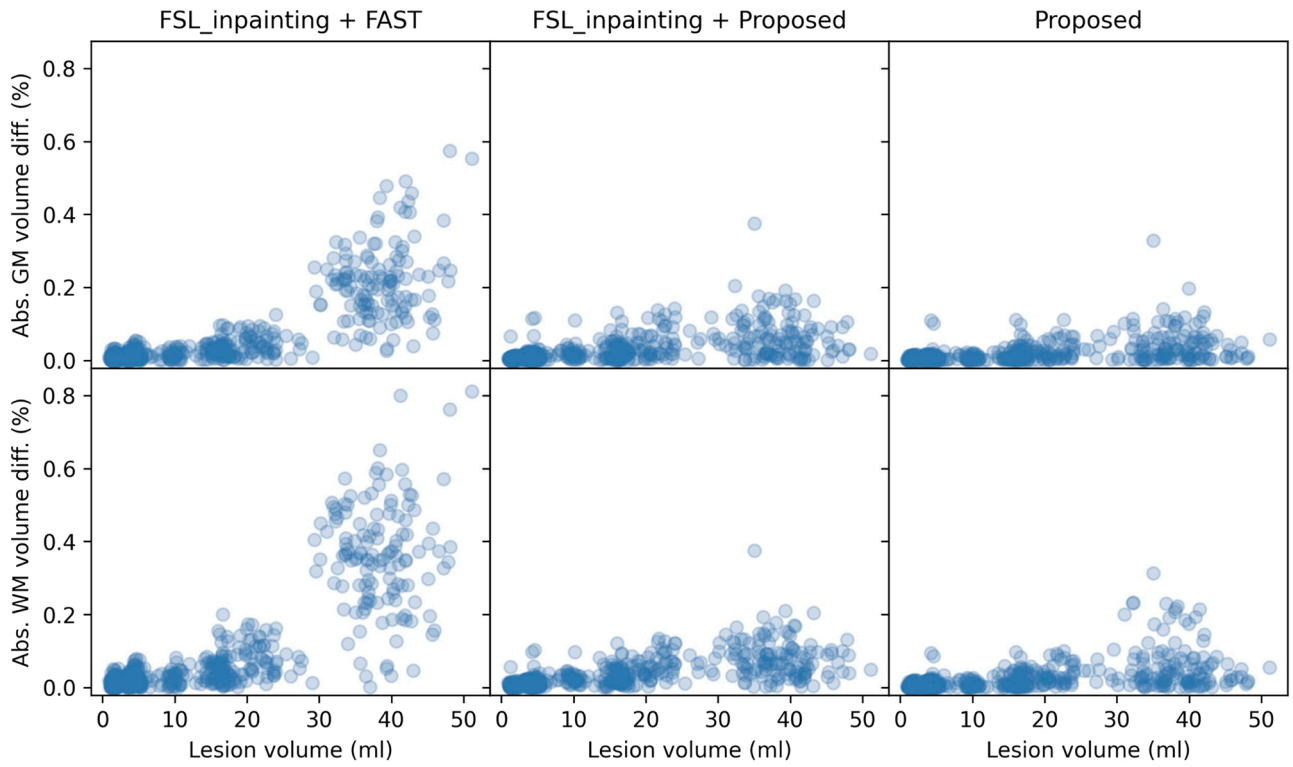


Fig. 5. Correlation of artificial lesion volume and the absolute volume differences (%) of the GM and WM of healthy and artificially lesioned images when using lesion inpainting.

the accuracy and precision of the learned tissue segmentation model and its lesion effect minimization performance on real WM lesions. Due to the way in which the proposed approach is trained, this requires a large database of MR images with manually annotated brain tissue and WM lesions of both healthy and pathological subjects. However, there is no such database, and our evaluation has therefore been limited to relative comparisons with FAST as the gold standard on artificially lesioned images. In this sense, our approach presents a lower WM lesion effect with and without inpainting with a Dice similarity coefficient of 94.6 $\pm$ 2.5% relative to FAST brain tissue segmentations. Unlike supervised learning methods using manually annotated segmentations for training, a higher DSC compared to that of the FAST segmentation is not indicative of better quality or accuracy, just of higher similarity. Unlike FAST, deep learning methods suffer from the domain adaptation issue where their performance is not guaranteed outside of the image domains used during training. In this sense, a different MR scanner or acquisition protocol than those used during training would likely lead to a decreased segmentation performance. In such cases, training a model from scratch on the target image domain only requires a set of healthy MR images from that domain to which WM lesion masks from publicly accessible pathological scans can be registered to train the proposed method. Another option is to use domain adaptation techniques that fine-tune pretrained network weights to optimize the model for the target domain.

In the proposed method, accurate WM lesion segmentation is required to obtain optimal results, and over or undersegmentation of the WM lesion would still introduce volume errors in the output segmentation. This could be an issue since manual lesion delineation or automated lesion segmentation is often performed on FLAIR MR images, while brain tissue volumetry is usually performed on T1-w MR images (Rovira et al., 2015). In this case, the FLAIR lesion segmentation mask is usually registered to the T1 image and might not encompass all abnormally appearing voxels in the target modality image. In the case of oversegmentation, the method can deal just as well with the inpainting and segmentation of larger occluded areas as long as they are to be segmented as WM. Due to the way the method was trained, any occluded voxel is assumed to be WM in its majority and will be segmented as such. If the WM lesion is undersegmented, the lesioned voxels are not inpainted, which introduces errors in neighboring tissue segmentation. However, the experimental results without inpainting show that the effect is still be smaller than that of FAST and confined to the undersegmented lesioned voxels neighborhood.

6. Conclusions

In this work, we focus on deep learning based tissue segmentation methods for brain volumetry and studied the error introduced by WM lesions. We have proposed a deep learning framework for brain tissue segmentation which is much less affected by WM lesions than the reference method used to train thanks to the use of small patches and a percentile-based input normalization. We have also proposed a multi-task double U-Net architecture, along with a training data generation procedure, which performs lesion inpainting and tissue segmentation in an end-to-end manner and can reduce the WM lesion effect to almost negligible levels. Reducing the effect of WM lesions is critical for accurate and reliable cross-sectional volumetry or longitudinal brain atrophy quantification. Typically, state-of-the-art atrophy quantification approaches are based either on boundary shift integration (Smith et al., 2002) or Jacobian integration (Boyes et al., 2006), both of which rely on prior accurate segmentation of brain tissue which needs to be robust against the influence of WM lesions. Automated brain volumetry methods are currently only used to evaluate the efficacy of experimental therapies and to correlate with treatment outcomes in clinical studies. Improving their accuracy would either strengthen the statistical significance of correlations or reduce the sample sizes needed to establish them. In routine clinical practice, the use of brain volumetry methods is discouraged for prognosis, such as assessing patient progression in MS (Rovira et al., 2015). These methods are unreliable when applied to a single subject instead of a large population due to the inherent technical issues and other confounding factors that severely affect brain volumetry methods. Improving the accuracy and reducing the error from

confounding factors such as WM lesions is critical to unlock brain volumetry as an imaging marker for the prognosis of patients with neurodegenerative diseases. In this sense, the proposed deep learning methodology is significantly less affected by WM lesions and can minimize the error they introduce in the measured tissue volumes.

CRediT authorship contribution statement

Albert Clèrigues: Conception and design of study, analysis and/or interpretation of data, Drafting the manuscript, Approval of the version of the manuscript to be published. **Sergi Valverde:** Conception and design of study, analysis and/or interpretation of data, Approval of the version of the manuscript to be published. **Joaquim Salvi:** Drafting the manuscript, revising the manuscript critically for important intellectual content, Approval of the version of the manuscript to be published. **Arnau Oliver:** Conception and design of study, Drafting the manuscript, revising the manuscript critically for important intellectual content, Approval of the version of the manuscript to be published. **Xavier Lladó:** Conception and design of study, Drafting the manuscript, revising the manuscript critically for important intellectual content, Approval of the version of the manuscript to be published.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

Albert Clèrigues holds an FPI grant from the Ministerio de Ciencia e Innovación with reference number PRE2018-083507. This work has been partially supported by DPI2020-114769RB-I00 from the Ministerio de Ciencia e Innovación. This work is also supported by ICREA under the ICREA Academia programme.

References

Armanious, K., Mecky, Y., Gatidis, S., Yang, B., 2019. Adversarial inpainting of medical image modalities. In: Proceedings of the ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 3267–3271.

Battaglini, M., Jenkinson, M., Stefano, N.D., 2012. Evaluating and reducing the impact of white matter lesions on brain volume measurements. Hum. Brain Mapp. 33, 2062–2071.

Bendfeldt, K., Kuster, P., Traud, S., Egger, H., Winklhofer, S., Mueller-Lenke, N., Naegelin, Y., Gass, A., Kappos, L., Matthews, P.M., Nichols, T.E., Radue, E.W., Borgwardt, S.J., 2009. Association of regional gray matter volume loss and progression of white matter lesions in multiple sclerosis — a longitudinal voxel-based morphometry study. NeuroImage 45, 60–67.

Boyes, R.G., Rueckert, D., Aljabar, P., Whitwell, J., Schott, J.M., Hill, D.L., Fox, N.C., 2006. Cerebral atrophy measurements using jacobian integration: comparison with the boundary shift integral. NeuroImage 32, 159–169.

Carass, A., Roy, S., Jog, A., Cuzzocreo, J.L., Magrath, E., Gherman, A., Button, J., Nguyen, J., Bazin, P.L., Calabresi, P.A., Crainiceanu, C.M., Ellingsen, L.M., Reich, D. S., Prince, J.L., Pham, D.L., 2017. Longitudinal multiple sclerosis lesion segmentation data resource. Data Brief 12, 346–350.

Chard, D.T., Jackson, J.S., Miller, D.H., Wheeler-Kingshott, C.A., 2010. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. J. Magn. Reson. Imaging 32, 223–228.

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., Mahbod, A., Wang, C., McKinley, R., Wagner, F., Muschelli, J., Sweeney, E., Roura, E., Lladó, X., Santos, M.M., Santos, W.P., Silva-Filho, A.G., Tomas-Fernandez, X., Urien, H., Bloch, I., Valverde, S., Cabezas, M., Vera-Olmos, F.J., Malpica, N., Guttmann, C., Vukusic, S., Edan, G., Dojat, M., Styner, M., Warfield, S.K., Cotton, F., Barillot, C., 2018. Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. Sci. Rep. 8, 13650.

Cortese, R., Battaglini, M., Sormani, M.P., Luchetti, L., Gentile, G., Inderyas, M., Alexandri, N., De Stefano, N., 2022. Reduction in grey matter atrophy in patients with relapsing multiple sclerosis following treatment with cladribine tablets. Eur. J. Neurol.

Di Filippo, M., Anderson, V.M., Altmann, D.R., Swanton, J.K., Plant, G.T., Thompson, A. J., Miller, D.H., 2010. Brain atrophy and lesion load measures over 1 year relate to clinical status after 6 years in patients with clinically isolated syndromes. J. Neurol. Neurosurg. Psychiatry 81, 204–208.

Ghione, E., Bergsland, N., Dwyer, M., Hagemeier, J., Jakimovski, D., Ramasamy, D., Hojnacki, D., Lizarraga, A., Kolb, C., Eckert, S., Weinstock-Guttman, B., Zivadinov, R., 2020. Disability improvement is associated with less brain atrophy development in multiple sclerosis. Am. J. Neuroradiol. 41, 1577–1583.

González-Villà, S., Valverde, S., Cabezas, M., Pareto, D., Vilanova, J.C., Ramió-Torrentà, L., Alex Rovira, Oliver, A., Lladó, X., 2017. Evaluating the effect of multiple sclerosis lesions on automatic brain structure segmentation. NeuroImage Clin. 15, 228–238.

Guha Roy, A., Conjeti, S., Navab, N., Wachinger, C., 2019. Quicknat: a fully convolutional network for quick and accurate segmentation of neuroanatomy. NeuroImage 186, 713–727.

Henschel, L., Conjeti, S., Estrada, S., Diers, K., Fischl, B., Reuter, M., 2020. Fastsurfer - a fast and accurate deep learning based neuroimaging pipeline. NeuroImage 219, 117012.

Iglesias, J.E., Liu, C.Y., Thompson, P.M., Tu, Z., 2011. Robust brain extraction across datasets and comparison with publicly available methods. IEEE Trans. Med. Imaging 30, 1617–1634.

Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of the International Conference on Machine Learning, pp. 448–456.

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5, 143–156.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17, 825–841.

Kuijf, H.J., Casamitjana, A., Collins, D.L., Dadar, M., Georgiou, A., Ghafoorian, M., Jin, D., Khademi, A., Knight, J., Li, H., Lladó, X., Biesbroek, J.M., Luna, M., Mahmood, Q., Mckinley, R., Mehrtash, A., Ourselin, S., Park, B.Y., Park, H., Park, S. H., Pezold, S., Puybareau, E., Bresser, J.D., Rittner, L., Sudre, C.H., Valverde, S., Vilaplana, V., Wiest, R., Xu, Y., Xu, Z., Zeng, G., Zhang, J., Zheng, G., Heinen, R., Chen, C., Flier, W.V.D., Barkhof, F., Viergever, M.A., Biessels, G.J., Andermatt, S., Bento, M., Berseth, M., Belyaev, M., Cardoso, M.J., 2019. Standardized assessment of automatic segmentation of white matter hyperintensities and results of the wmh segmentation challenge. IEEE Trans. Med. Imaging 38, 2556–2568.

Lansley, J., Mataix-Cols, D., Grau, M., Radua, J., Sastre-Garriga, J., 2013. Localized grey matter atrophy in multiple sclerosis: a meta-analysis of voxel-based morphometry studies and associations with functional disability. Neurosci. Biobehav. Rev. 37, 819–830.

Magon, S., Gaetano, L., Chakravarty, M.M., Lerch, J.P., Naegelin, Y., Stippich, C., Kappos, L., Radue, E.-W., Sprenger, T., 2014. White matter lesion filling improves the accuracy of cortical thickness measurements in multiple sclerosis patients: a longitudinal study. BMC Neurosci. 15.

Manjón, J.V., Romero, J.E., Vivo-Hernando, R., Rubio, G., Aparici, F., de la Iglesia-Vaya, M., Tourdias, T., Coupé, P., 2020. Blind mri brain lesion inpainting using deep learning. In: Proceedings of the International Workshop on Simulation and Synthesis in Medical Imaging 12417 LNCS, pp. 41–49.

Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the International Conference on Machine Learning, pp. 807–814.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017. Automatic differentiation in pytorch. Neural Inf. Process. Syst.

Pérez-Miralles, F., Sastre-Garriga, J., Tintoré, M., Arrambide, G., Nos, C., Perkal, H., Río, J., Edo, M., Horga, A., Castilló, J., Auger, C., Huerga, E., Rovira, A., Montalban, X., 2013. Clinical impact of early brain atrophy in clinically isolated syndromes. Mult. Scler. J. 19, 1878–1886.

Prados, F., Cardoso, M.J., Kanber, B., Ciccarelli, O., Kapoor, R., Wheeler-Kingshott, C.A. G., Ourselin, S., 2016. A multi-timepoint modality-agnostic patch-based method for lesion filling in multiple sclerosis. NeuroImage 139, 376–384.

Rajchl, M., Pawlowski, N., Rueckert, D., Matthews, P.M., Glocker, B., 2018. Neuronet: Fast and robust reproduction of multiple brain image segmentation pipelines. arXiv preprint arXiv:1806.04224.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention, 9351, pp. 234–241.

Rovira, À., Wattjes, M.P., Tintoré, M., Tur, C., Yousry, T.A., Sormani, M.P., Stefano, C., Filippi, M., Auger, C., Rocca, M.A., Barkhof, F., Fazekas, F., Kappos, L., Polman, C., Miller, D., Montalban, X., 2015. Magnims consensus guidelines on the use of mri in multiple sclerosis—clinical implementation in the diagnostic process. Nat. Rev. Neurol. 8 (11), 471–482.

Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P., Federico, A., De Stefano, N., 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. NeuroImage 17, 479–489.

Sotirchos, E.S., Gonzalez-Caldito, N., Dewey, B.E., Fitzgerald, K.C., Glaister, J., Filippatou, A., Ogbuokiri, E., Feldman, S., Kwakyi, O., Risher, H., Crainiceanu, C., Pham, D.L., Zijl, P.C.V., Mowry, E.M., Reich, D.S., Prince, J.L., Calabresi, P.A., Saidha, S., 2020. Effect of disease-modifying therapies on subcortical gray matter atrophy in multiple sclerosis. Mult. Scler. 26, 312–321.

Souza, R., Lucena, O., Garrafa, J., Gobbi, D., Saluzzi, M., Appenzeller, S., Rittner, L., Frayne, R., Lotufo, R., 2018. An open, multi-vendor, multi-field-strength brain mr dataset and analysis of publicly available skull stripping methods agreement. NeuroImage 170, 482–494.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826.

Tang, Z., Cabezas, M., Liu, D., Barnett, M., Barnett, W., Wang, C., 2021. Lg-net: lesion gate network for multiple sclerosis lesion inpainting. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 660–669.

Valverde, S., Oliver, A., Lladó, X., 2014. A white matter lesion-filling approach to improve brain tissue volume measurements. NeuroImage Clin. 6, 86–92.

Wardlaw, J.M., Smith, E.E., Biessels, G.J., Cordonnier, C., Fazekas, F., Frayne, R., Lindley, R.I., O'Brien, J.T., Barkhof, F., Benavente, O.R., et al., 2013. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. Lancet Neurol. 12, 822–838.

Xiong, H., Wang, C., Barnett, M., Wang, C., 2020. Multiple sclerosis lesion filling using a non-lesion attention based convolutional network. In: Proceedings of the International Conference on Neural Information Processing 12532 LNCS, pp. 448–460.

Zeiler, M.D., 2012. Adadelta: an adaptive learning rate method. arXiv preprint arXiv: 1212.5701.

Zhang, H., Bakshi, R., Bagnato, F., Oguz, I., 2020. Robust multiple sclerosis lesion inpainting with edge prior. Mach. Learn. Med. Imaging 120–129.

Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging 20, 45–57.