

MPRAGE to MP2RAGE UNI translation via generative adversarial network improves the automatic tissue and lesion segmentation in multiple sclerosis patients

Francesco La Rosa^{a,b,c,*}, Thomas Yu^{a,b,c}, Germán Barquero^{a,b,c}, Jean-Philippe Thiran^{a,d}, Cristina Granziera^{e,f}, Meritxell Bach Cuadra^{b,c,d}

^a Signal Processing Laboratory (LTS5), Ecole Polytechnique Fédérale de Lausanne, Switzerland

^b CIBM Center for Biomedical Imaging, Switzerland

^c Medical Image Analysis Laboratory (MIAL), Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

^d Department of Radiology, Lausanne University Hospital and University of Lausanne, Switzerland

^e Neurologic Clinic and Policlinic, Departments of Medicine, Clinical Research and Biomedical Engineering, University Hospital Basel and University of Basel, Basel, Switzerland

^f Translational Imaging in Neurology (ThINK) Basel, Department of Medicine and Biomedical Engineering, University Hospital Basel and University of Basel, Basel, Switzerland

ARTICLE INFO

Keywords:

MRI
MP2RAGE
MPRAGE
GAN
Multiple sclerosis

ABSTRACT

Background and objective: Compared to the conventional magnetization-prepared rapid gradient-echo imaging (MPRAGE) MRI sequence, the specialized magnetization prepared 2 rapid acquisition gradient echoes (MP2RAGE) shows a higher brain tissue and lesion contrast in multiple sclerosis (MS) patients. The goal of this work is to retrospectively generate realistic-looking MP2RAGE uniform images (UNI) from already acquired MPRAGE images in order to improve the automatic lesion and tissue segmentation.

Methods: For this task we propose a generative adversarial network (GAN). Multi-contrast MRI data of 12 healthy controls and 44 patients diagnosed with MS was retrospectively analyzed. Imaging was acquired at 3T using a SIEMENS scanner with MPRAGE, MP2RAGE, FLAIR, and DIR sequences. We train the GAN with both healthy controls and MS patients to generate synthetic MP2RAGE UNI images. These images were then compared to the real MP2RAGE UNI (considered as ground truth) analyzing the output of automatic brain tissue and lesion segmentation tools. Reference-based metrics as well as the lesion-wise true and false positives, Dice coefficient, and volume difference were considered for the evaluation. Statistical differences were assessed with the Wilcoxon signed-rank test.

Results: The synthetic MP2RAGE UNI significantly improves the lesion and tissue segmentation masks in terms of Dice coefficient and volume difference (p-values < 0.001) compared to the MPRAGE. For the segmentation metrics analyzed no statistically significant differences are found between the synthetic and acquired MP2RAGE UNI.

Conclusion: Synthesized MP2RAGE UNI images are visually realistic and improve the output of automatic segmentation tools.

1. Introduction

Magnetic resonance imaging (MRI) plays a crucial role in multiple sclerosis (MS) as it is the imaging technique of choice for diagnosing the disease and monitoring its progression [1]. MRI for MS includes both T1-weighted and T2-weighted sequences. T1-weighted images are

commonly preferred for visualizing the anatomy of the brain and quantifying normal appearing gray matter (GM) and white matter (WM), but are also helpful for detecting lesional tissue in MS patients [2]. Currently, one of the most used T1-weighted sequences at high magnetic field (3T) is the three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MPRAGE) [3]. This sequence offers

* Corresponding author. Signal Processing Laboratory (LTS5), Ecole Polytechnique Fédérale de Lausanne, Switzerland.

E-mail address: francesco.laros@epfl.ch (F. La Rosa).

<https://doi.org/10.1016/j.complbiomed.2021.104297>

Received 20 December 2020; Received in revised form 8 February 2021; Accepted 22 February 2021

Available online 26 February 2021

0010-4825/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

accurate anatomical images of the brain in a reasonable acquisition time (about 5 min) and is routinely included in the standard clinical protocols.

An extension of the MPRAGE is the so-called magnetization prepared 2 rapid acquisition gradient echoes (MP2RAGE) [4]. This specialized sequence combines two images acquired at different inversion times, creating T1-weighted uniform images (UNI) with excellent tissue contrast and self-correction for B1- bias field. Moreover, in addition to the UNI, a T1 relaxation map (T1 map) can also be concurrently obtained from the same acquisition. Although the T1 map provides important quantitative information, the UNI image is the one primarily used for both visual and automatic inspection. To the best of our knowledge, the T1 maps are not exploited for tissue or lesion segmentation by any automatic methods. For this reason, we focus on the UNI images, and throughout this work, by MP2RAGE we refer to its UNI image.

The MP2RAGE has been shown to yield an improved tissue segmentation compared to the MPRAGE with classical segmentation tools optimized for conventional T1 contrasts [5,6]. Additionally, further studies [7,8] have described its valuable application to MS patients, obtaining an improved lesion visualization and detection, in particular regarding cortical lesions (CLs). Currently, despite its promising added value, MP2RAGE remains mainly in research settings as changing the MRI clinical protocols is a lengthy process. It would thus be highly beneficial if MP2RAGE-like images could be estimated from current MPRAGE acquisitions in order to benefit from the enhanced tissue contrast of MP2RAGE without waiting for its clinical adoption.

In the last decade, a class of deep learning algorithms called generative adversarial networks (GANs) [9] have emerged as the state-of-the-art technique for generating new *synthetic* data. Their two main components are a generator, responsible for synthesizing new realistic data from a given input, and a discriminator, whose goal is to distinguish real and synthetic data. The generator and discriminator are trained simultaneously in competition with each other, and this results in realistic-looking data being produced. GANs are especially effective with images, exploiting convolutional neural networks (CNNs) for their generator and discriminator. Starting from the computer vision field, they are finding several broad applications and have been recently explored also for medical imaging [10]. Considering that data scarcity and class imbalance often represent an obstacle for training CNNs, GANs have arguably several potential applications also in the medical field. Their main drawback, however, is that in some cases they introduce artefacts or unrealistic details, which cannot be tolerated in the clinical context either for diagnosis or for follow up purposes.

GANs have been explored for MRI for image reconstruction [11], to increase the image resolution [12], augmenting and increasing dataset size [13], as well as generating new contrasts or converting MRI images to computed tomography (and vice-versa) [14]. Specifically for MS patients, they have been shown to be an effective method for data augmentation [15] and also for generating realistic-looking MRI contrasts. Recently, Finck et al. [16] have proposed a GANs for the generation of synthetic double inversion recovery (DIR) images starting from standard MRI acquisitions as T1, T2, and fluid-attenuated inversion recovery (FLAIR) of MS patients. Two independent readers then evaluated the images, showing that the synthetic DIR was able to depict significantly more MS lesions compared to the conventional FLAIR. This could have an important application for MS diagnosis. However, as argued by Hagiwara et al. in a recent editorial article [17], adding a synthetic contrast implicitly requires additional time by the experts to analyze it. They conclude mentioning that it would be interesting to explore how an automatic lesion segmentation method would perform with it. Notably, CNNs are inspired by the human brain's learning process, but their way of extracting and combining features does not necessarily reflect what experts do.

In this work, we propose a GAN that, through a pixel-by-pixel image translation process, synthesizes the MP2RAGE images corresponding to

the input MPRAGE images. The GAN is trained on a dataset consisting of 12 healthy controls and 8 MS patients with axial, sagittal, and coronal 2D slices. The trained model is then tested on 36 MS patients in the early stages of the disease. The synthesised images are evaluated in two ways. First, we compare the MPRAGE images to the synthesized MP2RAGE images using quality metrics. Second, we compare differences in the segmentation of these images using two different automated methods: a supervised MS lesion segmentation CNN [19] and an unsupervised tissue segmentation approach [20].

2. Materials and method

2.1. Subjects

In this study, we consider a cohort of 56 subjects (36 female/20 male, mean age 34 ± 9 , age range [20–61] years). Of these, 12 are healthy controls, and the remaining 44 MS patients diagnosed with relapsing-remitting MS according to the McDonald criteria [2]. These patients were at early stages of the disease: the mean disease duration was (1.9 ± 1.5) years and the Expanded Disability Status Scale (EDSS) scores ranged from 1 to 2 (mean 1.5 ± 0.3).

2.2. Imaging

Imaging was acquired on a 3T MRI whole-brain scanner (MAGNETOM Trio, Siemens Healthcare, Erlangen, Germany) with the following 3D sequences (acquired in the sagittal plane) with a resolution of $1 \times 1 \times 1.2 \text{ mm}^3$: 3D FLAIR (field-of-view = $256 \times 240 \times 176$, TR/TE/TI = 5000,394,1800 ms, flip angle = 120° , acquisition time = 6 min), 3D MPRAGE (field-of-view = $256 \times 240 \times 160$, TR/TE/TI = 2300,2,98,900 ms, flip angle = 7° , acquisition time = 5 min), DIR (field-of-view = $256 \times 240 \times 160$, TR/TE/TI1/TI2 = 10000,218,450/3650 ms, flip angle = 90° , acquisition time = 13 min), and MP2RAGE (field-of-view = $256 \times 240 \times 176$, TR/TE/TI1/TI2 = 5000,2,89,700,2500, flip angles = $4^\circ/5^\circ$, acquisition time = 8 min). In the following sections we will refer to FLAIR and MPRAGE as "conventional sequences", as these are part of the clinical protocol, and to DIR and MP2RAGE as "specialized sequences", as these are mainly used in a research setting.

The study was approved by the Ethics Committee of our institution, and all subjects gave written informed consent prior to participation.

2.3. Manual segmentation

A radiologist and a neurologist, with 7 and 11 years of experience in MS research respectively, detected white matter lesions (WMLs) and CLs by consensus on the MS patients' scans using all four imaging modalities and the three orthogonal planes. A trained technician then manually delineated all the lesions considering again both the conventional and advanced sequences using ITK-SNAP [21]. Importantly, pathological validation remains the ultimate gold standard for lesion detection, whereas this work does not aim at evaluating the MRI sensitivity compared to pathology.

2.4. Generative adversarial network

The goal of image-to-image translation networks is to learn the mapping between an input and an output image. GANs have been successfully proposed as a solution for this task, where the generator learns this mapping, while the competing discriminator pushes it to further improve. One of the first and top-performing image-to-image translation networks proposed is the pix2pix architecture [18]. This is a general-purpose GAN which includes a U-Net-like CNN as generator, and a PatchGAN as discriminator. It has been applied to a wide range of tasks, such as translating day photographs to night ones or sketches to photographs [18]. Pix2pix combines a common pixel-wise L1 loss with the adversarial loss of the discriminator, showing how the latter

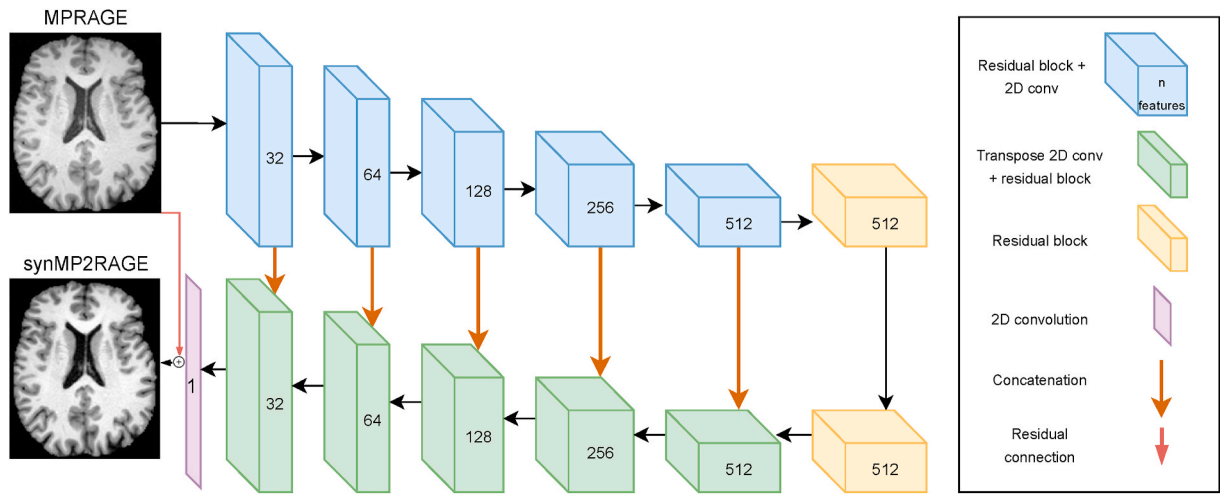


Fig. 1. Scheme of the generator of the GAN proposed. For more details see Table 1 of the Supplementary material.

consistently helps improve the results.

Proposed framework. Inspired by the pix2pix architecture [18], we propose a pixel-wise translation network that receives as input MPRAGE images and outputs realistic-looking MP2RAGE ones (synMP2RAGE). The original implementation is adapted with additional residual blocks in the generator increasing the overall number of parameters. Moreover, as proposed by Kupyn et al. [22], we include a global skip connection between the input and the output of the last layer of the generator. In this way, the CNN learns a residual correction to the input MPRAGE image. This empirically decreased the training time and improved the generator robustness. The architecture of the generator is illustrated in Fig. 1 (see the Supplementary material for more details). Contrary to pix2pix, our discriminator classifies at each iteration the entire image as either real or fake, and it is composed of five convolutional layers, each one followed by a Leaky ReLU activation function. The downsampling is learned through the convolutions and no max-pooling or fully-connected layers are present, as recommended in a previous work [23]. The complete fully-convolutional discriminator's architecture has about 50k trainable parameters (see Table 2 in the Supplementary material).

Loss functions. The first loss considered is the pixel-wise mean absolute error (L1 loss) between the images produced by the generator and the target ones, as used in most works in the literature [12,18,22]. The second loss is given by the ability of the generator of fooling the

discriminator. This is computed as the sigmoid cross-entropy (adversarial loss) between the discriminator output and an array of 0s and 1s, where the 0s represent fake images and the 1s real ones. MRI acquisitions, however, are considerably affected by noise, and the L1 and the adversarial losses are not sufficient to produce realistic-looking images. In order to account also for the visual quality of the produced MP2RAGE considering overall spatial features/textures, we added a perceptual loss. This is computed as the mean absolute error between the feature maps of the fourth convolutional layer of the pre-trained VGG-16 model produced by the real and the synthesized MP2RAGE. The fourth layer was chosen as it gave the best results experimentally and it is as well in accordance with a previous work on MRI [24]. Deeper layers extract more abstract features and did not seem to be beneficial for our scope. The perceptual loss, together with the adversarial one, is responsible for making the images look more realistic. Empirically, we observed that they prevented the network from over-smoothing the synthesized MP2RAGE, which is also observed to occur in the use of perceptual losses in computer vision [25,26]. Summing up, the total loss is given by:

$$L_{\text{Tot}} = (\alpha * L_1) + (\beta * L_{\text{perceptual}}) + L_{\text{adversarial}}$$

Where α and β were set to 150 and 5, respectively. Generator and discriminator are trained and updated simultaneously at each iteration with the relative losses, refer to Fig. 2 for a scheme of the framework.

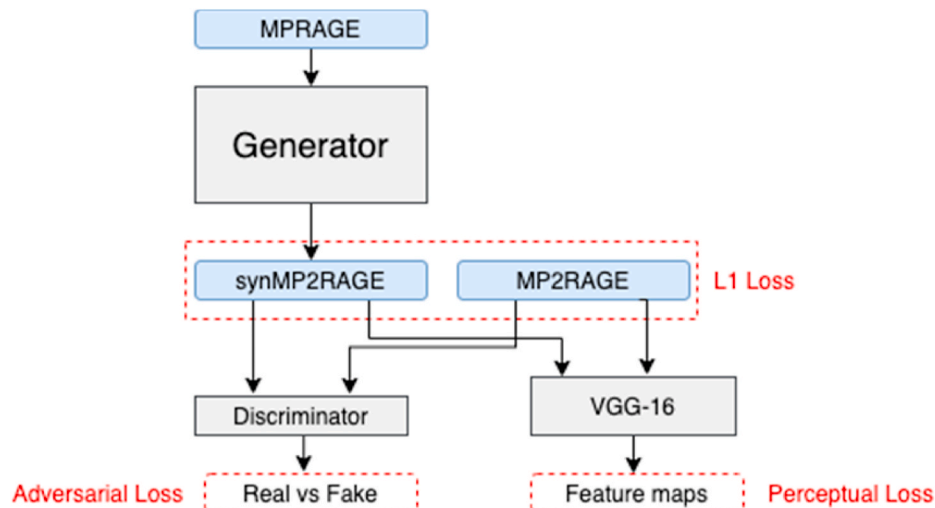


Fig. 2. Scheme of the generative framework proposed. The three loss functions used are highlighted in red.

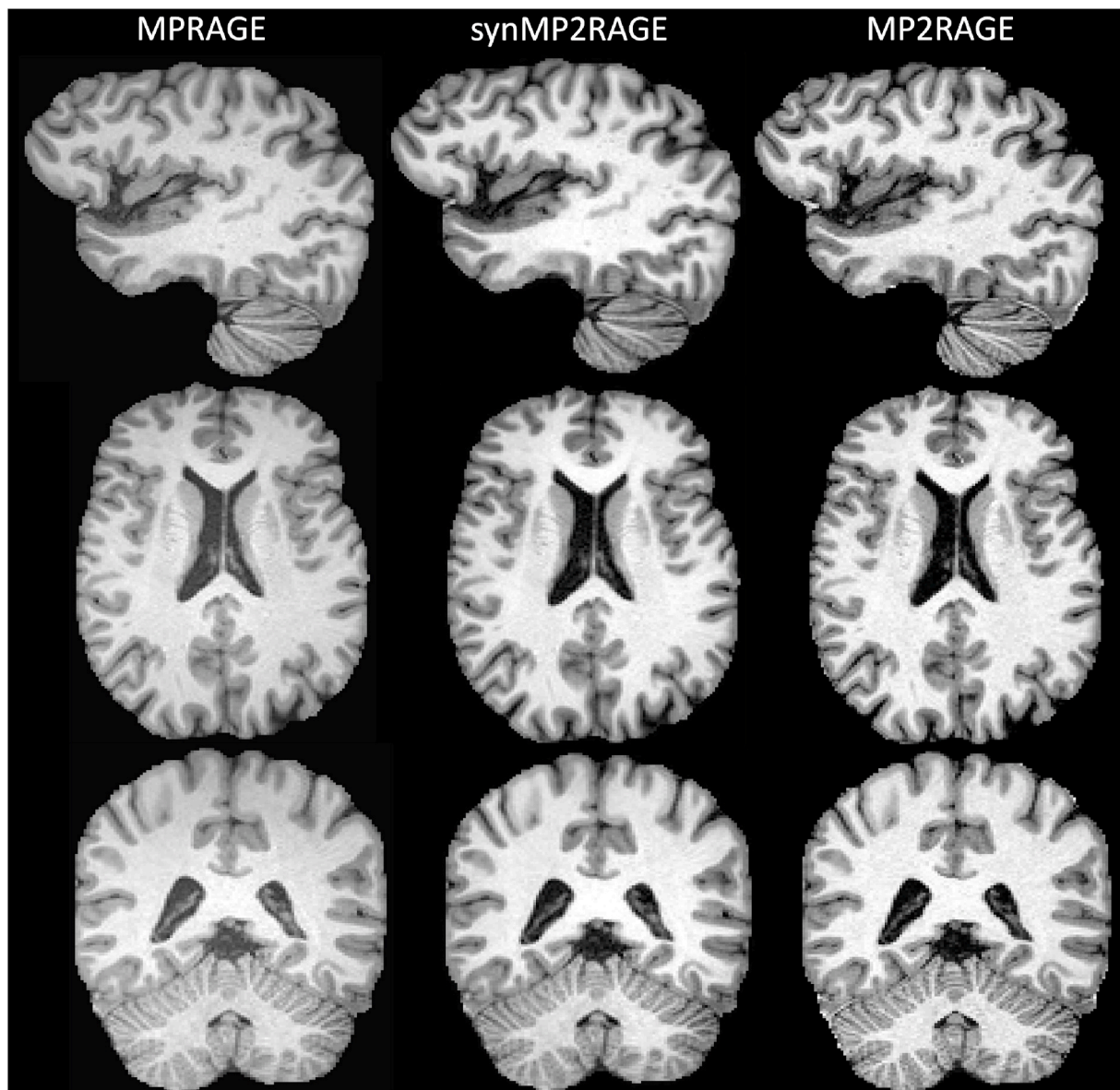


Fig. 3. Qualitative results showing the synMP2RAGE compared to the acquired one and the MP2RAGE.

2.5. Pre-processing

First, each MPRAGE acquisition in our dataset is rigidly registered to the corresponding MP2RAGE UNI using ELASTIX [27]. Second, the MPRAGE image is skull-stripped with ANTs [20] and the brain mask is applied to the MP2RAGE image. This step is necessary in order to remove the noise present outside the brain in the MP2RAGE. Finally, both volumes for each subject are normalized to zero mean and standard deviation of one.

2.6. Training details

For each input 3D MPRAGE volume, we extracted 150×150 pixel slices across the three orthogonal views of the brain (axial, coronal, and sagittal planes). The 9000 2D slices obtained in this way were then concatenated together. This particular size was chosen as it includes the entire brain for all subjects. Data augmentation is then performed in order to prevent overfitting. In particular, we randomly crop 128×128 or 64×64 images at each iteration. The latter ones are then resized to the 128×128 input size using bilinear interpolation. Moreover, random

flipping along both axes is applied. The images obtained are then fed as input to the generator with a mini-batch size of 1. The initial learning rate set is $1e-5$ with Adam [28] as optimizer for both the generator and the discriminator.

The generative framework has been developed in Tensorflow 2.1.0 [29] using one NVIDIA Tesla P100 GPU. The code is deployed as a Jupyter Google Colaboratory¹ notebook which simply runs on the internet browser of any computer taking advantage of the free GPU usage available in Colaboratory. The code is available on our research website.²

2.7. Inference

At inference time the MPRAGE is skull-stripped, normalized to zero mean and unit standard deviation, and zero-padded, obtaining 256×256 pixels slices in each of the three different planes. As the generator's

¹ www.colab.research.google.com.

² <https://github.com/FrancescoLR/synMP2RAGE>.

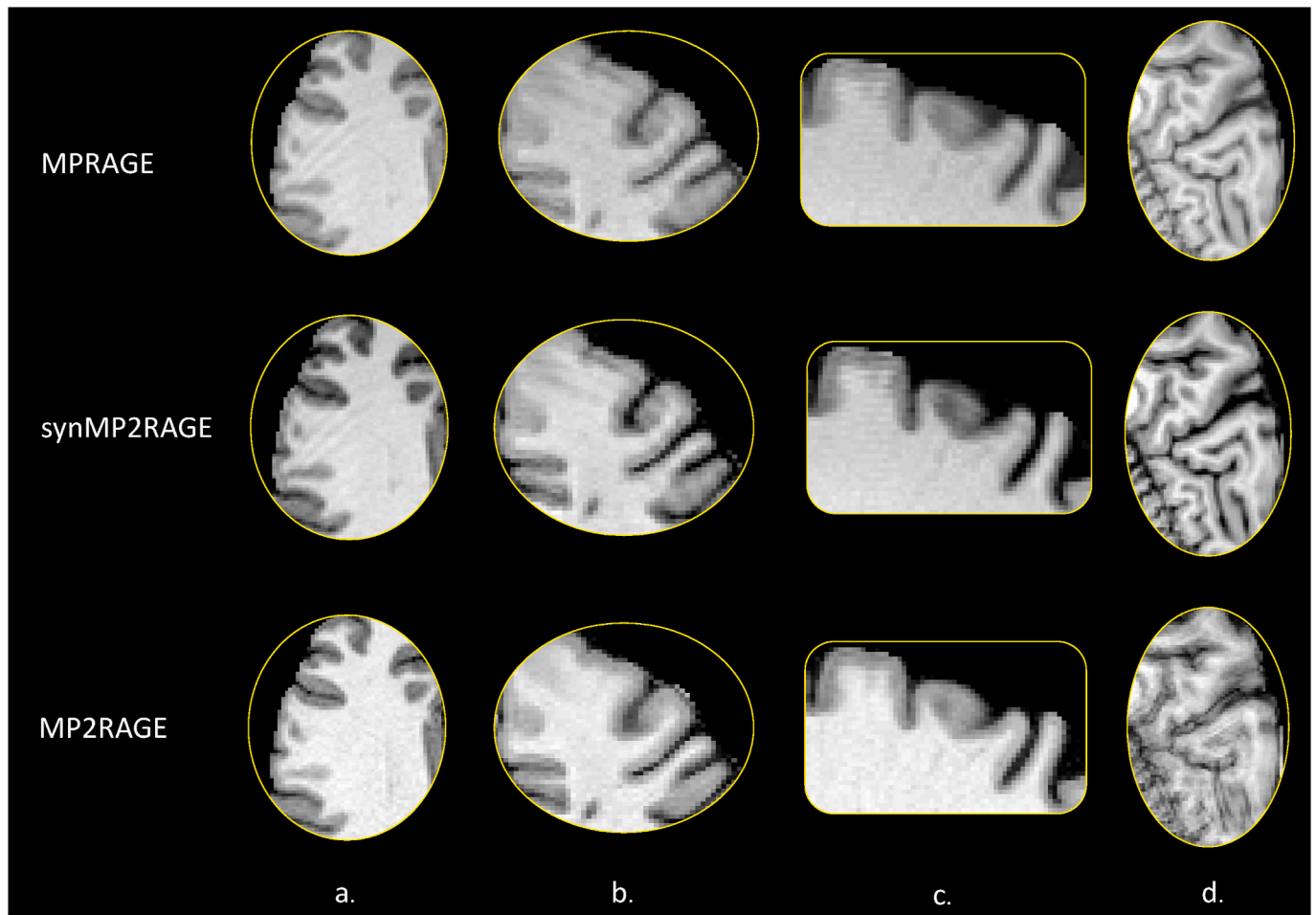


Fig. 4. Motion artefacts observed in the qualitative analysis. In a.,b.,c. different artefacts are present both in the MPRAGE and synMP2RAGE, but not in the MP2RAGE. In d. the MP2RAGE is affected by strong motion artefacts which are not observed in the MPRAGE and synMP2RAGE.

architecture is fully-convolutional, these inference input dimensions are arbitrary and can be chosen depending on the testing images. The generator is then tested separately with the transversal, coronal and sagittal images. The three output volumes are finally averaged to obtain the final synthetic MP2RAGE image. The inference process takes about 50 s per subject on a UNIX machine equipped with a GPU Tesla P-100 from NVIDIA.

3. Evaluation

3.1. Qualitative evaluation

In order to identify possible artificial artefacts introduced by the GAN, an image analysis expert, with more than 20 years of experience in brain MRI, carefully examined, slice-by-slice and in all three orthogonal planes, the 36 synMP2RAGE images of the testing dataset. In a second assessment, the artefacts found were then analyzed in comparison with the corresponding MPRAGE images.

3.2. Quantitative evaluation

In order to quantitatively compare the synMP2RAGE and the MPRAGE with the original MP2RAGE we computed three widely used reference-based similarity metrics [12,16]: peak signal-to-noise ratio (PSNR), normalized root mean square error (NRMSE) and mean structural similarity index (SSIM). The metrics were computed per subject, considering the skull-stripped images, and averaged across the entire

testing dataset.

3.3. Automatic segmentation

We propose to test two different automated segmentation approaches in order to objectively evaluate the benefits of analyzing synthetic MP2RAGE images compared to the commonly acquired MPRAGE. We considered two different methods:

- Atropos [20], an unsupervised approach for brain tissue segmentation distributed with ANTs. This is a Bayesian method that aims at solving the expectation-maximization algorithm modelling the class intensities with either parametric or non-parametric finite mixtures. We initialized the algorithm with k-means and selected 3 classes to be segmented: WM, GM and cerebrospinal fluid (CSF). The likelihood model chosen was a Gaussian and ran for 5 iterations. All other parameters were the default ones. Prior to running Atropos, all scans were skull-stripped and normalized for bias field inhomogeneities with ANTs.
- A CNN recently proposed for MS lesion segmentation [19]. This architecture is based on the 3D U-Net and was specifically adapted to detect both cortical and white matter lesions with high accuracy. We evaluated the CNN, setting all the default parameters, with a 3 folds cross-validation over the 36 testing cases. Four different combinations of MRI contrasts as input for the segmentation were considered: FLAIR-MPRAGE, FLAIR-MP2RAGE, FLAIR-DIR, and

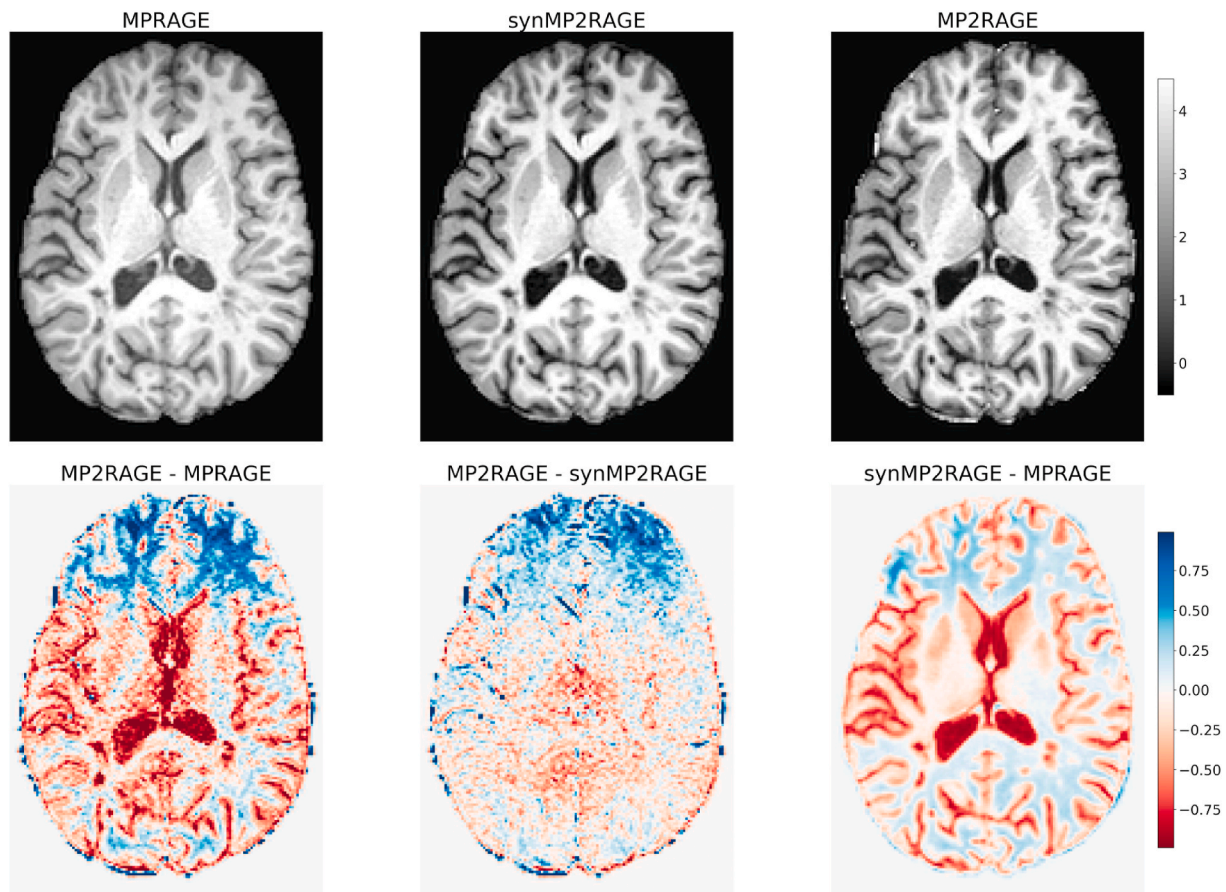


Fig. 5. First row: example of a 2D axial slice of MPRAGE, synMP2RAGE, and MP2RAGE. Second row: the three residuals between the images above. The intensity values are normalized with zero mean and unit variance.

FLAIR-synMP2RAGE. Prior to training, the second contrast was rigidly registered to the FLAIR space with ELASTIX25.

Automated segmentations were assessed computing the Dice coefficient and volume difference for the brain tissues and the detection rate, Dice coefficient, and volume difference for the cortical and white matter lesions.

3.4. Statistical analysis

For all metrics considered, the Wilcoxon signed-rank test was computed at the subject-wise level using the SciPy Python library [30]. Statistical differences were considered for p -values < 0.05 .

4. Results

The synthetic images inferred from the test set were evaluated both qualitatively and quantitatively. Qualitative results, comparing a slice of each of the three orthogonal planes of MPRAGE, MP2RAGE, and synMP2RAGE, are shown in Fig. 3. As can be observed, our generative approach synthesizes images that are consistent in the three planes and exhibit a visually evident increase in tissue contrast, with only a slight over-smoothing compared to the real MP2RAGE. In the qualitative evaluation, our image analysis expert judged the synMP2RAGE images of high quality, with good tissue contrast and low noise. The visual assessment revealed the presence of common MRI artefacts such as blurred areas, aliasing artefacts, and checkboard patterns in few slices of 24 out of 36 subjects. However, analyzing the corresponding MPRAGE images, it was verified that all artefacts were already present in the original acquisitions. In Fig. 4 we illustrate the different types of

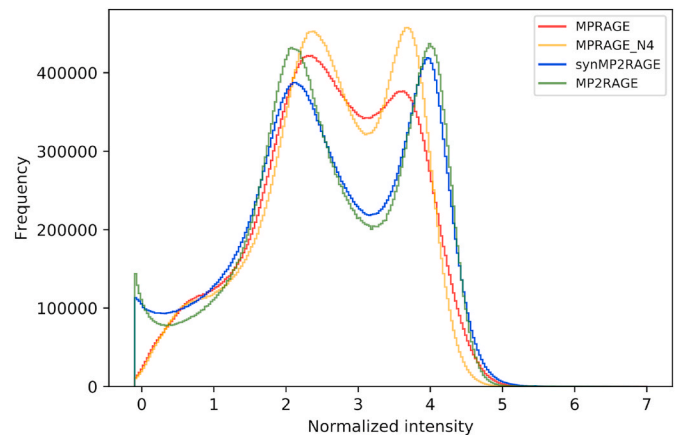


Fig. 6. Cumulative histogram over the 36 cases of the testing set comparing the real contrasts and the generated one. The intensity peak of the background just below zero is omitted for visualization purposes.

artefacts found. Interestingly, for one subject the MP2RAGE was affected by strong motion whereas the MPRAGE and generated synMP2RAGE seem fine. Overall, our expert concluded that the GAN primarily learns an intensity mapping, and no new artefacts are introduced or removed, except for the bias field which is much reduced or even removed.

Residual images are presented in Fig. 5. Compared to the MPRAGE, we can notice that synMP2RAGE improves both the lesion and tissue contrast, showing high residual values. Looking at the residual with MP2RAGE, however, we see that, as expected, the method is not perfect

Table 1

Quantitative metrics and their standard deviation obtained for MPRAGE and synMP2RAGE compared to the reference real MP2RAGE. For all metrics comparison p-values < 0.001. 1: the higher the better, ↓ the lower the better.

Contrast	PSNR 1	NRMSE ↓	SSIM 1
MPRAGE	29.49 ± 0.73	0.17 ± 0.01	0.97 ± 0.01
synMP2RAGE	31.39 ± 0.96	0.13 ± 0.01	0.98 ± 0.01

and between the prevailing random noise, some border voxels along the CSF have high residuals. Moreover, in Fig. 6 we report the cumulative histogram over the 36 testing cases for MPRAGE, MPRAGE with N4 bias field correction, MP2RAGE, and synMP2RAGE. It can be seen that the bias field correction contributes to increasing the prominence of the two peaks of GM and WM. By contrast, in the synMP2RAGE histogram the two peaks, besides having greater prominence, are also farther apart from each other compared to the MPRAGE. As a consequence, the synMP2RAGE histogram almost matches the one of the real MP2RAGE, confirming the added value of the synthesized images.

Results of the quantitative evaluation considering reference-based similarity metrics are reported in Table 1. We compare the metrics between synMP2RAGE and the ground truth MP2RAGE and those between

MPRAGE and the ground truth MP2RAGE. For all three metrics, synMP2RAGE outperforms the initial MPRAGE, achieving for the PSNR, NRMSE, and SSIM 31.39, 0.13, and 0.98, respectively. This shows that our method is transforming the MPRAGE input closer to the MP2RAGE.

Turning now to the automatic segmentation evaluation, the synMP2RAGE images were assessed in terms of both the lesion and tissue masks obtained. Visual examples of lesion and tissue segmentations for the different contrasts are shown in Fig. 7. Upon closer inspection of this figure, the improvements of the synMP2RAGE segmentations over those of the MPRAGE can be easily appreciated.

Regarding the automatic brain tissue segmentation results, Fig. 8 depicts the boxplots of the Dice coefficient for the three main brain tissue types (WM, GM, and CSF). For each tissue type, synMP2RAGE significantly outperformed the MPRAGE (all p-values < 0.0001), reaching in particular for the WM a median value of over 0.94.

Moving on now to consider the automatic MS lesion segmentation, we considered lesion-wise and voxel-wise metrics for WMLs and CLs. The top row of Fig. 9 shows the boxplots of lesion-wise true and false positives computed for each patient. No significant differences between any MRI contrast combination were found. The boxplots of the Dice coefficient and volume difference are presented in the bottom row of the same figure. For both metrics, the segmentation exploiting the synMP2RAGE significantly improves compared to both the MPRAGE and the DIR (p-values < 0.001), proving the added value of the synthesized images. Differences were not significant between synMP2RAGE and real

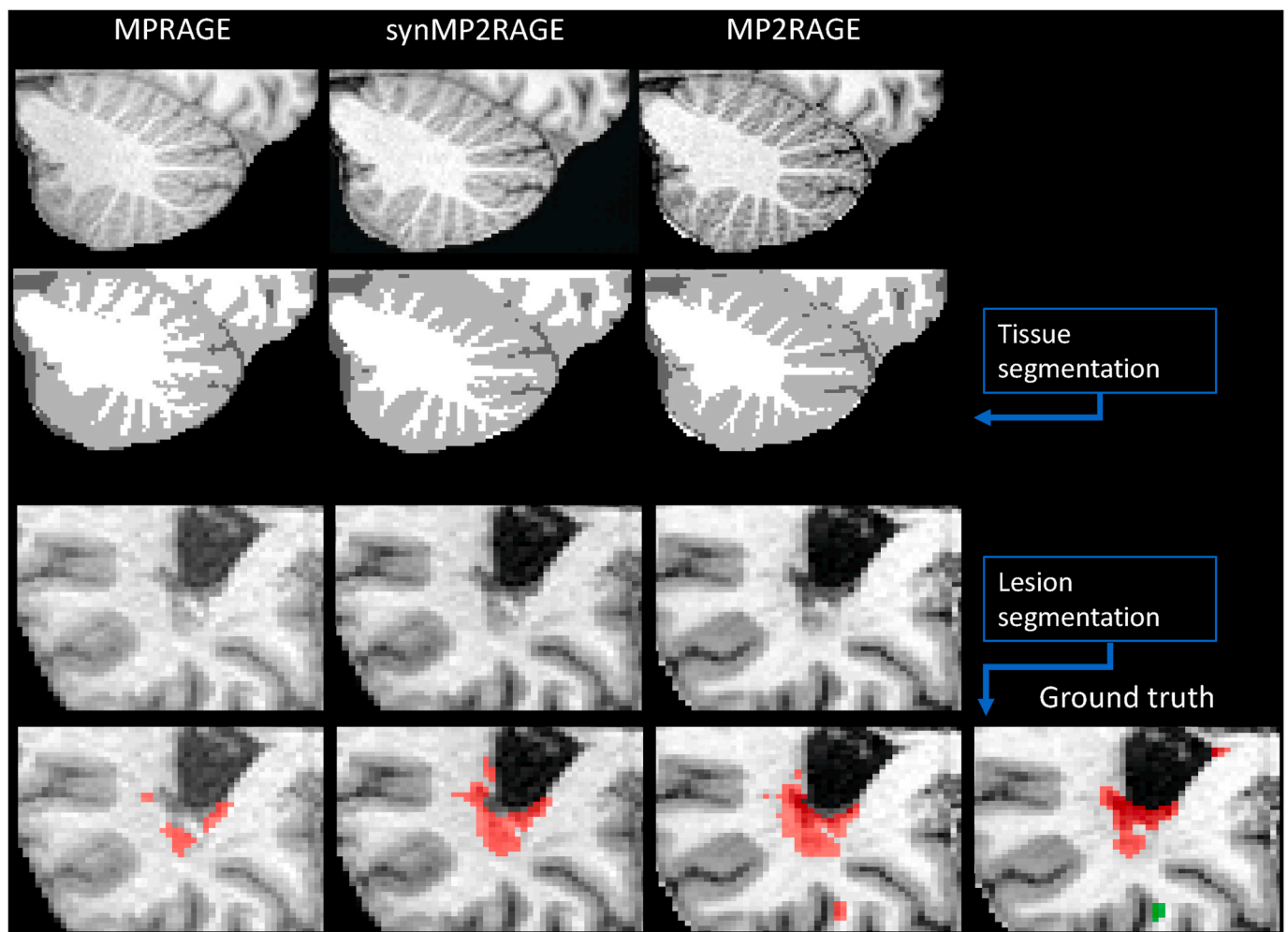


Fig. 7. Comparison of the tissue and lesion segmentation obtained with MPRAGE, synMP2RAGE, and MP2RAGE. In the first row: zoom-in slices of the three contrasts. In the second row the relative tissue segmentation: in white the WM, light gray the GM, and dark gray the CSF. In the last row, the lesion segmentation is compared to the ground truth: red for WMLs and green for CLs.

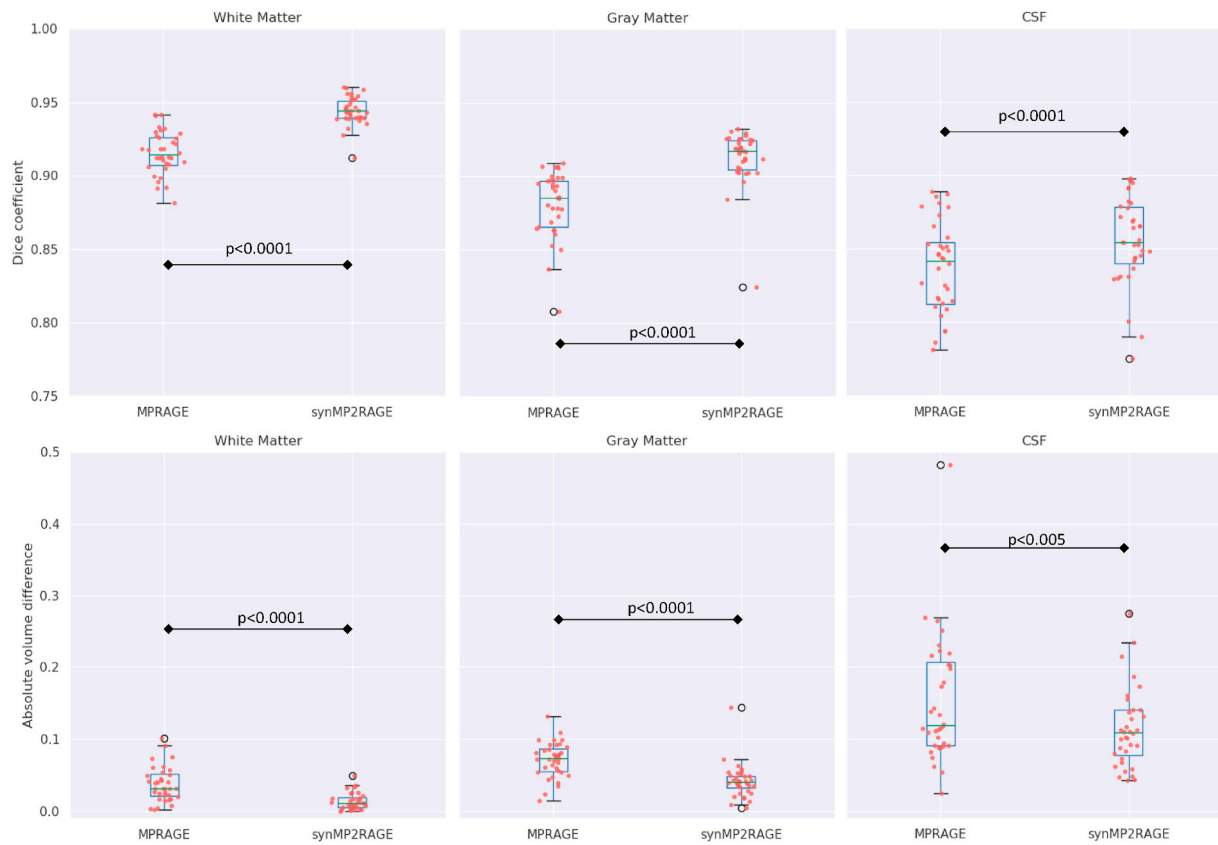


Fig. 8. Metrics of the automated brain tissue segmentation obtained using MPRAGE and synMP2RAGE and considering the MP2RAGE as ground truth. First row: Dice coefficient for WM, GM, and CSF. Second row: the absolute volume difference for the same tissue types.

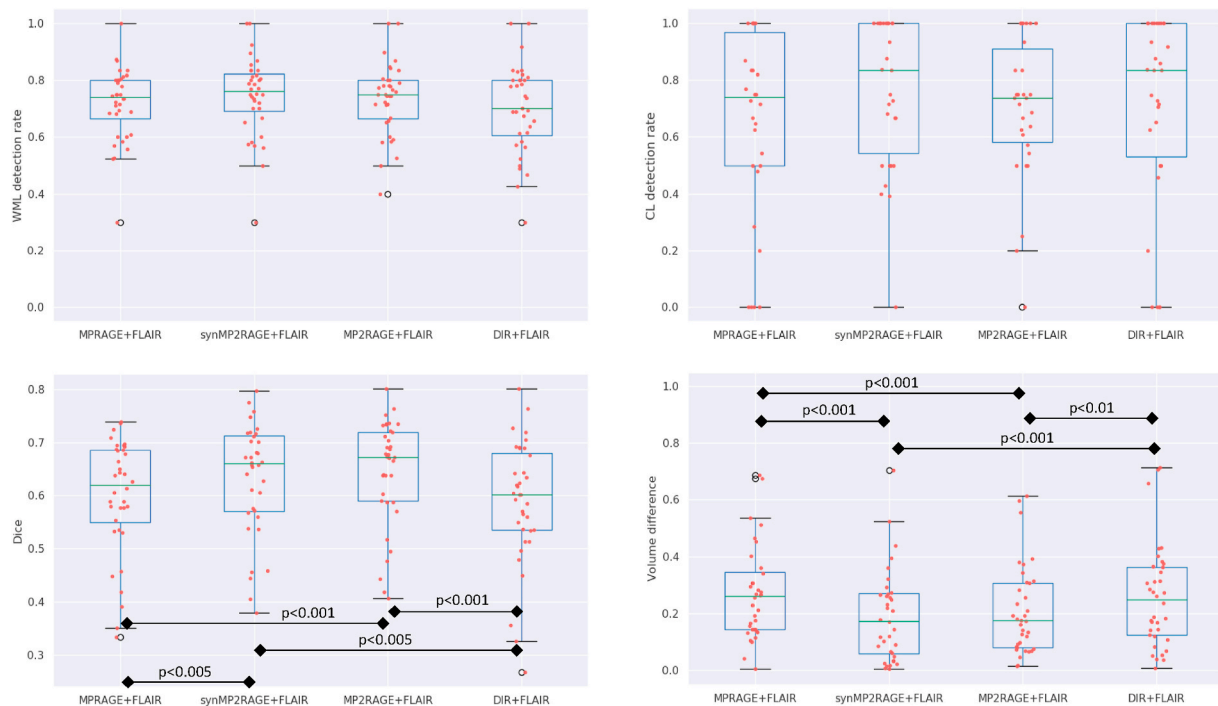


Fig. 9. First row: WML and CL detection rate for the combination of input MRI contrasts considered. Second row: Dice coefficient and absolute volume difference for the same MRI contrasts. All comparisons not shown are not significant.

Table 2

Cortical and white matter lesion detection rates computed lesion-wise over the entire testing cohort for the different contrasts, all considered together with FLAIR. The false positive rate is fixed to 30% in order to easily compare the detection rates.

	WML detection rate (n = 1641)	CL detection rate (n = 373)
MPRAGE	73%	71%
synMP2RAGE	73%	71%
MP2RAGE	75%	71%
DIR	71%	75%

MP2RAGE (p-values > 0.05). Finally, the cortical and WM lesion detection rates were computed lesion-wise over the 36 testing cases for the different contrasts are reported in Table 2 (see the Supplementary material for the metrics of unimodal FLAIR model). We can observe that once again in terms of detection rate there are very minimal differences.

5. Discussion

In this work we present a tailored generative adversarial network for translating MPRAGE images to realistic-looking MP2RAGE ones. The MP2RAGE is an extension of the MPRAGE MRI sequence which shows a higher brain tissue contrast and is helpful for depicting MS lesions. Regardless of these advantages and its limited acquisition time (about 8 min or even less if accelerated [31]), however, it is still not widely acquired in clinical routine MRI examinations. Our work aims at supporting MRI studies whenever MP2RAGE was not originally acquired. We evaluated the method on a test set of 36 MS patients. Qualitative results show that the generated synMP2RAGE is visually similar to the acquired MP2RAGE with only a slight over-smoothing (Fig. 3) and the GAN limits its task mainly to increase the tissue contrast (see the histogram in Fig. 6). Importantly, our generator did not introduce any artificial artefact compared to the acquired MPRAGE images, supporting our choice of using a combination of L1, adversarial and perceptual loss during training. A quantitative evaluation was then performed considering reference-based metrics, showing that synMP2RAGE outperformed MPRAGE on all metrics. Two different methods proposed in the literature were explored to evaluate the added value/applicability of the synthetic MP2RAGE images: automatic tissue and lesion segmentation in MS patients. In both cases, using the synMP2RAGE images offered significant improvements in terms of Dice overlap and volume difference (p-values < 0.05 in the patient-wise analysis) over using MPRAGE images. Importantly, in terms of lesion segmentation, synMP2RAGE does not significantly differ from using the ground truth MP2RAGE (p-values > 0.05).

There are several possible explanations supporting the fact that a GAN might be able to learn a mapping that translates MPRAGE image to MP2RAGE ones. For instance, a study has shown that while generating an exogenous MRI contrast obtained from a contrast agent might not be possible, synthesizing a missing contrast (such as T1 or T2) from others, is [32]. In our work, we aim at obtaining a variation of the MPRAGE contrast given as input to the generator. Both MPRAGE and MP2RAGE UNI are T1-weighted sequences with similar acquisition parameters (see Imaging subsection) and limited visual differences. Including a global skip connection in the GAN ensures that the correction produced is added to the input image, and therefore that the output does not greatly differ from the input. In this work, we do not consider synthesizing the MP2RAGE T1-map images, as these are slightly different visually and are currently not used by automated segmentation methods. However, if clinical studies will promote their usage, future work could focus on the generation of these quantitative images.

We believe that our method of translating MPRAGE acquisitions to synthetic MP2RAGE images has a high practical value for several reasons. Firstly, our evaluation answers the question in Hagiwara et al. [17] about whether a synthetic MRI contrast could also improve the

performance of an automated analysis tool: independently of the method considered, the synthetic MP2RAGE presented improved the output of both a CNN-based and a Bayesian segmentation approach. Secondly, as the MP2RAGE is now gradually being adopted for clinical use, generating its synthetic version for MRI studies where MP2RAGE was not originally acquired would allow homogenizing the datasets for retrospective analysis. Thirdly, experts might visually benefit from its increased tissue and lesion contrast as well.

Furthermore, to the best of our knowledge, we present the first comparison of conventional and specialized MRI sequences for the automatic segmentation of MS lesions using a CNN. Interestingly, no significant differences (p-values > 0.05) in the patient-wise nor in the lesion-wise analysis were found in terms of cortical and white matter lesion detection rate between MPRAGE, MP2RAGE, synMP2RAGE, and DIR. This is in contrast with previous works showing that specialized MRI contrasts improve the detection of CLs both visually and automatically [7,8]. We hypothesize that this is because of the intrinsic learning process of a CNN, particularly different from classical machine learning approaches previously applied to this task [8,33]. The evaluation set, however, is limited to 36 patients and it does not allow us to draw strong conclusions.

The generalisability of our results is subject to certain limitations. First, for all training and testing cases, the MPRAGE images were acquired with the same scanner and imaging protocol, including the same resolution. Therefore, the question of whether the proposed generative framework is able to learn a general mapping from different MPRAGE images to a common MP2RAGE remains open. Second, in this work we only examine MS patients in the early stages of the disease. The performance of our proposed method on patients with a high lesion load remains to be explored. Third, while our qualitative evaluation tends to indicate that no artefacts are introduced in the synMP2RAGE and that its usage supports automated segmentation tools, a radiological analysis of the generated image by experts would be needed in the context of evaluating the clinical value of synMP2RAGE.

In conclusion, we propose a GAN which successfully translates MPRAGE images to realistic-looking MP2RAGE ones of MS patients. An extensive evaluation of the synthesized images with automatic segmentation tools proved that the generated MP2RAGE significantly improves both the tissue and lesion segmentation. The framework is fast to run at inference time and publicly available, and can, therefore, be useful to MRI and clinical researchers for multiple tasks, even beyond neuroimaging studies of MS patients.

Acknowledgements

This project is supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie project TRABIT (agreement No 765148). We acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University of Lausanne (UNIL), École polytechnique fédérale de Lausanne (EPFL), University of Geneva (UNIGE) and Geneva University Hospitals (HUG). The work is also supported by the Leenaards and Jeantet Foundations. CG is supported by the Swiss National Science Foundation grant PP00P3-176984.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compbimed.2021.104297>.

References

- [1] M. Filippi, M.A. Rocca, O. Ciccarelli, et al., MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines, *Lancet Neurol.* 15 (3) (2016) 292–303, [https://doi.org/10.1016/S1474-4422\(15\)00393-2](https://doi.org/10.1016/S1474-4422(15)00393-2).

- [2] A.J. Thompson, B.L. Banwell, F. Barkhof, et al., Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria, *Lancet Neurol.* 17 (2) (2018) 162–173, [https://doi.org/10.1016/S1474-4422\(17\)30470-2](https://doi.org/10.1016/S1474-4422(17)30470-2).
- [3] J.P. Mugler, J.R. Brookeman, Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE), *Magn. Reson. Med.* 15 (1) (1990) 152–157, <https://doi.org/10.1002/mrm.1910150117>.
- [4] J.P. Marques, T. Kober, G. Krueger, W. van der Zwaag, P.-F. Van de Moortele, R. Gruetter, MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field, *Neuroimage* 49 (2) (2010) 1271–1281, <https://doi.org/10.1016/j.neuroimage.2009.10.002>.
- [5] G. Okubo, T. Okada, A. Yamamoto, et al., MP2RAGE for deep gray matter measurement of the brain: a comparative study with MPRAGE: MP2RAGE for Deep Gray Matter Measurement, *J. Magn. Reson. Imag.* 43 (1) (2016) 55–62, <https://doi.org/10.1002/jmri.24960>.
- [6] M.B. Cuadra, S. Gelin, A. Roche, Classical segmentation methods on novel MR imaging: a study of brain tissue segmentation of MP2RAGE Vs MPRAGE, Accessed May 14, 2020, <http://archive.ismrm.org/2013/0948.html>.
- [7] T. Kober, C. Granziera, D. Ribes, et al., MP2RAGE multiple sclerosis magnetic resonance imaging at 3 T, *Invest. Radiol.* 47 (6) (2012) 346–352, <https://doi.org/10.1097/RLI.0b013e31824600e9>.
- [8] M.J. Fartaria, G. Bonnier, A. Roche, et al., Automated detection of white matter and cortical lesions in early stages of multiple sclerosis: automated MS Lesion Segmentation, *J. Magn. Reson. Imag.* 43 (6) (2016) 1445–1454, <https://doi.org/10.1002/jmri.25095>.
- [9] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., Generative adversarial networks, *ArXiv14062661 Cs stat*, Published online June 10, 2014. Accessed May 5, 2020, <http://arxiv.org/abs/1406.2661>.
- [10] S. Kazemnia, C. Baur, A. Kuijper, et al., GANs for medical image analysis, *ArXiv180906222 Cs Stat*, Published online October 9, 2019. Accessed May 5, 2020, <http://arxiv.org/abs/1809.06222>.
- [11] P. Shende, M. Pawar, S. Kakde, A brief review on: MRI images reconstruction using GAN, in: 2019 International Conference on Communication and Signal Processing (ICCSP), 2019, <https://doi.org/10.1109/ICCSP.2019.8698083>, 0139–0142.
- [12] Y. Chen, A.G. Christodoulou, Z. Zhou, F. Shi, Y. Xie, D. Li, MRI super-resolution with GAN and 3D multi-level DenseNet: smaller, faster, and better, *ArXiv200301217 Cs Eess*, Published online March 6, <http://arxiv.org/abs/2003.01217>, 2020. Accessed May 5, 2020.
- [13] H.-C. Shin, N.A. Tenenholtz, J.K. Rogers, et al., Medical image synthesis for data augmentation and anonymization using generative adversarial networks, *ArXiv180710225 Cs Stat*, Published online September 13, 2018. Accessed May 5, 2020, <http://arxiv.org/abs/1807.10225>.
- [14] Y. Lei, J. Harms, T. Wang, et al., MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks, *Med Phys* (2019), <https://doi.org/10.1002/mp.13617> mp.13617.
- [15] M. Salem, S. Valverde, M. Cabezas, et al., Multiple sclerosis lesion synthesis in MRI using an encoder-decoder U-NET, *ArXiv190105733 Cs*, Published online January 17, 2019. Accessed May 5, <http://arxiv.org/abs/1901.05733>, 2020.
- [16] T. Finck, H. Li, L. Grundl, et al., Deep-learning generated synthetic double inversion recovery images improve multiple sclerosis lesion detection, *Invest. Radiol.* 55 (5) (2020) 318–323, <https://doi.org/10.1097/RLI.0000000000000640>.
- [17] A. Hagiwara, Y. Otsuka, M. Hori, et al., Improving the quality of synthetic FLAIR images with deep learning using a conditional generative adversarial network for pixel-by-pixel image translation, *Am. J. Neuroradiol.* 40 (2) (2019) 224–230, <https://doi.org/10.3174/ajnr.A5927>.
- [18] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-Image translation with conditional adversarial networks, *ArXiv161107004 Cs*, Published online November 26, 2018. Accessed May 5, <http://arxiv.org/abs/1611.07004>, 2020.
- [19] F. La Rosa, A. Abdulkadir, M.J. Fartaria, et al., Multiple sclerosis cortical and WM lesion segmentation at 3T MRI: a deep learning method based on FLAIR and MP2RAGE, *Neuroimage Clin* (2020) 102335, <https://doi.org/10.1016/j.nicl.2020.102335>.
- [20] B.B. Avants, N.J. Tustison, J. Wu, P.A. Cook, J.C. Gee, An open source multivariate framework for n-tissue segmentation with evaluation on public data, *Neuroinformatics* 9 (4) (2011) 381–400, <https://doi.org/10.1007/s12021-011-9109-y>.
- [21] P.A. Yushkevich, J. Piven, H.C. Hazlett, et al., User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability, *Neuroimage* 31 (3) (2006) 1116–1128, <https://doi.org/10.1016/j.neuroimage.2006.01.015>.
- [22] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, J. Matas, DeblurGAN: blind motion deblurring using conditional adversarial networks, *ArXiv171107064 Cs*, Published online April 3, 2018. Accessed May 7, <http://arxiv.org/abs/1711.07064>, 2020.
- [23] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, *ArXiv151106434 Cs*, Published online January 7, 2016. Accessed June 3, <http://arxiv.org/abs/1511.06434>, 2020.
- [24] G. Yang, S. Yu, H. Dong, et al., DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction, *IEEE Trans. Med. Imag.* 37 (6) (2018) 1310–1321, <https://doi.org/10.1109/TMI.2017.2785879>.
- [25] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, *ArXiv160308155 Cs*, Published online March 26, <http://arxiv.org/abs/1603.08155>, 2016. Accessed May 28, 2020.
- [26] M.S.M. Sajjadi, B. Schölkopf, M. Hirsch, EnhanceNet: single image super-resolution through automated texture synthesis, *ArXiv161207919 Cs Stat*, Published online July 30, <http://arxiv.org/abs/1612.07919>, 2017. Accessed May 28, 2020.
- [27] S. Klein, M. Staring, K. Murphy, M.A. Viergever, J. Pluim, Elastix: a toolbox for intensity-based medical image registration, *IEEE Trans. Med. Imag.* 29 (1) (2010) 196–205, <https://doi.org/10.1109/TMI.2009.2035616>.
- [28] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, *ArXiv1412.6980 Cs*, Published online January 29, <http://arxiv.org/abs/1412.6980>, 2017. Accessed July 27, 2020.
- [29] Abadi M, Barham P, Chen J, et al. TensorFlow: A System for Large-Scale Machine Learning. :21.
- [30] SciPy 1.0 Contributors, P. Virtanen, R. Gommers, et al., SciPy 1.0: fundamental algorithms for scientific computing in Python, *Nat. Methods* 17 (3) (2020) 261–272, <https://doi.org/10.1038/s41592-019-0686-2>.
- [31] E. Mussard, T. Hilbert, C. Forman, R. Meuli, J. Thiran, T. Kober, Accelerated MP2RAGE imaging using Cartesian phyllotaxis readout and compressed sensing reconstruction, *Magn. Reson. Med.* (2020) mrm.28244, <https://doi.org/10.1002/mrm.28244>. Published online March 16.
- [32] D. Lee, W.-J. Moon, J.C. Ye, Assessing the importance of magnetic resonance contrasts using collaborative generative adversarial networks, *Nat Mach Intell* 2 (1) (2020) 34–42, <https://doi.org/10.1038/s42256-019-0137-x>.
- [33] M.J. Fartaria, A. Roche, R. Meuli, C. Granziera, T. Kober, M. Bach Cuadra, Segmentation of cortical and subcortical multiple sclerosis lesions based on constrained partial volume modeling, in: M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D.L. Collins, S. Duchesne (Eds.), *Medical Image Computing And Computer Assisted Intervention – MICCAI 2017*. Vol 10435. Lecture Notes in Computer Science, Springer International Publishing, 2017, pp. 142–149, https://doi.org/10.1007/978-3-319-66179-7_17.