Deep-Learning Generated Synthetic Double Inversion Recovery Images Improve Multiple Sclerosis Lesion Detection

Abstract

Objectives: To implement a Deep-Learning tool to produce synthetic double-inversion recovery (synthDIR) images and compare their diagnostic performance to conventional sequences in patients with multiple sclerosis (MS).

Materials and Methods: For this retrospective analysis, 100 MS patients [65 female, 37 (22-68) years] were randomly selected from a prospective observational cohort between 2014 and 2016. In a subset of 50 patients, an artificial neural network *(DiamondGAN)* was trained to generate a synthetic DIR (synthDIR) from standard acquisitions (T1, T2 and FLAIR). With the resulting network, synthDIR was generated for the remaining 50 subjects. These images as well as conventionally acquired DIR (trueDIR) and FLAIR images were assessed for MS lesions by two independent readers, blinded to the source of the DIR image. Lesion counts in the different modalities were compared using a Wilcoxon signed-rank test and inter-rater analysis was performed. Contrast-to-noise ratios (CNR) were compared for objective image quality.

Results: Utilization of synthDIR allowed to detect significantly more lesions compared to the use of FLAIR images (31.4 ± 20.7 vs. 22.8 ± 12.7, $p<0.001$). This improvement was mainly attributable to an improved depiction of juxtacortical lesions (12.3 ± 10.8 vs. 7.2 ± 5.6,

p<0.001). Inter-rater reliability was excellent in FLAIR 0.92 (95% CI 0.85; 0.95), synthDIR 0.93 (95% CI 0.87; 0.96) and trueDIR 0.95 (95% CI 0.85; 0.98).

CNR in synthDIR exceeded that of FLAIR (22.0 ± 6.4 vs. 16.7 ± 3.6, p=0.009), no significant difference was seen in comparison to trueDIR (22.0 ± 6.4 vs. 22.4 ± 7.9, p=0.87).

Conclusion: Computationally generated DIR images improve lesion depiction compared to the use of standard modalities. This method demonstrates how artificial intelligence can help improving imaging in specific pathologies.

Abbreviations

MS: Multiple Sclerosis

FLAIR: Fluid-Attenuated Inversion Recovery

DIR: Double Inversion Recovery

GAN: Generative Adversarial Network

trueDIR: physically acquired DIR

synthDIR: synthetically generated DIR

CNR: Contrast-to-Noise ratio

NAWM: Normal Appearing White Matter

ICC: Intraclass Correlation Coefficient

EDSS: Expanded Disability Status Scale

RRMS: Relapsing-Remitting Multiple Sclerosis

AI: Artificial Intelligence

Introduction

Multiple sclerosis (MS) is the most common non-traumatic cause of disability in young adults with accelerating rates of new diagnoses over the past decades (1). Recently, data has accumulated demonstrating neuronal loss and early-onset atrophic brain changes in patients suffering from MS (2, 3). As the regenerative potential of brain tissue is per-se limited and decreases with age, it is important to attenuate disease activity as early as possible in order to avoid non-repairable damage. Both, delayed initial diagnosis as well as an undetected progression under treatment can negatively impact the prognosis of patients (4, 5). MRI has proven to be a valuable tool in detecting and evaluating progression of MS-related brain lesions and is tightly integrated into current guidelines (6). While fluid-attenuated inversion recovery (FLAIR) is arguably the most widespread sequence used to assess brain lesions, technical advances have led to the development of double inversion-recovery (DIR) acquisition with a higher sensitivity for lesion detection than conventional or fluid-attenuated T2-weighted acquisitions (7). As it has recently become clear that (juxta)cortical plaques play an important role in MS and appear in early stages of the disease, the particular strength of DIR to detect these kinds of lesions is becoming more and more relevant from a clinical point of view (8-11). Beyond the better conspicuity of (juxta)cortical lesions, DIR facilitates automated detection of lesions compared to standard sequences and thus seems better suited for segmentation tasks (12, 13). However, the time-consuming acquisition of DIR has hindered the widespread implementation of this technique in daily routine. In view of this limitation, synthesizing DIR from routinely acquired MR sequences with the aid of deep-learning tools, such as recently proposed generative adversarial networks (GAN) seems promising (14). Reliable image synthesis with the here-proposed *DiamondGAN* (generative

adversarial network with a diamond-shaped topology) has already been validated from a technical perspective (15).

An important aspect of *DiamondGAN* is its ability to augment image information via multi-to-one mapping, i.e. to use several input sequences (in this case T1, T2 and FLAIR) to generate one output modality (in this case DIR). This enables *DiamondGAN* to learn synergistic combinations of image information from multiple input sequences and thus improves the output. In a prior work describing the technical background of *DiamondGAN*, we have shown DIR images of healthy controls, either acquired or synthesized, to 14 neuroradiologists and asked for their visual evaluation of both modalities (15). Blinded to the origin of these images, they were unable to differentiate if the shown modality was acquired (trueDIR) or synthetic (synthDIR).

Whether GANs only create realistic looking images or whether these images indeed carry a surplus of clinically relevant information is however yet to be shown. We therefore investigated the hypothesis that synthDIR augments background image information from FLAIR, T1 and T2 and thus improves lesion depiction in MS patients.

Materials and Methods

Scan Acquisition

This study design was approved by the local institutional review board and informed consent was obtained from all patients. MR datasets from 100 randomly selected patients with diagnosed MS were retrospectively collected from a prospectively collected observation cohort from 08.2014 to 03.2016 and included T1- (2:25 minutes acquisition time), T2- (3:17 minutes acquisition time), FLAIR (3:55 minutes acquisition time) and DIR (6:31 minutes acquisition time) sequences. Scans were performed on a 3T scanner (Philips Achieva 3.0T, Philips Healthcare, Eindhoven, Netherlands). Identical parameters in all patients were chosen for T1 (repetition time= 9ms, echo time= 4.0ms, flip angle= 8°, acquired in the sagittal plane with isotropic voxel size of $1mm^3$), T2 (repetition time= 4000ms, echo time= 35.0ms, flip angle= 90°, acquired in the sagittal plane with isotropic voxel size of $1mm^3$), FLAIR (repetition time= 10000ms, echo time= 140.0ms, inversion time= 2750ms, flip angle= 90°, acquired in the sagittal plane with isotropic voxel size of $1mm^3$) and DIR (repetition time= 5500ms, echo time= 321.6ms, inversion time= 2550ms and 2990ms, flip angle= 90°, acquired in the sagittal plane with isotropic voxel size of $1mm^3$).

DIR synthesis with GAN

The basic principle of *DiamondGAN* is to synthesize a target MRI modality T given a set of n input modalities $X = \{x_i | i = 1, \dots, n\}$. The goal of the synthesis task is to learn a modality

generator *G* such that: $G(X) = T$. The *DiamondGAN* pipeline contains two networks: a generator *G* and a discriminator *D*, built on conventional GAN techniques (16). G contains two generators G_1 and G_2 which simultaneously learn the mappings from $X \rightarrow T$ and $T \rightarrow X$ respectively. *D* consists of two discriminators D_1 and D_2. D_1 discriminates the real images from *source* domain and synthetic images from G_1 while D_2 discriminates the real images from *target* domain and synthetic images from G_2. In this adversarial learning process, the four networks are simultaneously optimized to generate high-quality images. *G* and *D* are variants of convolutional neural networks and optimized with multiple loss functions in an end-to-end fashion, as explained before (15). One helpful property of publicly-available *DiamondGAN (https://github.com/hongweilibran/DiamondGAN/)* is that it does not require the input and output to be strictly spatially aligned, by mapping the inputs to latent spaces and optimizing the generator networks by a cycle-consistency loss function as explained in (15). Conventional regression approaches require the input and output to be strictly spatially aligned. However, in practice, registration methods cannot guarantee such pixel-to-pixel alignment properly between the input and output image spaces (17). A schematic overview illustrating the architecture of *DiamondGAN* is given in figure 1.

We hypothesize that a large portion of information in an individual MR sequence is also contained (albeit possibly hidden) in other sequences and that DIR can be reconstructed given the combination of FLAIR, T1 and T2. The 3D volumes of these sequences are parsed into 2D axial slices. The concatenation of FLAIR, T1 and T2 axial slices is fed to train the network while synthDIR slices are the output. Technically, the network does not necessarily require 3D acquisitions for the sequences, but the acquisition should be consistent (either 2D or 3D) among the input modalities. Fifty MS patients with complete sequences are used to train the

model. The generator model contains 7,218,987 parameters and is trained for 12 hours with a high-end graphic card (NVIDIA Titan V, Santa Clara, USA). In the inference stage, we take FLAIR, T1 and T2 slices as the input and synthesize the DIR slices. Then the DIR slices are normalized by histogram matching and spatially concatenated into 3D volumes. Since the axial slices on top or bottom do not contain brain structures, we synthesize the middle slices with brain structure after empirically setting a starting and ending threshold. Exemplary sets of FLAIR, synthDIR and trueDIR with their respective lesion segmentations are given in figure 2.

Expert readings

In accordance with the current imaging criteria to diagnose MS, we did not distinguish between juxtacortical and cortical lesions (18). For simplicity, we will therefore refer to both types of lesions as juxtacortical lesions.

All 150 data sets (50 test patients with FLAIR, trueDIR and synthDIR) were visually assessed by two neuroradiologists (both with 3 years of experience) for the number of juxtacortical, periventricular, infratentorial and subcortical white matter lesions, having a minimum diameter of 3mm in any direction. Manual lesion count was done independently using an open-source 3D image analysis tool (19) and the order of investigated modalities was randomly altered to prevent a learning effect. The readers were blinded for the nature of DIR modalities (trueDIR or synthDIR) and had no clinical background information on patients except for the fact that prior diagnosis of MS had been made. Each lesion in synthDIR was retrospectively validated in trueDIR by one rater to exclude that false-positive lesions had been generated during the synthetization process.

Lesion contrast

To assess image quality of the modalities, we calculated the contrast-to-noise ratio (CNR) for a randomly selected subset of 15 patients: in each patient, 1 or 2 representative lesions were manually segmented on the T2 image (to avoid bias) and equally sized regions of interest were placed in the contralateral normal-appearing white matter (NAWM) using open-source 3D image analysis tool (19). From the coregistered modalities, CNR was calculated for FLAIR, synthDIR and trueDIR as:

$$CNR = \frac{Mean_{SignalLesions} - Mean_{SignalNAWM}}{SD_{NAWM}}$$

Statistical analysis

Same counts of lesions in FLAIR (that was used as input modality) and synthDIR was defined as null hypothesis. The normality of distribution was violated as tested by the D'Agostino-Pearson test. Lesion counts from the three investigated modalities were compared with a Wilcoxon signed-rank test, CNR between the investigated modalities was compared with a paired Student's t-test.

Inter-rater reliability was assessed with the intraclass correlation coefficient (ICC) (use of single measurements for absolute agreement in a two-way random model).

Statistical computations were performed with software (SPSS Statistics for Windows, version 25.0; IBM, Armonk, NY). $P < 0.05$ was considered statistically significant.

Results

Patient characteristics for the training and test set are given in table 1. No significant differences in clinical parameters were observed.

Inter-rater reliability was excellent with intraclass correlation coefficients (ICC) between both raters ranging from 0.92 (95% CI: 0.85; 0.95) for FLAIR to comparable levels of 0.93 (95% CI: 0.87; 0.96) and 0.95 (95% CI: 0.85; 0.98) for synthDIR and trueDIR, respectively. ICCs between both readers as a function of lesion localization are given in table 2.

Quantitative assessment of image contrast (lesion vs. NAWM) in a subset of 15 randomly selected patients revealed a significantly better CNR for synthDIR compared to FLAIR (22.0 ± 6.4 vs. 16.7 ± 3.6, p=0.009), matching the CNR of trueDIR (22.0 ± 6.4 vs. 22.4 ± 7.9, p=0.87) (figure 3).

The mean count of MS-specific lesions (juxtacortical + periventricular + infratentorial) per patient was significantly higher in synthDIR than FLAIR (31.4 ± 20.7 vs. 22.8 ± 12.7, p<0.001) (location-dependent lesion counts for both readers are given in table 3, counts of MS-specific lesions are further shown in figure 4). Worth mentioning is the fact that the improved performance of synthDIR compared to FLAIR could be primarily attributed to better detection of juxtacortical lesions (12.3 ± 10.8 vs. 7.2 ± 5.6, p<0.001). Both observations held true when analysing lesion counts from the second reader (table 3). Consequently, a shift in the proportion of juxtacortical and subcortical lesions was noted as the relative share of juxtacortical lesions from the total count increased from 16.9% in FLAIR

to 23.3% in synthDIR ($p<0.001$) while the share of subcortical lesions decreased analogously from 41.2% in FLAIR to 37.3% in synthDIR ($p<0.001$).

Although a location-dependent heterogeneity could be noted, a tendency for improved lesion detection in synthDIR compared to FLAIR held true irrespective if lesions were juxtacortical, periventricular, subcortical or infratentorial (counts for both readers are given in table 3). Physically acquired trueDIR trumped both, FLAIR (36.1 ± 24.3 vs. 22.8 ± 12.7, $p<0.001$) and synthDIR (36.1 ± 24.3 vs. 31.4 ± 20.7, $p<0.001$) in depicting MS-specific lesions. Retrospective visual cross-validation showed that there were no lesions in synthDIR that could not be detected in trueDIR, hence the possibility of artificial lesion generation by the GAN could be excluded.

Discussion

We hypothesized that training a GAN with sequences routinely acquired in MS imaging allows to create synthetic, high lesion-to-background contrast DIR images. These generated images enabled the readers to find significantly more lesions compared to the standard sequence FLAIR.

Further supporting the ability of *DiamondGAN* to synergistically combine input image information to create its output is the fact that the increase in lesions found in the synthetic DIR compared to FLAIR was mostly driven by juxtacortical lesions. The DIR acquisition is best known for its ability to detect this type of lesion, as it has become increasingly clear that juxtacortical lesions play an important role for diagnosis and prognosis of MS patients (20-22). By combining information from FLAIR, T1 and T2 images, *DiamondGAN* is able to replicate this strength of DIR images with potentially profound ramifications as initial diagnosis of MS strongly relies on the robust differentiation between MS-specific (i.e. juxtacortical) non-MS-specific (i.e. subcortical) lesions .

Beyond the methodological innovation of using complimentary aspects of different acquisitions (T1, T2, FLAIR) to generate output (synthDIR), such pronounced differences in lesion detection can have clinical implications for the monitoring of a disease that is tightly linked to the dynamics of inflammatory lesions. Recent studies have highlighted the potential of DIR in monitoring MS progression and stress the urgency to use this acquisition more broadly and prospectively wean our dependency on gadolinium scans (22). In light of the time-consuming physical acquisition of DIR, artificial intelligence (AI) could be the key to facilitate its wider implementation in MS imaging. Even as the physically acquired DIR still

outperformed the synthetic DIR it needs to be said that the information content in synthDIR was nevertheless superior to that of input FLAIR and can potentially be improved with future improvements of GAN training and input combinations..

Beyond the subjective lesion analysis, quantification of CNR confirms the data augmentation taking place within *DiamondGAN*, a publicly available AI-tool, and provides proof for the similitude between synthDIR and trueDIR in depicting white-matter lesions.

While ongoing scientific advances, be it the improvement of image resolution, use of experimental methods such as MR-fingerprinting or inclusion of spectroscopy findings, pave the way for improving diagnostics in MS, the utilization of AI in chronic-inflammatory brain disease has in the vast majority of cases been restricted to quantitative analyses such as lesion segmentation (23-26). The here-presented study provides a different approach as it does not focus on interpreting available images but explores the prospect of augmenting intrinsic, yet not necessarily visual information within an MR dataset. Similar approaches may further open the door for data homogenization through synthetization of a ''standardized'' MR dataset from heterogenous input data or generation of artificial imaging sets to feed Deep Learning algorithms.

A further advantage of GANs is their applicability to existing data. This is in contrast to technical developments such as new sequences or hardware, which can only be applied prospectively and offers a unique chance to retrospectively validate findings that might become apparent after implementation of a more sensitive imaging protocol (27).

One general limitation of this study is the relatively small sample size of only 100 patients. Further, the single-centre setting with all scans originating from one MR scanner, rendering statements about the generalizability of our findings impossible. Using GANs for data homogenization by synthesizing a standardized input dataset irrespective of the source data

is however one potential remedy worth exploring. Additionally, the input acquisitions used in our study (T1, T2 and FLAIR) all depict MS lesions, thus a large portion of information from the individual acquisition is redundant. Therefore, it remains to be studied whether satisfying synthetic images can be derived already from a subset of these acquisitions. Moreover, the value of alternative input acquisitions than the ones used in the present study has yet to be investigated.

In summary, we have demonstrated the ability of artificial neural networks to create high contrast images from standard input, thereby significantly improving lesion detection in MS patients. Future studies investigating generalizability and optimal sequence combinations for image synthesis seem warranted.

Acknowledgments

MM, MB and BW are supported by a DFG grant within the Priority Programme Radiomics.

References

1. Collaborators GBDMS. Global, regional, and national burden of multiple sclerosis 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 2019;18(3):269-85.
2. Calabrese M, Atzori M, Bernardi V, et al. Cortical atrophy is relevant in multiple sclerosis at clinical onset. J Neurol. 2007;254(9):1212-20.
3. Carassiti D, Altmann DR, Petrova N, et al. Neuronal loss, demyelination and volume change in the multiple sclerosis neocortex. Neuropathol Appl Neurobiol. 2018;44(4):377-90.
4. Rudick RA, Polman CH. Current approaches to the identification and management of breakthrough disease in patients with multiple sclerosis. Lancet Neurol. 2009;8(6):545-59.
5. Fernandez O, Fernandez V, Arbizu T, et al. Characteristics of multiple sclerosis at onset and delay of diagnosis and treatment in Spain (the Novo Study). J Neurol. 2010;257(9):1500-7.
6. Filippi M, Rocca MA, Ciccarelli O, et al. MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines. Lancet Neurol. 2016;15(3):292-303.
7. Wattjes MP, Lutterbey GG, Gieseke J, et al. Double inversion recovery brain imaging at 3T: diagnostic value in the detection of multiple sclerosis lesions. AJNR Am J Neuroradiol. 2007;28(1):54-9.
8. Popescu BF, Bunyan RF, Parisi JE, et al. A case of multiple sclerosis presenting with inflammatory cortical demyelination. Neurology. 2011;76(20):1705-10.
9. Redpath TW, Smith FW. Technical note: use of a double inversion recovery pulse sequence to image selectively grey or white brain matter. Br J Radiol. 1994;67(804):1258-63.
10. Seewann A, Kooi EJ, Roosendaal SD, et al. Postmortem verification of MS cortical lesion detection with 3D DIR. Neurology. 2012;78(5):302-8.
11. Calabrese M, Filippi M, Gallo P. Cortical lesions in multiple sclerosis. Nat Rev Neurol. 2010;6(8):438-44.
12. Geurts JJ, Pouwels PJ, Uitdehaag BM, et al. Intracortical lesions in multiple sclerosis: improved detection with 3D double inversion-recovery MR imaging. Radiology. 2005;236(1):254-60.
13. Fartaria MJ, Bonnier G, Roche A, et al. Automated detection of white matter and cortical lesions in early stages of multiple sclerosis. J Magn Reson Imaging. 2016;43(6):1445-54.
14. Nie D, Trullo R, Lian J, et al. Medical Image Synthesis with Deep Convolutional Adversarial Networks. IEEE Trans Biomed Eng. 2018;65(12):2720-30.
15. Hongwei Li JCP, Anjany Sekuboyina, Florian Kofler, Jianguo Zhang, Jan S. Kirschke, Benedikt Wiestler, Bjoern Menze. DiamondGAN: Unified Multi-Modal Generative Adversarial Networks for MRI Sequences Synthesis. Shen D. et al. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. MICCAI 2019. Lecture Notes in Computer Science, vol 11767. Springer, Cham. 2019.

16. Goodfellow I P-AJ, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. Adv Neural Inf Process Syst. 2014.
17. Alexander DC, Zikic D, Ghosh A, et al. Image quality transfer and applications in diffusion MRI. Neuroimage. 2017;152:283-98.
18. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. 2018;17(2):162-73.
19. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage. 2006;31(3):1116-28.
20. Moriarty DM, Blackshaw AJ, Talbot PR, et al. Memory dysfunction in multiple sclerosis corresponds to juxtacortical lesion load on fast fluid-attenuated inversion-recovery MR images. AJNR Am J Neuroradiol. 1999;20(10):1956-62.
21. Stadelmann C, Albert M, Wegner C, Bruck W. Cortical pathology in multiple sclerosis. Curr Opin Neurol. 2008;21(3):229-34.
22. Eichinger P, Schon S, Pongratz V, et al. Accuracy of Unenhanced MRI in the Detection of New Brain Lesions in Multiple Sclerosis. Radiology. 2019;291(2):429-35.
23. Llufriu S, Kornak J, Ratiney H, et al. Magnetic resonance spectroscopy markers of disease progression in multiple sclerosis. JAMA Neurol. 2014;71(7):840-7.
24. Granberg T, Uppman M, Hashim F, et al. Clinical Feasibility of Synthetic MRI in Multiple Sclerosis: A Diagnostic and Volumetric Validation Study. AJNR Am J Neuroradiol. 2016;37(6):1023-9.
25. Commowick O, Istace A, Kain M, et al. Objective Evaluation of Multiple Sclerosis Lesion Segmentation using a Data Management and Processing Infrastructure. Sci Rep. 2018;8(1):13650.
26. Kazuhiro K, Werner RA, Toriumi F, et al. Generative Adversarial Networks for the Creation of Realistic Artificial Brain Magnetic Resonance Images. Tomography. 2018;4(4):159-63.
27. Eichinger P, Hock A, Schon S, et al. Acceleration of Double Inversion Recovery Sequences in Multiple Sclerosis With Compressed Sensing. Invest Radiol. 2019;54(6):319-24.

Tables

	Training set	Testing set	p
n	50	50	1
Age	38.5 ± 9.2	36.4 11.0	0.30
Sex	18 (36)	17 (34)	0.84
% RRMS	45 (90%)	46 (92%)	0.73
Disease duration	5.7 ± 4.6	5.8 ± 4.0	0.91
EDSS	1.5 (1.0; 2.5)	1.5 (0; 2.0)	0.61

Table 1. Patient characteristics. Given are key clinical parameters for the training and testing set of the included 100 patients. RRMS: relapsing-remitting multiple sclerosis. EDSS: Expanded Disability Status Scale

	Periventricular			Juxtacortical		
	FLAIR	synthDIR	trueDIR	FLAIR	synthDIR	trueDIR
R1 counts	13.6 ± 9.35	16.7 ± 12.8	18.8 ± 14.3	7.2 ± 5.55	12.3 ± 10.8	14.7 ± 13.1
R2 counts	10.7 ± 8.36	13.0 ± 12.7	17.8 ± 15.1	3.7 ± 3.45	7.4 ± 9.6	11.1 ± 13.1
ICC	0.93 (0.88; 0.96)	0.95 (0.91; 0.97)	0.96 (0.93; 0.98)	0.82 (0.68; 0.90)	0.90 (0.82; 0.94)	0.95 (0.91; 0.97)
	Infratentorial			Subcortical		
	FLAIR	synthDIR	trueDIR	FLAIR	synthDIR	trueDIR
R1 counts	2.0 ± 1.68	2.4 ± 2.2	2.6 ± 2.2	12.1 ± 10.2	15.1 ± 10.8	15.8 ± 11.7
R2 counts	0.9 ± 1.26	1.3 ± 2.2	1.4 ± 1.6	14.6 ± 11.3	16.5 ± 12.2	18.8 ± 13.2
ICC	0.48 (0.07; 0.70)	0.87 (0.78; 0.93)	0.81 (0.66; 0.89)	0.92 (0.87; 0.96)	0.89 (0.81; 0.94)	0.91 (0.83; 0.95)

Table 2 Lesion count (mean ± standard deviation) as well as intraclass correlation coefficients (ICC with 95% CI) for both readers and all acquistions/locations. R = reader.

	MS-specific lesions (PV+JC+IT)	p	Periventicular lesions	p	Juxtacortical lesions	p	Infratentorial lesions	p	Subcortical lesions	p
Reader 1										
FLAIR vs. synthDIR	22.8 ± 12.7 31.4 ± 20.7	<0.001*	13.6 ± 9.35 16.7 ± 12.8	<0.001*	7.2 ± 5.55 12.3 ± 10.8	<0.001*	2.0 ± 1.68 2.4 ± 2.2	0.017*	12.1 ± 10.2 15.1 ± 10.8	<0.001*
FLAIR vs. trueDIR	22.8 ± 12.7 36.1 ± 24.3	<0.001*	13.6 ± 9.35 18.8 ± 14.3	<0.001*	7.2 ± 5.55 14.7 ± 13.1	<0.001*	2.0 ± 1.68 2.6 ± 2.2	<0.001*	12.1 ± 10.2 15.8 ± 11.7	<0.001*
synthDIR vs. trueDIR	31.4 ± 20.7 36.1 ± 24.3	<0.001*	16.7 ± 12.8 18.8 ± 14.3	0.002*	12.3 ± 10.8 14.7 ± 13.1	0.004*	2.4 ± 2.2 2.6 ± 2.2	0.14	15.1 ± 10.8 15.8 ± 11.7	0.10
Reader 2										
FLAIR vs. synthDIR	15.3 ± 10.4 21.7 ± 20.8	0.026*	10.7 ± 8.36 13.0 ± 12.7	0.29	3.7 ± 3.45 7.4 ± 9.6	0.0011*	0.86 ± 1.26 1.3 ± 2.2	0.13	14.6 ± 11.3 16.5 ± 12.2	0.074
FLAIR vs. trueDIR	15.3 ± 10.4 30.3 ± 23.8	<0.001*	10.7 ± 8.36 17.8 ± 15.1	<0.001*	3.7 ± 3.45 11.1 ± 13.1	<0.001*	0.86 ± 1.26 1.4 ± 1.6	0.0027*	14.6 ± 11.3 18.8 ± 13.2	<0.001*
synthDIR vs. trueDIR	21.7 ± 20.8 30.3 ± 23.8	<0.001*	13.0 ± 12.7 17.8 ± 15.1	<0.001*	7.4 ± 9.6 11.1 ± 13.1	<0.001*	1.3 ± 2.2 1.4 ± 1.6	0.31	16.5 ± 12.2 18.8 ± 13.2	0.0018*

Table 3. Location-dependent lesion count differences, discerned for periventricular, juxtacortical, infratentorial and subcortical lesions, as well as a composite of all MS-specific lesions (juxtacortical + periventricular + infratentorial). Significant results are highlighted with *. PV=periventricular; JC=juxtacortical; IT=infratentorial. Counts are given for readers 1 and 2.

Figures

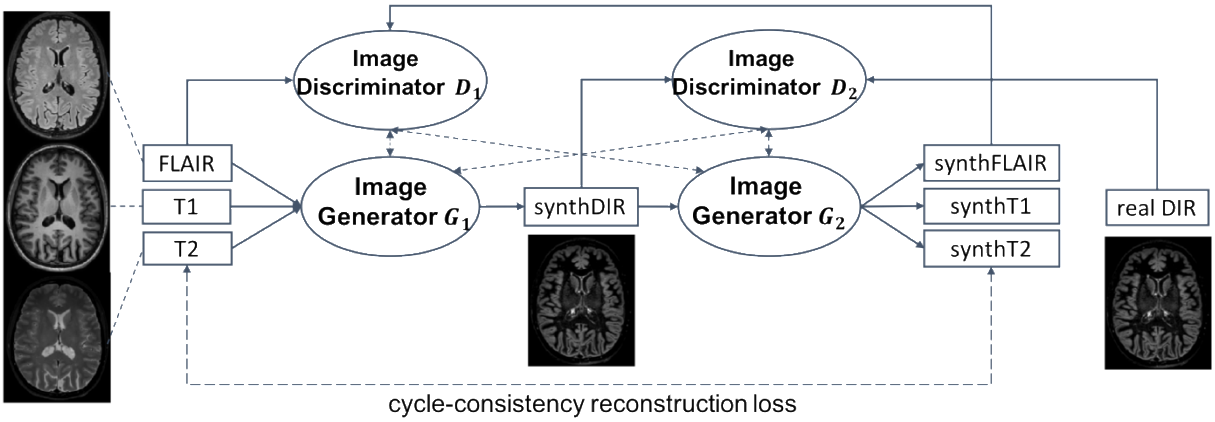

Figure 1: A schematic view of the proposed image synthesis system using generative adversarial networks, including two image generators and two image discriminators. The four networks are simultaneously optimized to generate high-quality images.

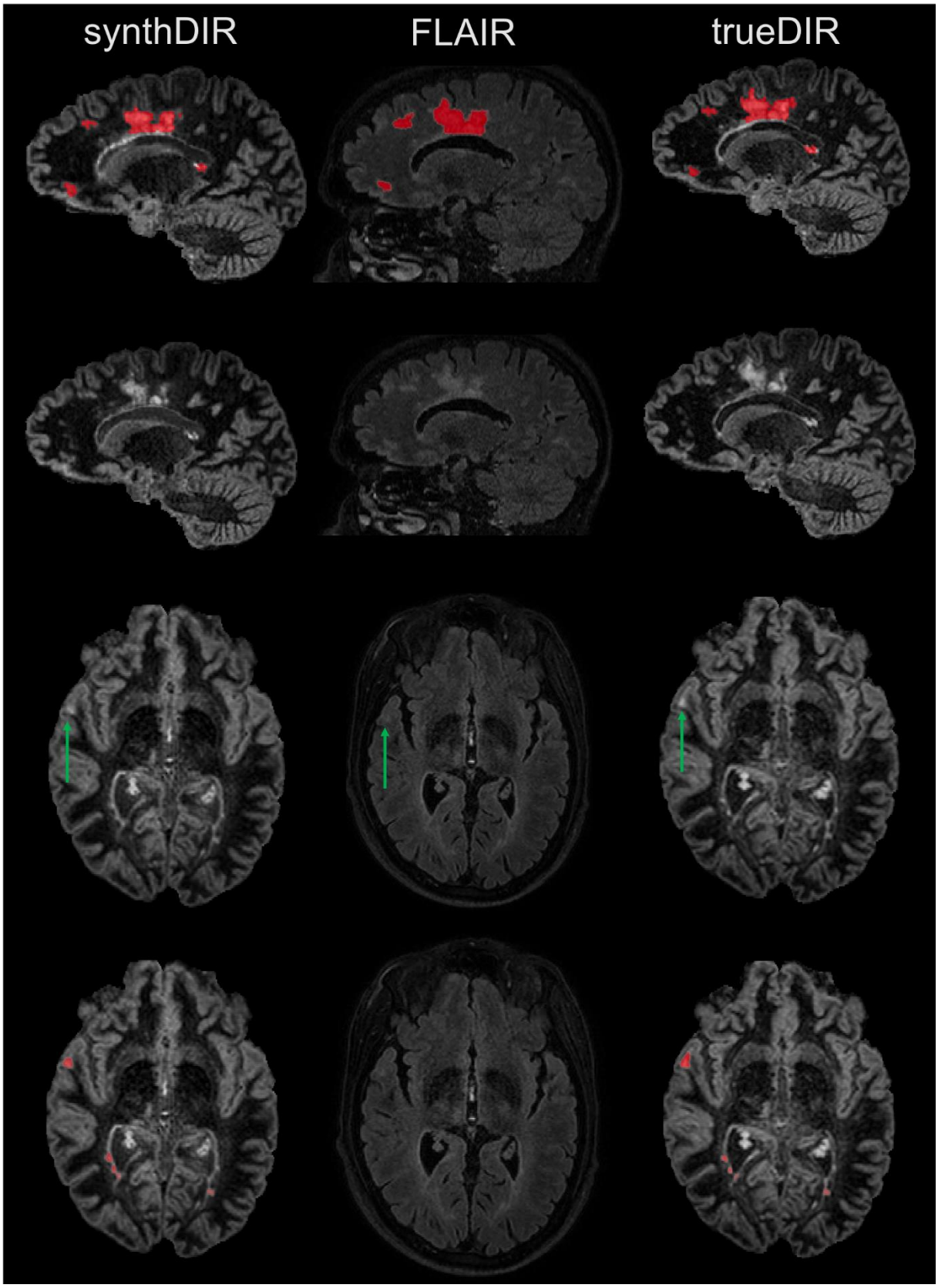

Figure 2. Exemplary images of FLAIR, synthDIR and trueDIR from the same patient. Given are sets of slices in the sagittal and axial plane with their respective lesion segmentations. Notable is the improved ability to detect juxtacortical lesions (green arrows) in synthDIR compared to (input) FLAIR.

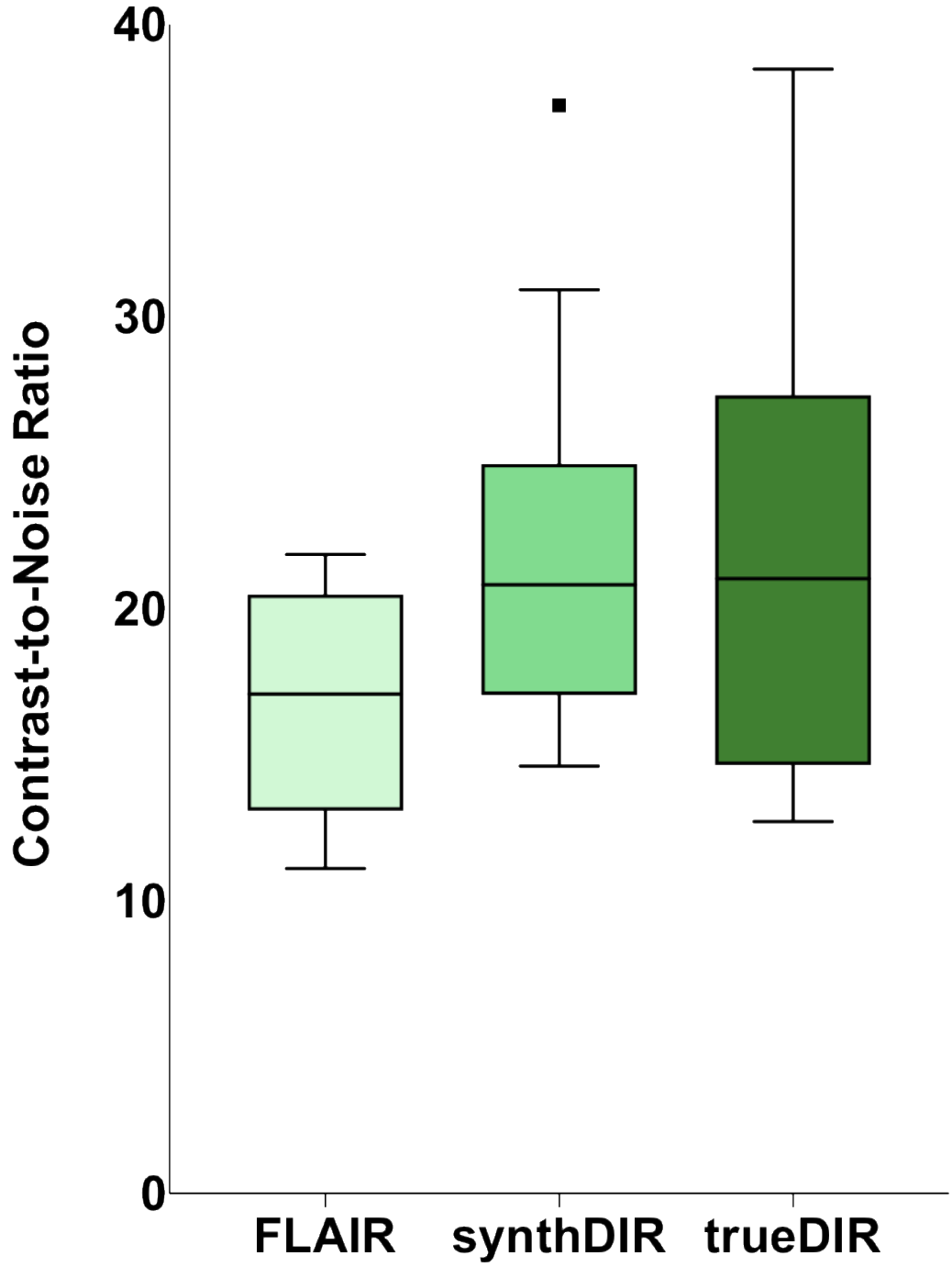

Figure 3. Contrast-to-noise ratios for FLAIR, synthDIR and trueDIR.

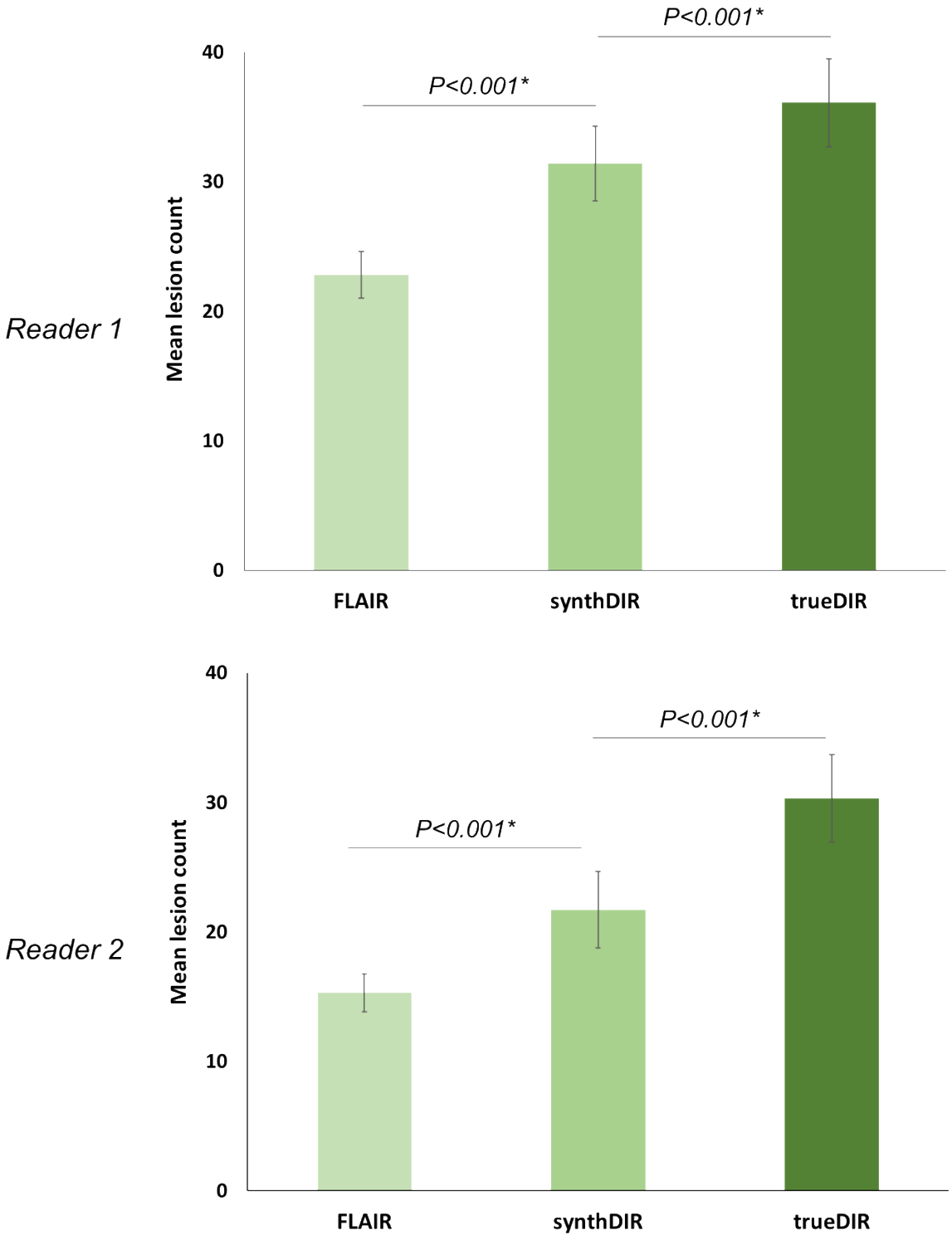

Figure 4. Mean counts of specific lesions (± standard error of the mean) for FLAIR, synthDIR and trueDIR. Counts are given for both readers. Significant differences are highlighted with *.