# Super-resolution reconstruction of multi-slice T2-W FLAIR MRI improves multiple sclerosis lesion segmentation

# Super-Resolution Reconstruction of Multi-Slice T2-W FLAIR MRI Improves Multiple Sclerosis Lesion Segmentation

Diana L. Giraldo[1,2], Quinten Beirinckx[1,2], Arnold J. Den Dekker[1,2], Ben Jeurissen[1,2,3], Jan Sijbers[1,2]

***Abstract*— Due to acquisition time constraints, T2-w FLAIR MRI of Multiple Sclerosis (MS) patients is often acquired with multi-slice 2D protocols with a low through-plane resolution rather than with high-resolution 3D protocols. Automated lesion segmentation on such low-resolution (LR) images, however, performs poorly and leads to inaccurate lesion volume estimates. Super-resolution reconstruction (SRR) methods can then be used to obtain a high-resolution (HR) image from multiple LR images to serve as input for lesion segmentation. In this work, we evaluate the effect on MS lesion segmentation of three SRR approaches: one based on interpolation, a state-of-the-art self-supervised CNN-based strategy, and a recently proposed model-based SRR method. These SRR strategies were applied to LR acquisitions simulated from 3D T2-w FLAIR MRI of MS patients. Each SRR method was evaluated in terms of image reconstruction quality and subsequent lesion segmentation performance. When compared to segmentation on LR images, the three considered SRR strategies demonstrate improved lesion segmentation. Furthermore, in some scenarios, SRR achieves a similar segmentation performance compared to segmentation of HR images.**

***Clinical relevance*— This study demonstrates the positive impact of super-resolution reconstruction from T2-w FLAIR multi-slice MRI acquisitions on segmentation performance of MS lesions.**

## I. Introduction

Multiple sclerosis (MS) is a chronic autoimmune disease characterized by inflammation and demyelination in the central nervous system leading to progressive disability. Diagnosis of MS patients relies on the detection of lesions in the brain or spinal cord, which are visible in T2-weighted (T2-w) magnetic resonance images (MRI). Particularly, T2-w fluid-attenuated inversion recovery (FLAIR) is the core sequence for MS diagnosis and monitoring [1], when it comes to assessing the number, volume, and location of lesions.

The gold standard for lesion segmentation is still the manual delineation by trained specialists, a task that is time-consuming and prone to inter-operator variability [2]. Although several automated techniques for lesion segmentation have been proposed, with some showing promising results [3], none of them is yet recommended to be used in clinical practice [1]. One reason is that most automated tools for MRI analysis have been designed for and evaluated with acquisition protocols that are common in research but not in clinical scenarios [4]. Due to acquisition time constraints in the clinics, multi-slice 2D T2-w FLAIR images with orthogonal slice orientations are often acquired, which are faster than 3D acquisitions and provide sufficient in-plane resolution for visual assessment by radiologists. However, these 2D acquisitions use a low through-plane resolution, introducing partial volume effects and leading to poor results from automated segmentation tools. Moreover, when multiple images are acquired, these are not necessarily aligned due to involuntary patient motion. In this scenario, super-resolution reconstruction (SRR) methods can be applied to obtain a high-resolution (HR) image from multiple low-resolution (LR) images, likely improving the outcomes of automated lesion segmentation. However, SRR methods can also affect the extraction of pathology-related features by blurring tissue boundaries or by introducing false structures [5], therefore they need to be carefully evaluated when applied to MS patients' images.

In this work, we evaluate the impact on automated lesion segmentation and image reconstruction quality of three SRR methods that do not depend on additional sequences nor are required to be trained with external data: the first one based on iterative rigid registration, interpolation, and averaging; the second one is based on self-supervised training of a convolutional neural network [6]; and the third one is a model-based SRR method with joint motion estimation [7]. The three SRR methods considered were applied to LR images simulated from HR T2-w FLAIR images of MS patients; then, reconstructed images were used as inputs for automated lesion segmentation of lesions [8], [9], and segmentation performance was evaluated with respect to a consensus of manual delineations of lesions.

## II. Methods

The methodology for this evaluation consists of the following steps: first, we simulated LR images from HR images following two commonly used multi-slice 2D acquisition protocols; second, we applied SRR to obtain HR reconstructions from simulated LR acquisitions; third, we segmented lesions on the reconstructed HR images using two automated tools; and finally, we assessed lesion segmentation performance and compared it against lesion segmentations on simulated LR and ground truth HR images.

### A. MRI data

In this work, we used data from 15 MS patients from the MICCAI 2016 challenge dataset [2]. This dataset contains, for each patient, a HR T2-w FLAIR MRI and a

[1]imec-Vision Lab, University of Antwerp.
[2]$\mu$NEURO Research Centre of Excellence, University of Antwerp.
[3]Lab for Equilibrium Investigations & Aerospace, University of Antwerp.
Corresponding author e-mail: diana.giraldofranco@uantwerpen.be

consensus lesion segmentation computed from seven manual delineations. Images were acquired in three different centers following a 3D protocol, resulting in three different voxel sizes: $1.1\times0.5\times0.5$ mm$^3$, $1.25\times1.04\times1.04$ mm$^3$, and $0.7\times0.74\times0.74$ mm$^3$. Detailed information about data acquisition and compliance with ethical standards can be found in the dataset publication [2]. Raw T2-w FLAIR images from the dataset were denoised with adaptive non-local means [10], and bias-field corrected with N4 algorithm [11]. We adjusted both T2-w FLAIR and lesion masks to have isotropic voxels of 1 mm$^3$.

### *B. Simulation of multi-slice LR images*

Simulation of LR images consisted of applying a sequence of operators to the HR ground truth that model the multi-slice acquisition process [12]: unintended motion, image warping by a known geometric transformation, spatially invariant blurring with a point spread function that models the slice selection and in-plane sampling, downsampling, and corruption by Rician noise. We performed the simulations according to two different acquisition protocols that have been commonly used in the clinics:

- **Protocol A:** two 2D images acquired with axial and sagittal slice orientations, slice thickness $= 5$ mm, slice spacing $= 6$ mm, and in-plane resolution of $1\times1$ mm$^2$.
- **Protocol B:** three 2D images acquired with axial, sagittal, and coronal slice orientations, slice thickness $= 3$ mm, slice spacing $= 3$ mm, in-plane resolution of $1\times1$ mm$^2$.

It should be noted that both protocols thatconstitute different levels of difficulty for the reconstruction task, with Protocol A a more challenging protocol than Protocol B due to thicker slices, fewer orthogonal acquisitions, and the presence of slice gaps. An example of simulated LR T2-w FLAIR images is shown in Fig. 1.

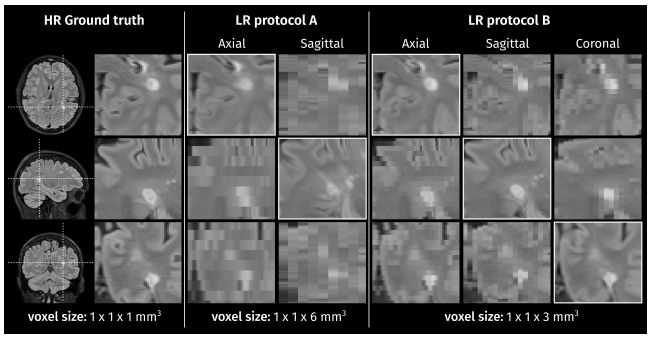


Fig. 1. Example of simulated LR acquisition protocols from one HR T2-w FLAIR (left).

### *C. Super-resolution reconstruction*

The three strategies considered for this evaluation have two advantages: they do not require additional sequences or external data to be trained, and either can deal with or are not affected by unintended rigid motion between LR images. These are:

- **Interpolation**: An iterative approach in which LR images are rigidly registered to a reference image and interpolated to match the desired resolution, then the average across interpolated images becomes the reference image for the next iteration. We implemented this strategy with cubic interpolation, 3 iterations, and we selected the LR image with axial slice orientation as the reference for the first iteration.
- **SMORE**: *"Synthetic Multi-Orientation Resolution Enhancement"* [6], it is an anti-aliasing and super-resolution algorithm based on self-trained convolutional neural networks to increase the through-plane resolution of multi-slice images. Since the method is devised to do the self-training and prediction by extracting small patches from only one LR image, we adapted the implementation by sequentially training the networks with the multiple LR images available per patient and protocol, and then reconstructing the HR image by applying the trained networks to the axial LR image. This method focuses on learning local structures rather than global structures, therefore it should not be affected by rigid motion between LR images.
- **Model-based SRR**: A recent, model-based super-resolution reconstruction method with joint inter-image motion estimation developed for improving quantitative MRI mapping [7]. This method uses a Bayesian Maximum a Posteriori (MAP) estimation framework with a total variation prithator for the HR image. Motion parameters and HR image are jointly estimated using a cyclic block-coordinate descent approach.

Implementation of these methods can be shared by the authors upon request.

### *D. Automated lesion segmentation*

To evaluate the impact of SRR on subsequent MS lesion segmentation, we applied two automated lesion segmentation algorithms to the HR outcomes of SRR methods, the simulated LR, and the HR ground truth images. To the best of our knowledge, there exist two lesion segmentation methods, with publicly available implementations, that accept T2-w FLAIR as the only input:

- **LST-lpa [8]:** the lesion prediction algorithm (lpa) in the Lesion Segmentation Toolbox (LST) for SPM. This algorithm is based on a logistic regression model that includes a lesion belief map and a spatial covariate that takes into account voxel-specific changes in lesion probability.
- **SAMSEG [9]:** a recently published lesion segmentation add-on to SAMSEG routine in Freesurfer, which allows the simultaneous segmentation of white matter lesions and 41 structures by decoupling computational models of anatomy from models of the imaging process.

### *E. Evaluation*

For each SRR strategy, we evaluated the image reconstruction quality with respect to the HR ground truth image using the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) calculated within a brain mask. We assessed lesion segmentation performance

with respect to the consensus mask by computing the Dice score, sensitivity, precision, and the error of lesion volume (LV) estimation. The latter was calculated as the difference between the voxel count in the predicted lesion mask and in the consensus mask, because both masks have 1 $mm^3$ voxels. Evaluation metrics were compared using paired Wilcoxon signed rank tests.

## III. Results

### A. Reconstruction image quathatthatlity

A summary of reconstruction image quality measures, PSNR and SSIM, is presented in Table I. Reconstruction quality from Protocol B is superior to that from Protocol A for all three SRR methods. This is expected because the LR images in Protocol B have a higher through-plane resolution and this protocol also includes an image with coronal slice orientation. When images are reconstructed from Protocol A, the three SRR methods result in similar mean PSNR ($\approx 24$ dB) with SSIM being slightly (but significantly) higher for model-based SRR compared to interpolation and SMORE. Results for reconstructions from Protocol B also show that model-based SRR outperforms Interpolation and SMORE in terms of PSNR and SSIM.

TABLE I
MEAN AND STANDARD DEVIATION OF PSNR AND SSIM FOR EACH SRR METHOD.

| LR protocol | Super-resolution method | PSNR (dB) | SSIM |
|---|---|---|---|
| A | Interpolation | $24.1 \pm 1.8$ | $0.66 \pm 0.02$ |
| | SMORE | $24.0 \pm 1.8$ | $0.66 \pm 0.02$ |
| | Model-based SRR | $24.0 \pm 2.0$ | $0.69 \pm 0.06\star$ |
| B | Interpolation | $29.4 \pm 1.5$ | $0.89 \pm 0.02$ |
| | SMORE | $28.2 \pm 2.2$ | $0.84 \pm 0.05$ |
| | Model-based SRR | $33.5 \pm 4.0\star$ | $0.91 \pm 0.11\star$ |

$\star$ Significantly better than the other two methods ($p < 0.01$).

### B. Automated lesion segmentation performance

The distributions of Dice scores, sensitivity, precision, and errors of LV estimates for both automated lesion segmentation methods are shown in Fig. 2. Examples of resulting lesion segmentation masks over reconstructed T2-w FLAIR images are shown in Figure 3.

When applying LST-lpa, the three SRR strategies lead to higher Dice scores ($p \leq 0.0001$), higher precision ($p \leq 0.0001$), and smaller absolute errors in LV estimates ($p \leq 0.0001$) compared to segmentations using LR images, with almost no significant differences in sensitivity except for the case when interpolation is used to reconstruct images from Protocol A. In the case of reconstructions from protocol A, model-based SRR leads to significantly higher Dice scores ($p \leq 0.004$), higher sensitivity ($p \leq 0.0002$), and smaller absolute error of LV ($p \leq 0.05$) than interpolation, with interpolation leading to higher precision ($p \leq 0.0004$) than SMORE and model-based SRR, indicating more conservative lesion segmentations. If reconstructed images from Protocol B are given as inputs to LST-lpa, interpolation also produces

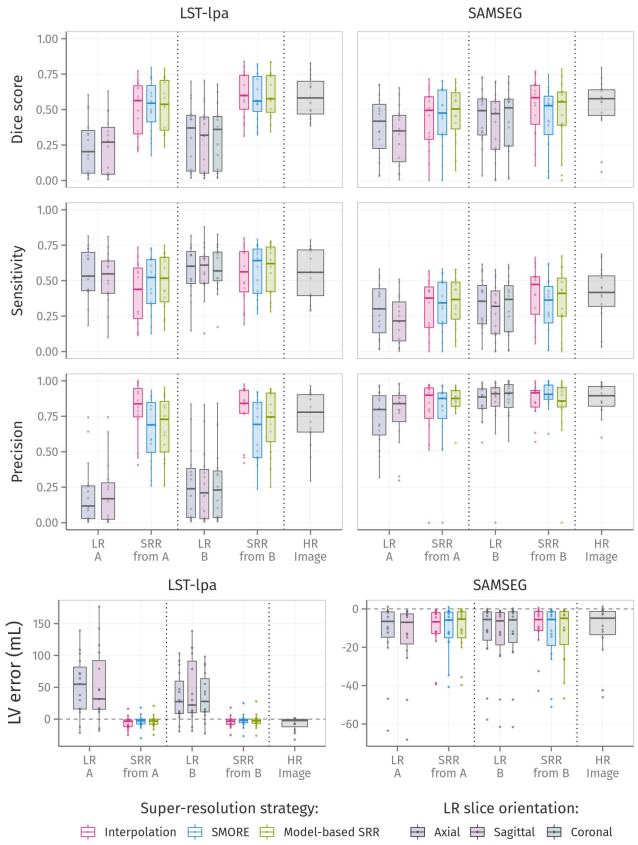


Fig. 2. Box plots of Dice scores, sensitivity, precision, and lesion volume (LV) estimation error of automated lesion segmentation on simulated low-resolution (LR), super-resolution reconstructions (SRR), and high-resolution (HR) ground truth T2-w FLAIR images.

segmentations with lower sensitivity ($p \leq 0.04$) and higher precision ($p \leq 0.0001$) than SMORE and model-based SRR. For Protocol B, none of the three SRR methods results on Dice scores or LV estimates significantly different from what is obtained if the segmentation method is applied to the ground truth image.

In the case of SAMSEG, all three SRR methods improve the Dice score ($p \leq 0.03$) and sensitivity ($p \leq 0.02$) of lesion segmentation compared to segmentations using LR images with sagittal orientation. For reconstructions from Protocol A, there are no significant differences in segmentation performance metrics between SRR strategies, with all of them underperforming segmentation using the HR ground truth image in terms of Dice score and sensitivity. When images are reconstructed form Protocol B, interpolation and model-based SRR lead to segmentations with significantly better Dice scores ($p \leq 0.04$) and sensitivity ($p \leq 0.05$) than SMORE, with SAMSEG segmentation on SMORE reconstructions still underperforming segmentation on the HR image in terms of LV estimation, Dice score, and sensitivity.

This evaluation also shows that, when applied to T2-w FLAIR MRI, SAMSEG tends to produce more conservative lesion masks than LST-lpa. Furthermore, the difference between applying SAMSEG to LR images and applying it to HR images is not as large as with LST-lpa, suggesting some robustness to resolution and blurring but little refine-

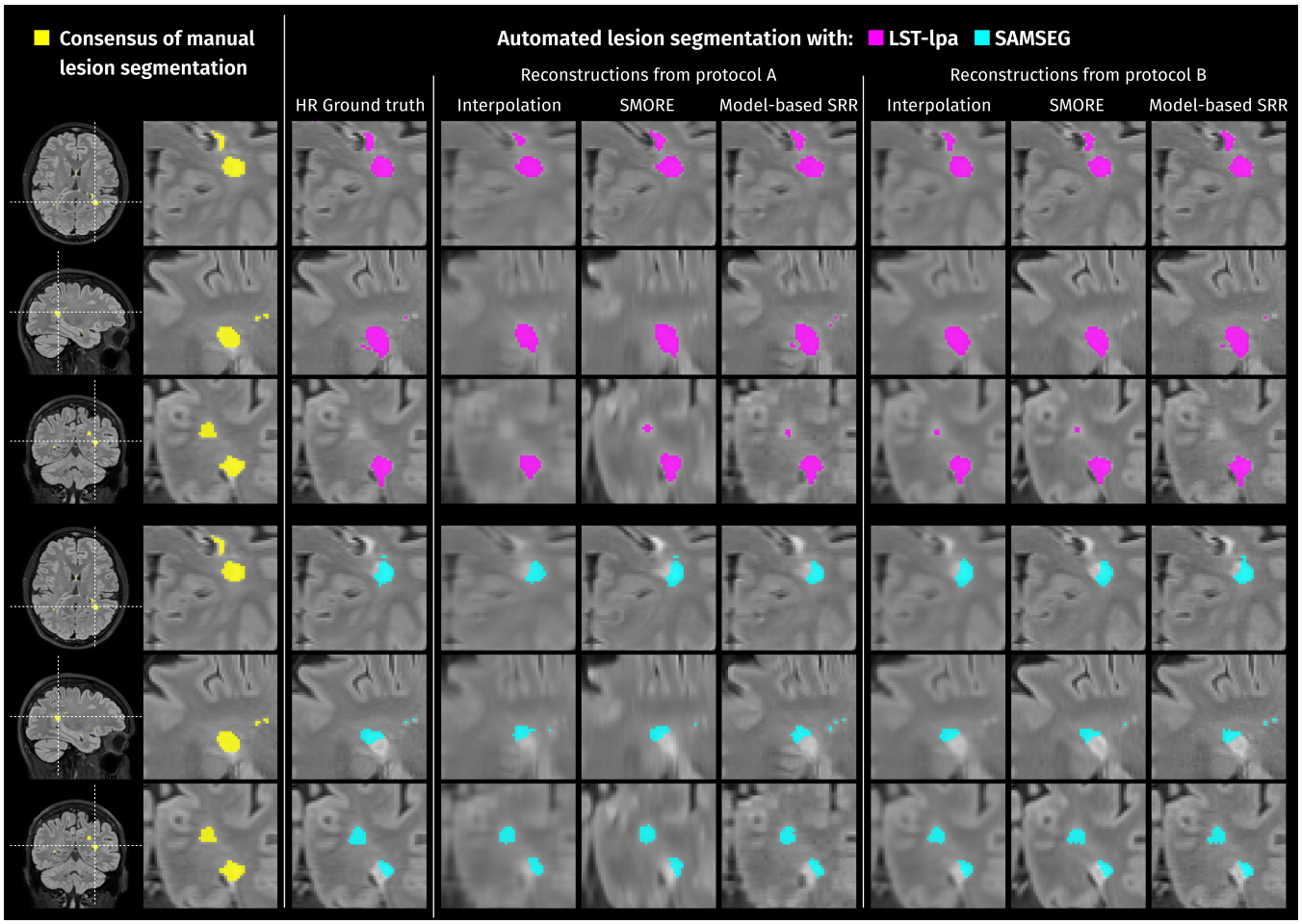


Fig. 3. Example of automated lesion segmentation methods applied to reconstructions from both simulated LR protocols with the three SRR strategies considered.

ment when lesion boundaries are sharp (See bottom row in Fig. 3), mitigating the positive effect of SRR before lesion segmentation.

## IV. Conclusion

The three SRR strategies evaluated in this work, namely, interpolation, SMORE, and model-based SRR, improve the automated segmentation of MS lesions from LR multi-slice T2-w FLAIR MR images by reconstructing a HR image before lesion segmentation. Acquisitions with little spatial information like Protocol A benefit more from model-based SRR or SMORE than from interpolation, and acquisitions with Protocol B benefit from the three methods to a similar extent. The integration of SRR in automated processing pipelines would facilitate the use of data acquired in the clinics in the investigation of MRI-based markers of disease outcomes.

## Acknowledgment

This research received funds from the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" program.

## References

[1] M. P. Wattjes *et al.*, "2021 MAGNIMS–CMSC–NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis," *Lancet Neurol.*, vol. 20, no. 8, pp. 653–670, 2021.

[2] O. Commowick *et al.*, "Multiple sclerosis lesions segmentation from multiple experts: The MICCAI 2016 challenge dataset," *NeuroImage*, vol. 244, p. 118589, 2021.

[3] A. Carass *et al.*, "Evaluating white matter lesion segmentations with refined sørensen-dice analysis," *Scientific Reports*, vol. 10, no. 1, 2020.

[4] Z. Mendelsohn *et al.*, "Commercial volumetric MRI reporting tools in multiple sclerosis: a systematic review of the evidence," *Neuroradiology*, vol. 65, no. 1, pp. 5–24, Nov. 2023.

[5] S. Bhadra *et al.*, "On hallucinations in tomographic image reconstruction," *IEEE Trans. Med. Imaging*, vol. 40, no. 11, pp. 3249–3260, 2021.

[6] C. Zhao *et al.*, "SMORE: A self-supervised anti-aliasing and super-resolution algorithm for MRI using deep learning," *IEEE Trans. Med. Imaging*, vol. 40, no. 3, pp. 805–817, 2021.

[7] Q. Beirinckx *et al.*, "Model-based super-resolution reconstruction with joint motion estimation for improved quantitative MRI parameter mapping," *Comput. Med. Imaging Graph.*, vol. 100, p. 102071, 2022.

[8] P. Schmidt, "Bayesian inference for structured additive regression models for large-scale problems with applications to medical imaging," Ph.D. dissertation, Ludwig-Maximilians-Universität München, 2017.

[9] S. Cerri *et al.*, "A contrast-adaptive method for simultaneous whole-brain and lesion segmentation in multiple sclerosis," *NeuroImage*, vol. 225, p. 117471, 2021.

[10] J. V. Manjón *et al.*, "Adaptive non-local means denoising of MR images with spatially varying noise levels," *J. Magn. Reson. Imaging.*, vol. 31, no. 1, pp. 192–203, 2010.

[11] N. J. Tustison *et al.*, "N4ITK: Improved N3 bias correction," *IEEE Trans. Med. Imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.

[12] D. H. J. Poot *et al.*, "General and efficient super-resolution method for multi-slice MRI," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, vol. 6361, 2010, pp. 615–622.