Perceptual super-resolution in multiple sclerosis MRI

Diana L. Giraldo[1,2,3]*, Hamza Khan[4,5,6], Gustavo Pineda[3], Zhihua Liang[1,2], Alfonso Lozano-Castillo[7], Bart Van Wijmeersch[8], Henry C. Woodruff[6,9], Philippe Lambin[6,9], Eduardo Romero[3], Liesbet M. Peeters[4,5] and Jan Sijbers[1,2]

[1]Imec-Vision Lab, University of Antwerp, Antwerp, Belgium, [2]µNEURO Research Center of Excellence, University of Antwerp, Antwerp, Belgium, [3]Computer Imaging and Medical Applications Laboratory—Cim@Lab, Universidad Nacional de Colombia, Bogotá, Colombia, [4]University MS Center, Biomedical Research Institute, Hasselt University, Hasselt, Belgium, [5]Data Science Institute (DSI), Hasselt University, Hasselt, Belgium, [6]The D-Lab, Department of Precision Medicine, GROW-Research Institute for Oncology and Reproduction, Maastricht University, Maastricht, Netherlands, [7]Department of Diagnostic Imaging, Hospital Universitario Nacional, Universidad Nacional de Colombia, Bogotá, Colombia, [8]Noorderhart, Revalidatie en Multiple Sclerose, Pelt, Belgium, [9]Department of Radiology and Nuclear Imaging, GROW-Research Institute for Oncology and Reproduction, Maastricht University Medical Centre+, Maastricht, Netherlands

Introduction: Magnetic resonance imaging (MRI) is crucial for diagnosing and monitoring of multiple sclerosis (MS) as it is used to assess lesions in the brain and spinal cord. However, in real-world clinical settings, MRI scans are often acquired with thick slices, limiting their utility for automated quantitative analyses. This work presents a single-image super-resolution (SR) reconstruction framework that leverages SR convolutional neural networks (CNN) to enhance the through-plane resolution of structural MRI in people with MS (PwMS).

Methods: Our strategy involves the supervised fine-tuning of CNN architectures, guided by a content loss function that promotes perceptual quality, as well as reconstruction accuracy, to recover high-level image features.

Results: Extensive evaluation with MRI data of PwMS shows that our SR strategy leads to more accurate MRI reconstructions than competing methods. Furthermore, it improves lesion segmentation on low-resolution MRI, approaching the performance achievable with high-resolution images.

Discussion: Results demonstrate the potential of our SR framework to facilitate the use of low-resolution retrospective MRI from real-world clinical settings to investigate quantitative image-based biomarkers of MS.

1 Introduction

Multiple sclerosis (MS) is a chronic autoimmune neurodegenerative disease characterized by inflammation and demyelination of nerve axons in the central nervous system. This damage leads to the formation of lesions, which are the most important markers of disease activity (Kolb et al., 2022). Diagnosis and monitoring of people with MS (PwMS) relies on the acquisition of Magnetic Resonance Imaging (MRI), particularly T2-weighted (T2-W) fluid-attenuated inversion recovery (FLAIR), to assess lesions in the white matter (WM) (Wattjes et al., 2021). Although current clinical guidelines recommend acquiring high-resolution (HR) T2-W FLAIR MRI using 3D sequences (Wattjes et al., 2021), in clinical settings images have been often acquired with 2D sequences, where the resulting "3D" image is rather a stack of thick 2D slices with highly anisotropic voxels. Such multi-slice images are faster to acquire, are less prone to motion artifacts, and

their in-plane resolution is often sufficient for visual inspection by radiologists. However, their poor through-plane resolution hampers their use for precise quantitative analyzes (e.g., radiomics) as most of the automated methods proposed for lesion segmentation and morphometric analyzes require HR images with isotropic voxels (Danelakis et al., 2018; Carass et al., 2020; Mendelsohn et al., 2023). In this scenario, super-resolution (SR) methods aiming to improve the spatial resolution of acquired MRI would facilitate the use of real-world MRI and clinical data of PwMS to investigate MS biomarkers.

Super-resolution methods aim to estimate an unknown HR MRI from one or more acquired low-resolution (LR) MRIs. Model-based methods assume an explicit imaging model of the MRI acquisition and seek a numerical solution of the ill-posed inverse problem by introducing regularization terms to constrain the solution space (Shilling et al., 2009; Poot et al., 2010; Gholipour et al., 2015; Beirinckx et al., 2022). These model-based SR methods, however, require multiple multi-slice LR images to reconstruct one HR MRI. Hence, they are not suitable for single-image SR. Even when two multi-slice LR images (e.g., with orthogonal slice direction) are available, their performance is limited (Giraldo et al., 2023). Alternatively, learning-based single-image SR approaches, for example based on convolutional neural networks (CNN), have demonstrated impressive performance in natural images, due to their ability to learn the relation between the LR and HR images from data (Dong et al., 2014; Johnson et al., 2016; Ledig et al., 2017; Lim et al., 2017; Blau et al., 2019; Wang et al., 2021). A straightforward procedure is to train a CNN generative model with paired LR-HR samples, enabling the model to learn a SR mapping that can be applied to unseen LR data with generally fast inference times. Despite their demonstrated effectiveness on natural images, the performance of these SR approaches heavily relies on the variability and characteristics of the data the models were trained and tested with. CNN models trained with large datasets of natural images do not capture the statistical properties of medical images, including specific acquisition conditions, artifacts and types of noise. In medical imaging, preserving anatomical and pathological features is crucial to avoid misleading diagnostic judgments. Therefore, applying CNN SR models to medical images requires fine-tuning these models to the specific application domain. This process ensures the models capture domain-specific features and maintain diagnostic integrity.

In the context of brain MRI SR, several works have presented and evaluated CNNs for single-image SR of structural MRI sequences, primarily T1-W and T2-W scans. Sanchez and Vilaplana (2018) proposed a 3D Generative Adversarial Network (GAN) architecture inspired by the SRGAN model (Ledig et al., 2017). Their method achieved promising quantitative results and demonstrated the potential of GANs for MRI SR compared to conventional cubic spline interpolation. Pham et al. (2019) presented a 3D CNN based on residual learning, with the underlying assumption that it is easier to find a mapping from the missing high-frequency information to HR, instead of finding a direct mapping from LR to HR. A similar residual learning approach was adopted by Du et al. (2020), who proposed a 2D CNN network for SR reconstruction of multi-slice T1-W and T2-W MRI, although it was mainly evaluated on synthetic brain images.

Addressing the challenge of collecting training data, Zhao et al. (2019, 2021) introduced a self-supervised approach based on 2D CNNs, which leveraged the high in-plane resolution of acquired MRI to train a SR CNN for increasing the through-plane resolution. This method showcased improved performance over previous self-supervised methods when evaluated for the SR of multi-slice T2-W MRI. However, the requirement to train or fine-tune a CNN model each time the method is applied to a new image poses an important practical limitation in terms of computational cost and processing time. A CNN model trained to convert a brain MRI of any orientation, resolution and contrast into an HR T1-W MRI was recently presented (Iglesias et al., 2021, 2023). While this approach holds promise for facilitating standard morphometric analyzes by inpainting normal-appearing tissue in pathological areas, it does not facilitate the quantitative analysis of lesions in PwMS.

Most CNN strategies for MRI SR rely on minimizing either the mean squared error ($\mathcal{L}_2$ loss) (Pham et al., 2017, 2019; Du et al., 2020) or mean absolute error ($\mathcal{L}_1$ loss) (Zhao et al., 2021; Iglesias et al., 2023) between the model output and the ground truth HR image during model training. While this minimization approach results in improved reconstruction accuracy measures such as peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), it can produce images with over-smoothed textures and blurry boundaries (Timofte et al., 2017; Blau et al., 2019). To address this limitation, previous works have explored the use of GANs (Sanchez and Vilaplana, 2018; Zhang et al., 2022), which encourage the model to generate more realistic-looking images by incorporating an adversarial loss. However, this adversarial approach carries the risk of hallucinating structures or introducing artifacts in generated images. Another promising approach to improve the perceptual quality in SR is the use of loss functions that compare high-level image features rather than relying solely on pixel-wise similarities. These high-level features are often extracted from intermediate layers of pre-trained neural networks (Johnson et al., 2016). Perceptual losses, which transfer semantic knowledge from the pre-trained loss network, have demonstrated their effectiveness in improving the perceptual quality of single-image SR for natural images (Johnson et al., 2016; Wang et al., 2019). Recently, Zhang et al. (2022) incorporated a perceptual loss into a GAN framework for MRI SR, which was trained with T1-W brain MRI of PwMS and then tested on T2-W FLAIR.

Current state-of-the-art approaches for single-image SR of multi-slice MRI (Du et al., 2020; Zhao et al., 2021; Zhang et al., 2022) have been evaluated in scenarios where the upsampling scale factor between the slice thickness of LR and HR ranges from 2 to 6. As expected, the performance of SR reconstruction decreases as the input slice thickness increases, leading to greater challenges in faithfully recovering anatomical structures and details. The larger resolution gap between LR input and HR target exacerbates the risk of hallucinating artificial features, compromising the diagnostic quality and reliability of SR outputs, a critical concern when applying SR to brain MRI of PwMS. A particularly challenging scenario, that is prevalent in real-world retrospective image datasets of PwMS, involves MRI acquisitions with a thick slice spacing of 6 mm. Moreover, these multi-slice MRI scans are often acquired with slice gaps, a factor that is frequently overlooked when generating

LR training data for SR models but can significantly influence their performance (Han et al., 2023).

In this paper, we present a SR framework that leverages SR CNN architectures to enhance the resolution of multi-slice structural MRI of PwMS, reducing the slice spacing from 6 to 1 mm. Our strategy involves fine-tuning SR models with image patches extracted from T2-W FLAIR and T1-W MRI data of PwMS. The fine-tuning of SR models is guided by a content loss $\mathcal{L}_c$ which simultaneously promotes perceptual quality and pixel-wise reconstruction accuracy, ensuring realistic textures and well-defined tissue boundaries in the reconstructed HR images. Our framework reconstructs HR MRI volumes from single LR inputs, enabling more accurate downstream 3D analysis. Through a comprehensive evaluation using multi-center MRI data, we demonstrate that our SR framework outperforms existing methods in terms of reconstruction accuracy, and additionally improves the performance of automated lesion segmentation on T2-W FLAIR MRI, a highly relevant task in the context of multiple sclerosis.

We summarize the contributions of this work as follow:

- We present a single-image SR framework for multi-slice MRI based on the adaptation of CNN architectures to the domain of structural MRI in MS.
- We fine-tune two state-of-the-art SR CNN models, namely the EDSR (Lim et al., 2017) and the RealESRGAN (Wang et al., 2021), using a perceptual loss to recover realistic features, and the mean absolute error ($\mathcal{L}_1$ loss) to control the reconstruction accuracy.
- We incorporate information from the MRI physics model by simulating LR data from HR MRI with an acquisition model that accounts for a slice selection profile, including slice gap, commonly found in clinical multi-slice MRI data.
- We evaluate our framework for MRI SR reconstruction and compare its performance against existing MRI SR methods using MRI datasets of PwMS from different centers.
- We evaluate the impact of our SR strategy in a relevant downstream task: the automated segmentation of white matter lesions on reconstructed T2W-FLAIR images.

We named our framework PRETTIER, a name encapsulating its purpose: "Perceptual super-REsoluTion in mulTIple sclERosis." The code to apply PRETTIER is available at: https://github.com/diagiraldo/PRETTIER.

2 Materials and methods

An overview of the workflow is presented in Figure 1. The process begins with MRI data preparation, consisting of LR MRI simulation and extraction of paired LR-HR patches. Then, in the learning step, CNN models for image SR are fine-tuned using these patches, with final weights selected based on the minimum content loss in a validation set. We then use each fine-tuned CNN model to reconstruct HR MRI volumes from single LR MRI inputs, combining outputs from applying the model in different slice directions. The evaluation step uses an independent set of structural MRI from PwMS and comprises three parts: First, we evaluate the fine-tuning of CNN models with MRI patches. Second,

we evaluate the SR framework by reconstructing MRI volumes. Finally, we assess the impact of SR reconstruction on the automated segmentation of WM lesions.

2.1 MRI data

We gathered six different datasets containing 429 HR structural MRIs, T2-W FLAIR, and T1-W, of 192 PwMS. The numbers of unique subjects, images, and data partition are shown in Table 1.

- **ISBI2015** (Carass et al., 2017): 61 T1-W images from the test data for the longitudinal MS lesion segmentation challenge held during the ISBI 2015 conference. All images were acquired on a 3T Philips MRI scanner.
- **Lesjak-3D** (Lesjak et al., 2017): 30 T2-W FLAIR images accompanied by their lesion annotation resulting from the consensus of three experts. Images were acquired on a 3T Siemens Magnetom Trio MR system at the University Medical Center Ljubljana.
- **MSSEG1** (Commowick et al., 2021b): 53 T2-W FLAIR and 53 T1-W images from the MS lesions segmentation challenge held in MICCAI 2016. For a subset of 15 subjects (subset A) the dataset also provides the WM lesion annotation resulting from the consensus of seven experts. Images were acquired in four different centers with 1.5T and 3T MR scanners from different vendors (Siemens, Philips, and GE).
- **MSSEG2** (Commowick et al., 2021a): 80 T2-W FLAIR images from the MS new lesions segmentation challenge held in MICCAI 2021. This dataset contains scans of 40 PwMS acquired in two different time-points. Images were acquired with 11 different 1.5 T and 3T MR scanners.
- **MSPELT**: 56 T2-W FLAIR and 41 T1-W images of 41 PwMS from the Noorderhart—Revalidatie and MS Centrum in Pelt, Belgium. All images were acquired in a 1.5 T MR scanner (Philips Achieva dStream). The use of this pseudonymized retrospective dataset was approved by the ethical commission of the University of Hasselt.
- **HUN**: 28 T2-W FLAIR and 27 T1-W images of 14 PwMS from the Hospital Universitario Nacional (HUN) in Bogota, Colombia. This dataset contains, for each subject, scans at two different visits and the manual annotation of WM lesions verified by an expert neuroradiologist (20 years of experience) for each visit. All images were acquired in a 1.5 T MR scanner (Philips Multiva). Participants provided voluntary and informed consent, and the collection of images received ethical approval from the HUN ethics committee.

Detailed information about data acquisition and compliance with ethical standards for the four public datasets (ISBI2015, Lesjak-3D, MSSEG-1, and MSSEG-2) can be found in their corresponding publications.

2.1.1 MRI pre-processing

The raw versions of HR T2-w FLAIR and T1-W images were denoised with adaptive non-local means (Manjón et al., 2010), and
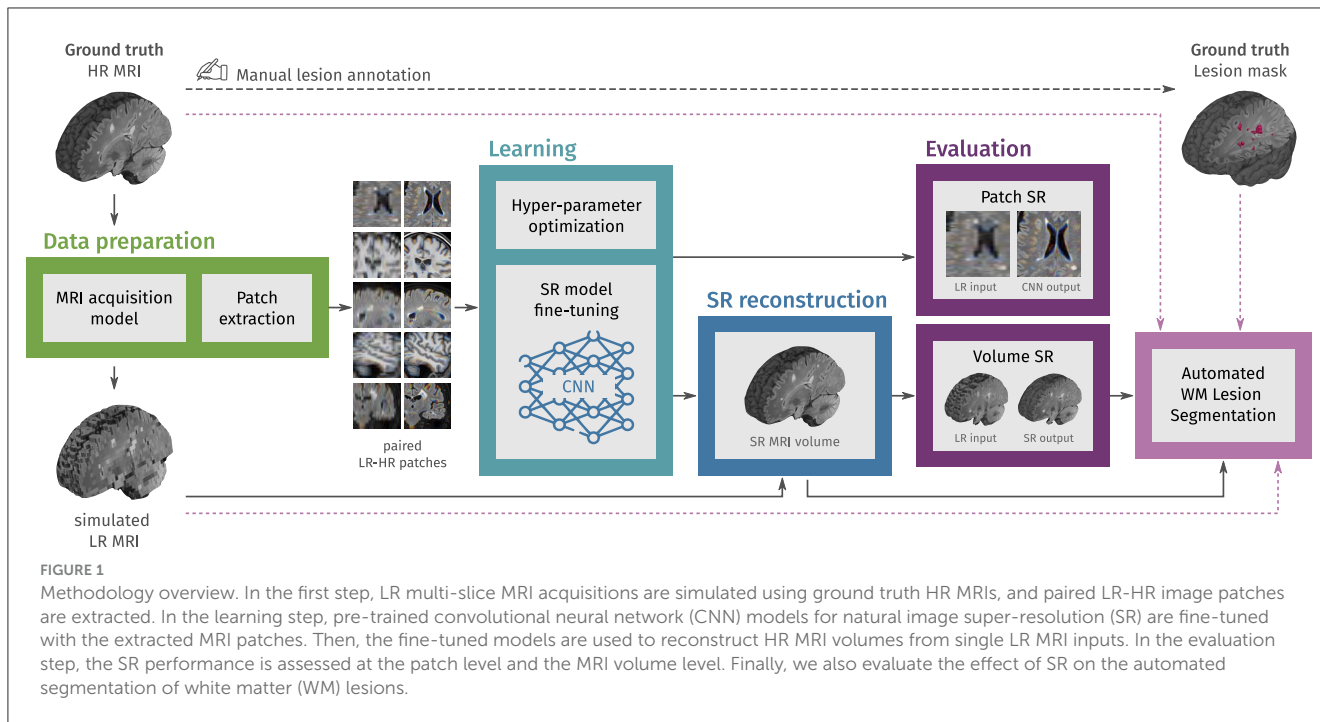
FIGURE 1
Methodology overview. In the first step, LR multi-slice MRI acquisitions are simulated using ground truth HR MRIs, and paired LR-HR image patches are extracted. In the learning step, pre-trained convolutional neural network (CNN) models for natural image super-resolution (SR) are fine-tuned with the extracted MRI patches. Then, the fine-tuned models are used to reconstruct HR MRI volumes from single LR MRI inputs. In the evaluation step, the SR performance is assessed at the patch level and the MRI volume level. Finally, we also evaluate the effect of SR on the automated segmentation of white matter (WM) lesions.

TABLE 1 Overview of the multi-centric dataset used in this work.

Dataset	Source	# subjects	# FLAIR images	# T1 images	Lesion annotation	Data partition
ISBI 2015	Public	14	-	61	-	Learning
Lesjak-3D	Public	30	30	-	✓	Learning
MSSEG1 B	Public	38	38	38	-	Learning
MSSEG2	Public	40	80	-	-	Learning
MSPELT	Clinical	41	56	41	-	Learning
MSSEG1 A	Public	15	15	15	✓	Evaluation
HUN	Clinical	14	28	27	✓	Evaluation
Total		192	247	182		

The two datasets used for evaluation, MSSEG1 A and HUN, have manual delineation of lesions to evaluate automated lesion segmentation.

bias-field corrected with N4 algorithm (Tustison et al., 2010). For each image, a brain mask was estimated using the HD-BET tool (Isensee et al., 2019). All images, brain masks and ground truth lesion masks were adjusted to have isotropic voxels of 1 mm^3 using cubic interpolation for MRIs and nearest neighbor interpolation for the binary masks.

2.1.2 Data partition

The set of 192 subjects was partitioned into learning and evaluation sets with the consideration of having ground truth segmentations and a balanced number of image modalities in the evaluation set (MSSEG1 A and HUN). As shown in Table 1, this first partition resulted in 163 subjects (204 T2-W FLAIR and 140 T1-W MRIs) in the learning set, and 29 subjects (43 T2-W FLAIR

and 42 T1-W MRIs) in the evaluation set. Images in the learning set were further partitioned into training and validation using a stratified scheme: for each dataset in the learning set, image sessions were randomly split into training and validation with a proportion of 7:3. The validation set, resulting from this second partition, was used to optimize the training hyper-parameters and select the best model weights during fine-tuning.

2.2 Data preparation

2.2.1 Simulation of LR images

We obtained pairs of LR-HR MRI volumes $(\mathbf{y_i}, \mathbf{x})$ by applying the multi-slice MRI acquisition model $\mathcal{M}_i$ (Poot et al., 2010) to the

HR ground truth image:

$$\mathbf{y_i} = \mathcal{M}_i(\mathbf{x}) = \mathcal{DBR}_i(\mathbf{x}) \tag{1}$$

Where $\mathcal{R}_i$ is a rotation operator that accounts for the slice direction of the multi-slice LR image, $\mathcal{B}$ is the blurring operator with a filter that accounts for the slice-selection profile, and $\mathcal{D}$ is the downsampling operator. For each HR image $\mathbf{x}$, we simulated three LR images $\{y_i\}_{i=1}^3$ with orthogonal slice orientations (axial, sagittal, and coronal), with slice thickness of 5 mm and slice spacing of 6 mm (a slice gap of 1 mm). The simulated LR images have highly anisotropic voxels of $1 \times 1 \times 6$ mm^3, which are common in retrospective clinical multi-slice MRI.

2.2.2 Extraction of patches

Most CNN models for image SR receive as inputs 2D images with three color channels (R: red, G: green, B: blue). For model fine-tuning, we randomly extracted pairs of 3-channel patches from each pair of LR-HR MRI by taking triplets of 2D patches that are contiguous in the third dimension. As we considered only fully convolutional CNN models, there are no restrictions in terms of patch size. To keep the memory requirements relatively low, we extracted HR patches of 96×96 while ensuring most of the patch area correspond to brain tissue. Therefore, the LR patches were 96×16 or 16×96, depending on the slicing orientation they were taken from their corresponding LR volume.

For the learning stage, 18 pairs of LR-HR patches were extracted from each LR-HR MRI pair resulting in 13, 344 training samples and 3, 168 validation samples. For the evaluation stage, 9 patch pairs extracted from each MRI pair, resulting in 2, 040 pairs of LR-HR patches.

2.3 Learning

The learning stage consists of fine-tuning two CNN models that have shown excellent performance for SR of natural images and have been previously tested for MRI SR. The first model is the *Enhanced Deep Residual Networks for Single Image Super-Resolution* (EDSR) (Lim et al., 2017), which ranked first in the NTIRE 2017 super-resolution challenge (Timofte et al., 2017). Some works, including a self-supervised approach, have also relied on the EDSR model to perform SR of structural brain MRI (Zhao et al., 2019, 2021; Fiscone et al., 2024). The second model is the *Real Enhanced Super-Resolution Generative Adversarial Network* (RealESRGAN) (Wang et al., 2021), which is an improved version of the ESRGAN (Wang et al., 2019). The ESRGAN achieved the best perceptual index in the 2018 PIRM challenge on perceptual image super-resolution (Blau et al., 2019). RealESRGAN showed superior performance than the ESRGAN in a recent work evaluating GAN-based approaches for SR of brain and knee MRI (Guerreiro et al., 2023).

2.3.1 CNN architectures

- **EDSR** (Lim et al., 2017): it builds upon the SRResNet (Ledig et al., 2017) by removing the batch normalization

from residual blocks and adding residual scaling to stabilize training. These modifications allow the use of more filters to improve performance without increasing the required computational resources. We used the architectural configuration and pre-trained weights provided by the authors.[1]

- **RealESRGAN** (Wang et al., 2021): its generator is composed by residual-in-residual dense blocks, similar as in ESRGAN (Wang et al., 2019), but the discriminator is a U-Net that provides per-pixel feedback (Schonfeld et al., 2020). We used the architectural configuration and pre-trained weights provided by the authors.[2]

2.3.2 Loss function

The content loss $\mathcal{L}_c$ or objective function to minimize during fine-tuning of the CNN models is the combination of the perceptual loss $\mathcal{L}_{\text{perceptual}}$ and the $\mathcal{L}_1$ loss, therefore:

$$\mathcal{L}_c = \mathcal{L}_{\text{perceptual}} + \eta \mathcal{L}_1$$
$$\mathcal{L}_{\text{EDSR}} = \mathcal{L}_c \tag{2}$$
$$\mathcal{L}_{\text{RealESRGAN}} = \mathcal{L}_c + \lambda\, \mathcal{L}_{\text{GAN}}$$

Where η accounts for the relative importance of the $\mathcal{L}_1$ loss, which measures the image pixel-wise differences. The perceptual loss $\mathcal{L}_{\text{perceptual}}$ compares high-level visual features between the model output and the ground truth. These features can be extracted from a pre-trained network, allowing the knowledge transfer from this loss network to the SR CNN (Johnson et al., 2016). Here, we calculate the perceptual loss using five layers of a pre-trained VGG-19 (Simonyan and Zisserman, 2015), following the approach used in the initial training of the RealESRGAN (Wang et al., 2021). The adversarial loss $\mathcal{L}_{\text{GAN}}$, used during RealESRGAN fine-tuning, is given by the binary cross entropy with sigmoid function, applied to discriminator output. The weighting factor λ modulates the contribution of the adversarial component to the overall loss during generator training.

2.3.3 Other training details

Both models, EDSR and RealESRGAN, were fine-tuned using the ADAM optimizer (Kingma and Ba, 2015) and cosine annealing with warm restarts (Loshchilov and Hutter, 2017) as learning rate scheduler to prevent over-fitting. The $\mathcal{L}_1$ loss weight η was 1, following a previous work that used the same content loss (Wang et al., 2021). The batch size was set to 6. Optimal training hyperparameters were selected using the tree-structured Parzen estimator (Bergstra et al., 2011) within the Optuna (Akiba et al., 2019) framework with the content loss $\mathcal{L}_c$ in the validation set as the objective function. The set of hyperparameters that were optimized included the initial learning rates lr_0, schedulers parameters T_0 and T_{mult}, and for RealESRGAN, the adversarial loss weight λ. The optimal set of hyperparameters for EDSR was: $lr_0 = 2 \times 10^{-4}$, $T_0 = 8$, and $T_{\text{mult}} = 2$. In the case of RealESRGAN, $\lambda = 0.05$, the generator hyperparameters were $lr_0 = 1 \times 10^{-4}$, $T_0 = 6$ and

---

1   https://github.com/sanghyun-son/EDSR-PyTorch/

2   https://github.com/xinntao/Real-ESRGAN/

$T_{\text{mult}} = 1$, the discriminator hyperparameters were $lr_0 = 5 \times 10^{-6}$, $T_0 = 8$, and $T_{\text{mult}} = 2$. Fine-tuning was run for 100 epochs, and the best weights for each model were selected based on the minimum content loss $\mathcal{L}_c$ in the validation set.

2.4 SR reconstruction of MRI volumes

After fine-tuning a CNN model for SR of MRI patches, the next step is to leverage the model for reconstructing HR MRI volumes from single LR MRI inputs. For this purpose, we applied the fine-tuned model to LR slices along each of the two in-plane dimensions using a sliding window approach. Specifically, we take 3 contiguous LR slices as the 3-channel input of the SR model, then we combine the HR outputs corresponding to the same slice location using a weighted average. This sliding window approach mitigates the stacking artifacts that could arise from single-slice inference. As this process is done for each in-plane dimension, the result is a pair of volumes consisting of stacked HR slices. These intermediate volumes are then averaged to produce the final SR reconstructed MRI volume.

2.5 Evaluation

We evaluated MRI SR at two levels: first, a patch-based assessment to evaluate the fine-tuned SR CNN models on unseen MRI patches. Second, a volume-based assessment to evaluate the SR reconstruction of MRI volumes. Additionally, we apply automated lesion segmentation methods to SR reconstructed T2-W FLAIR MRIs to assess the impact of our SR approach on downstream tasks.

2.5.1 SR of MRI patches

For the patch-based assessment, we used paired LR-HR patches extracted from MRIs in the evaluation set. The patches used in this evaluation were also RGB patches extracted following the approach described in subsection 2.2.2. We compared the output of fine-tuned SR models applied to LR patches against ground truth HR patches using an extended set of image quality measures. These included the widely used PSNR and SSIM, as well as five additional measures:

- Visual information fidelity (VIF) (Sheikh and Bovik, 2006), which combines the reference image information and the mutual information between the reference and the distorted image. Its calculation relies on a statistical model for natural scenes, a model for image distortions, and a model of the human visual system.
- Feature similarity index (FSIM) (Zhang et al., 2011), which characterizes the image local quality by combining the image phase congruency and the image gradient magnitude.
- Visual saliency-induced index (VSI) (Zhang et al., 2014), which uses a visual saliency map as a feature to characterize local quality and as a weighting factor when combining it with gradient and chrominance feature maps.

- Haar perceptual similarity index (HaarPSI) (Reisenhofer et al., 2018), which utilizes both high- and low-frequency Haar wavelet coefficients to assess local similarities and weigh local importance. It can be seen as a simplification of the FSIM.
- Deep image structure and texture similarity (DISTS) (Ding et al., 2022), which combines texture similarity and structure similarity, both computed with feature maps extracted from a pre-trained VGG16.

While PSNR and SSIM compare pixel-wise accuracy, the five additional metrics evaluate visual features extracted from the images using hand-crafted filters or pre-trained CNNs. Recent studies (Mason et al., 2020; Kastryulin et al., 2023) show that these visual feature metrics (VIF, FSIM, VSI, HaarPSI, and DISTS) are better correlated with diagnostic quality perceived by radiologists than PSNR and SSIM. This better correlation suggests that visual feature metrics may capture more relevant aspects for diagnostic interpretation than simple pixel-wise comparisons. We computed the metrics for patch evaluation using the PyTorch Image Quality (PIQ) package (Kastryulin et al., 2022).

2.5.2 SR reconstruction of MRI volumes

We evaluated the quality of SR reconstructed MRI volumes with respect to ground truth HR MRI using the PSNR and SSIM. These two measures were chosen for their straightforward applicability to 3D images. We opted to compute these measures within a brain mask to ensure they accurately reflect the quality of SR in diagnostically relevant areas, avoiding inflation from the uniformity of background air. We also included a comparison with two state-of-the-art methods for structural MRI SR that have publicly available implementations:

- SMORE-v4: the *"Synthetic Multi-Orientation Resolution Enhancement"* (Remedios et al., 2023) method is a single-image SR algorithm devised to increase the through-plane resolution of multi-slice MRI. It is based on the self-training of a CNN model for super-resolution using patches extracted from the HR plane. Then, the self-trained models are applied to LR slices to obtain a HR MRI volume.
- SOUP-GAN: the *"Super-resolution Optimized Using Perceptual-tuned GAN"* (Zhang et al., 2022) is a single-image 3D SR framework to produce thinner slices of MRI. In this work, authors trained a GAN using a perceptual loss calculated from slices in the axial, sagittal, and coronal view.

2.5.3 Impact of SR on automated lesion segmentation

The impact of SR on WM lesion segmentation was evaluated by applying two automated lesion segmentation methods to the SR reconstructed T2-W FLAIR volumes. These two methods were chosen because they accept T2-w FLAIR as the only input and are publicly available:

- **LST-lpa** (Schmidt, 2017): the lesion prediction algorithm (lpa) is part of the Lesion Segmentation Toolbox (LST) for SPM. It is a statistical method 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** (Cerri et al., 2021): the lesion segmentation add-on to SAMSEG routine in Freesurfer, it allows the simultaneous segmentation of white matter lesions and 41 structures by decoupling computational models of anatomy from models of the imaging process.

We applied these two segmentation methods also to the LR and the HR T2-weighted FLAIR images, being the latter a reference benchmark for segmentation performance.

Finally, we included the results of applying the recently proposed WMH-SynthSeg (Laso et al., 2024) to the simulated LR T2-w FLAIR images. WMH-SynthSeg is an automated method aiming to segment WM hyper-intensities and 36 anatomical brain regions from MRI of any resolution and contrast. Regardless of the input, WMH-SynthSeg produces a HR segmentation volume with 1 mm isotropic voxels, going directly from LR images to HR lesion segmentation.

3 Results

3.1 SR of MRI patches

The ability of fine-tuned models, namely EDSR and RealESRGAN, to upsample LR MRI patches was evaluated with seven image quality measures: PSNR, SSIM, VIF, FSIM, VSI, HaarPSI, and DISTS. The mean and standard deviation of each image measure along patches extracted from the evaluation set is presented in Table 2. The fine-tuned EDSR is superior to the fine-tuned RealESRGAN in all quality measure for both sequences, T2-W FLAIR and T1-W MRI. Meanwhile, both fine-tuned CNNs outperform bicubic interpolation by at least 1 dB in PSNR and 0.06 in SSIM.

Figure 2 shows four examples of super-resolution of LR patches using the fine-tuned models. Compared to bicubic interpolation, notable improvements are evident in tissue boundaries, sulci shape, and lesion surroundings. It should be noted that MRI patches for evaluation were extracted following the same approach used for training: three contiguous 2D LR patches taken as one RGB image. An extended comparison including examples with the pre-trained models is shown in Supplementary Figure 1.

3.2 SR reconstruction of MRI volumes

The quality of SR reconstructed MRIs was evaluated with the 3D versions of PSNR and SSIM, calculated within a brain mask. Table 3 presents the mean and standard deviation of these metrics for our SR framework, PRETTIER, compared to SMORE and SOUP-GAN. Our approach using the fine-tuned EDSR (PRETTIER-EDSR), consistently outperforms all other methods across all MRI sequences and evaluation sets (see Supplementary Figure 2 for distributions per dataset). PRETTIER with the fine-tuned RealESRGAN yields higher PSNR and SSIM than SMORE when reconstructing T2-W FLAIR images but lower when reconstructing T1-W MRI. Notably, both PRETTIER and SMORE substantially outperform SOUP-GAN.

Qualitative comparisons for T2-W FLAIR and T1-W are shown in Figures 3, 4, respectively. While quantitative results indicate a relatively modest increase of PSNR and SSIM over SMORE, qualitative results show noticeable improvements such as better-defined lesion contours in T2-W FLAIR (Figure 3) and more anatomically coherent tissue boundaries in T1-W MRI (Figure 4). Meanwhile, SOUP-GAN appears to introduce artifacts and textures that are not present in the ground truth HR image, which might explain its lower performance metrics compared to the other methods.

3.3 Effect of SR on automated lesion segmentation

The performance of automated lesion segmentation was quantitatively evaluated in terms of the Dice score, sensitivity (also known as recall), precision, and error of lesion volume estimation. In Table 4, we present the mean and standard deviation of the first three segmentation performance measures for LST-lpa and SAMSEG when applied to LR, SR reconstructed, and HR T2-W FLAIR images, and for WMH-SynthSeg applied only to LR images. Due to non-gaussianity and presence of outliers, we present the distributions of lesion volume estimation errors in Figure 5, a similar plot for Dice score, sensitivity, and precision is presented in Supplementary Figure 3. These results confirm that applying PRETTIER improves lesion segmentation over LR images and brings it closer to segmentation performance in ground truth HR images. Furthermore, in the scenario of segmenting WM lesions on LR multi-slice T2-W FLAIR of PwMS, a better Dice score is achieved by applying first PRETTIER-EDSR and then segmenting lesions with LST-lpa than by applying WMH-SynthSeg directly on the LR image.

Figure 6 shows an example of WM lesion segmentation on T2-W FLAIR, applied to LR inputs, our SR reconstructions, and HR images. This example illustrates how applying SR enhances the automated lesion segmentation when the input is a LR image (red mask in the figure). The automated segmentation on the HR image (green mask) serves as an upper bound, indicating the best segmentation performance attainable for each method, which might still be far from the ground truth segmentation (blue mask).

4 Discussion

In this work, we presented PRETTIER, a framework to enhance the through-plane resolution of multi-slice structural MRIs containing MS lesions. Evaluation results with independent datasets demonstrate successful SR reconstruction which leads to improved performance of automated lesion segmentation.

There are four key aspects that contribute to the successful domain adaptation of 2D CNN models in our approach. First, the content loss function guiding the fine-tuning process promotes the recovery of high-level features with the perceptual loss term, and also the pixel-wise reconstruction accuracy with the $\mathcal{L}_1$ term. This loss formulation, with $\mathcal{L}_1$ weight $\eta = 1$, leads to outputs with high perceptual quality, as shown in the evaluation with metrics beyond PSNR and SSIM (see Table 2), as well as in the qualitative results exhibiting well-defined tissue and lesion boundaries (see

TABLE 2 Mean and standard deviation of image quality measures (IQM) of fine-tuned CNN models, RealESRGAN and EDSR, per MRI sequence.

MRI sequence	IQM	Bicubic interpolation	Fine-tuned RealESRGAN	Fine-tuned EDSR
T2-W FLAIR	PSNR (dB)	22.7 ± 1.2	23.7 ± 1.2	**24.1 ± 1.3**
	SSIM	0.64 ± 0.06	0.70 ± 0.05	**0.72 ± 0.06**
	VIF	0.28 ± 0.04	0.33 ± 0.05	**0.35 ± 0.05**
	FSIM	0.82 ± 0.03	0.87 ± 0.02	**0.88 ± 0.02**
	HaarPSI	0.64 ± 0.03	0.72 ± 0.03	**0.74 ± 0.04**
	VSI	0.92 ± 0.01	0.94 ± 0.01	**0.95 ± 0.01**
	DISTS	0.30 ± 0.02	0.14 ± 0.02	**0.13 ± 0.02**
T1-W	PSNR (dB)	21.0 ± 2.0	23.0 ± 1.8	**23.7 ± 1.9**
	SSIM	0.71 ± 0.05	0.79 ± 0.05	**0.81 ± 0.05**
	VIF	0.32 ± 0.05	0.39 ± 0.05	**0.41 ± 0.06**
	FSIM	0.83 ± 0.03	0.88 ± 0.03	**0.89 ± 0.03**
	HaarPSI	0.65 ± 0.05	0.75 ± 0.04	**0.77 ± 0.04**
	VSI	0.93 ± 0.02	0.95 ± 0.01	**0.96 ± 0.01**
	DISTS	0.28 ± 0.03	0.13 ± 0.02	**0.12 ± 0.02**

Results of bicubic interpolation are also included for comparison. The best values for each metric are indicated in bold.

TABLE 3 Mean and standard deviation of PSNR and SSIM for MRI SR reconstruction methods, calculated within a brain mask.

MRI sequence	IQM	SMORE	SOUP-GAN	PRETTIER– RealESRGAN	PRETTIER– EDSR
T2-W FLAIR	PSNR brain	24.6 ± 1.3	21.6 ± 1.3	25.0 ± 1.4	**25.5 ± 1.4**
	SSIM brain	0.76 ± 0.04	0.65 ± 0.03	0.78 ± 0.03	**0.80 ± 0.04**
T1-W	PSNR brain	27.4 ± 1.7	23.7 ± 1.7	26.7 ± 1.7	**28.5 ± 1.8**
	SSIM brain	0.86 ± 0.03	0.76 ± 0.02	0.84 ± 0.04	**0.89 ± 0.03**

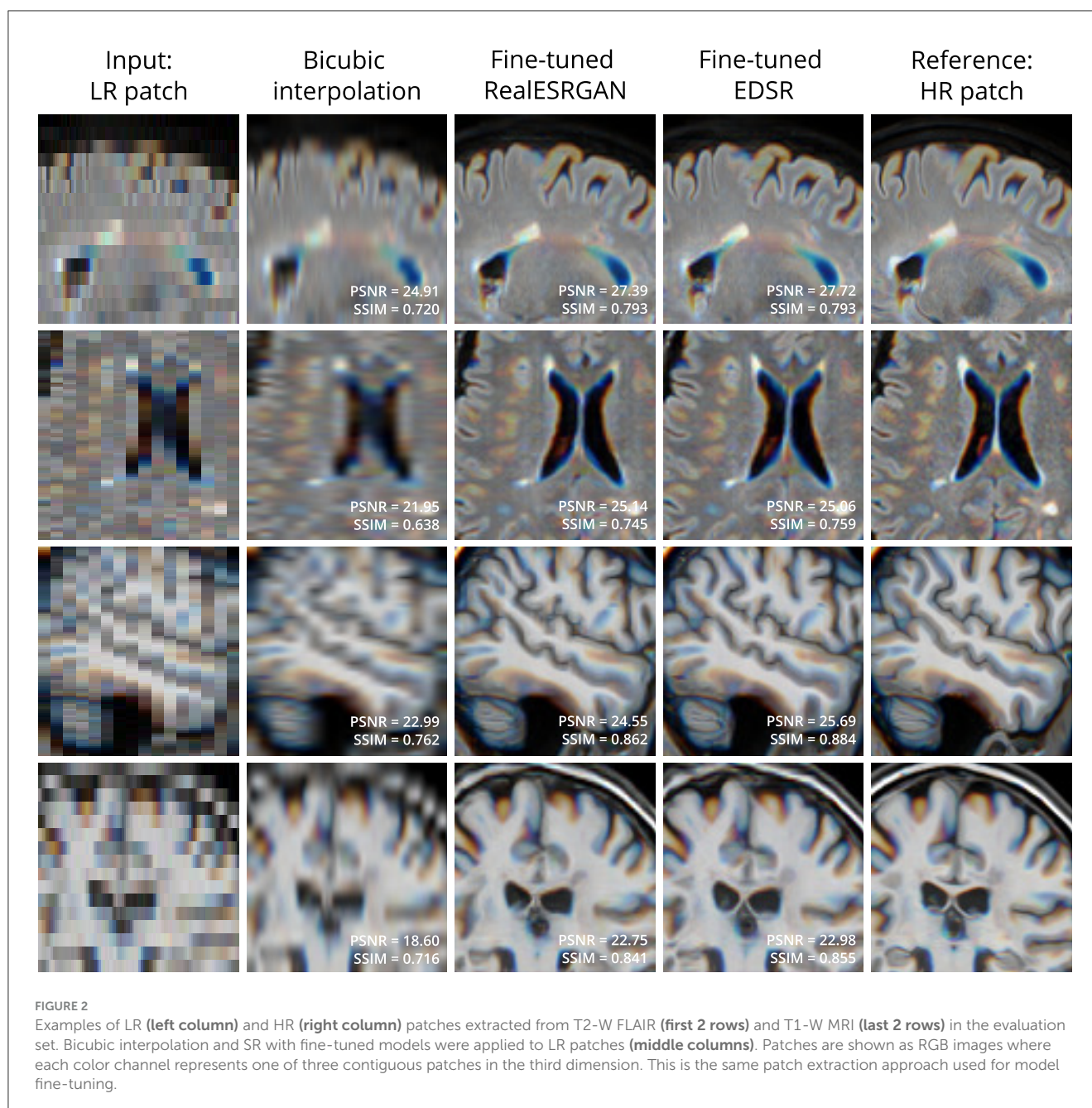The best values for each metric are indicated in bold.

TABLE 4 Mean and standard deviation of Dice score, sensitivity, and precision calculated for automated lesion segmentation on T2-W FLAIR.

Method	Input	Dice score	Sensitivity	Precision
LST-lpa	LR	0.47 ± 0.15	0.59 ± 0.17	0.47 ± 0.22
	PRETTIER-RealESRGAN	0.52 ± 0.15	0.60 ± 0.19	0.55 ± 0.22
	PRETTIER-EDSR	0.53 ± 0.15	0.60 ± 0.18	0.55 ± 0.22
	HR	0.58 ± 0.16	0.69 ± 0.16	0.57 ± 0.23
SAMSEG	LR	0.41 ± 0.17	0.32 ± 0.16	0.67 ± 0.16
	PRETTIER-RealESRGAN	0.46 ± 0.21	0.38 ± 0.21	0.71 ± 0.16
	PRETTIER-EDSR	0.47 ± 0.21	0.40 ± 0.21	0.71 ± 0.17
	HR	0.53 ± 0.20	0.43 ± 0.19	0.79 ± 0.14
WMH-SynthSeg	LR	0.35 ± 0.14	0.37 ± 0.14	0.40 ± 0.23

LST-lpa (Schmidt, 2017) and SAMSEG (Cerri et al., 2021) were applied to simulated LR images, the outputs of our SR framework (PRETTIER), and to ground truth HR images. WMH-SynthSeg (Laso et al., 2024) was only applied to LR images as its purpose is to directly produce HR segmentations. These results show that our SR framework improves the segmentation performance over LR images, approaching the performance of segmentation over HR images.

Figure 2 and Supplementary Figure 1). Second, we use patches extracted from two different MRI modalities (T2-W FLAIR and T1-W MRI) with three different slice orientations (axial, sagittal, and coronal). This variability in our training dataset exposes the models to a wide range of anatomical and contrast variations, potentially enhancing their generalizability, as suggested by the results in our similarly diverse evaluation dataset. Third, instead of using simple downsampling or k-space truncation as done in some existing literature (Sanchez and Vilaplana, 2018; Pham et al., 2019; Du et al., 2020), we obtain pairs of LR-HR images by applying a physics-informed model of multi-slice MRI acquisition to HR images which takes into account the slice selection profile.

FIGURE 2
Examples of LR **(left column)** and HR **(right column)** patches extracted from T2-W FLAIR **(first 2 rows)** and T1-W MRI **(last 2 rows)** in the evaluation set. Bicubic interpolation and SR with fine-tuned models were applied to LR patches **(middle columns)**. Patches are shown as RGB images where each color channel represents one of three contiguous patches in the third dimension. This is the same patch extraction approach used for model fine-tuning.
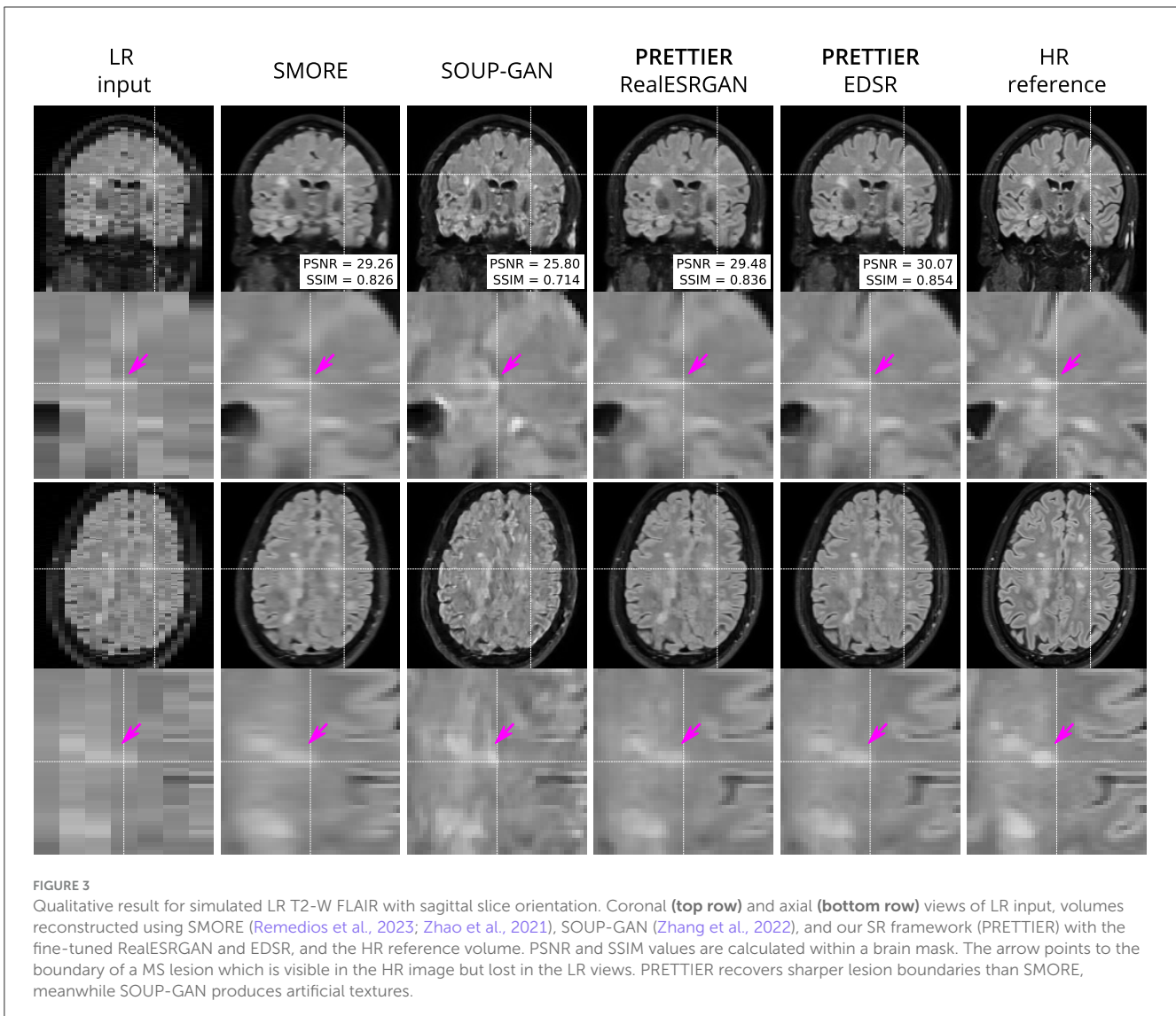
Fourth, we incorporate information from adjacent slices as color channels in the inputs and outputs of 2D CNN models, allowing them to leverage the 3D information in MRI while benefiting from architectural advances in natural image SR.

Two considerations led us to employ 2D instead of 3D CNN architectures. First, 2D models have lower computational costs during training and inference compared to their 3D counterparts. Second, adopting a 2D approach offers greater flexibility in leveraging advances from the vast literature on natural image SR. Many of the cutting-edge architectures and strategies for SR have been primarily developed and optimized for 2D images. By operating with 2D models, we could adapt and fine-tune any of these models to our MRI SR framework. Specifically, in this work we used two SR models: the EDSR (Lim et al., 2017) and the RealESRGAN (Wang et al., 2021).

Patch-based evaluation of fine-tuned models demonstrates that EDSR consistently outperforms RealESRGAN across datasets, MRI contrasts and evaluation metrics (see Table 2). It is worth noting the architectural differences between these models. EDSR has 32 residual blocks with 256 features in each convolutional layer, amounting to over 40 million trainable parameters. Meanwhile, the generator in RealESRGAN has 23 residual-in-residual dense blocks with 64 initial features per residual dense block, resulting in ∼16.7 million parameters, a lighter model size than EDSR.
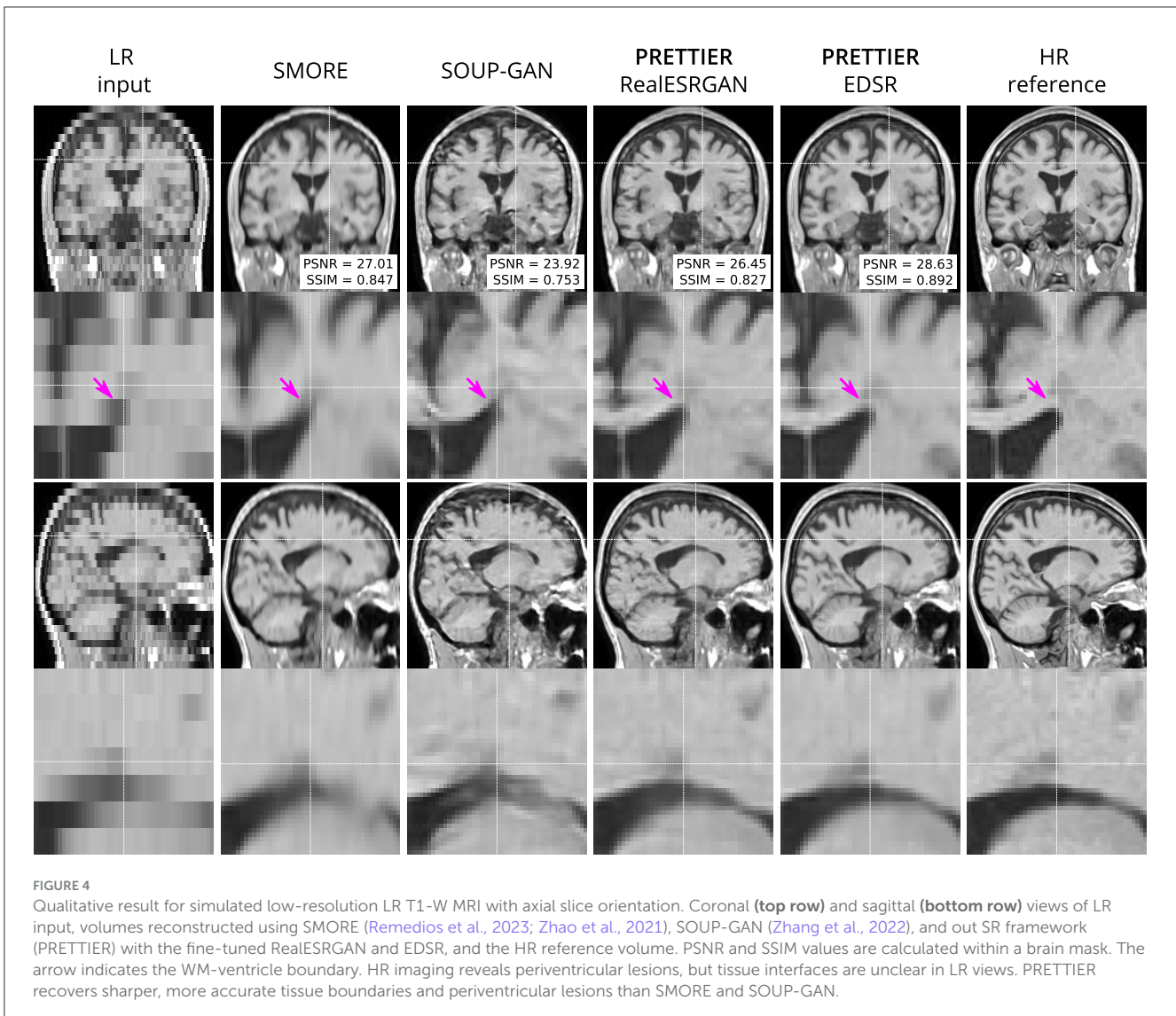
We quantitatively evaluate our framework for SR reconstruction of MRI volumes using PSNR and SSIM metrics, comparing against two state-of-the-art methods for MRI SR: SMORE (Remedios et al., 2023; Zhao et al., 2021) and SOUP-GAN (Zhang et al., 2022). The key feature of SMORE is its self-supervised training without relying on external data. However, this comes

FIGURE 3
Qualitative result for simulated LR T2-W FLAIR with sagittal slice orientation. Coronal **(top row)** and axial **(bottom row)** views of LR input, volumes reconstructed using SMORE (Remedios et al., 2023; Zhao et al., 2021), SOUP-GAN (Zhang et al., 2022), and our SR framework (PRETTIER) with the fine-tuned RealESRGAN and EDSR, and the HR reference volume. PSNR and SSIM values are calculated within a brain mask. The arrow points to the boundary of a MS lesion which is visible in the HR image but lost in the LR views. PRETTIER recovers sharper lesion boundaries than SMORE, meanwhile SOUP-GAN produces artificial textures.

at a significant computational cost as a CNN model must be trained for each new input image. In contrast, our approach leverages trained models that have been fine-tuned with data we gathered from external datasets, allowing faster and less resource demanding application. Evaluation on the independent dataset shows our fine-tuned EDSR model outperforms SMORE across metrics for both T2-W FLAIR and T1-W MRI (see Table 3 and Supplementary Figure 2). While those quantitative results also show our approach with the fine-tuned RealESRGAN slightly underperforms SMORE on T1-W, the qualitative example (Figure 4) reveals sharper tissue boundaries more alike the HR ground truth. Our SR framework shares some similarities with SOUP-GAN (Zhang et al., 2022), as both approaches rely on models trained with a perceptual loss. Specifically, SOUP-GAN employs a scale-attention architecture trained via an adversarial approach. However, our evaluation results demonstrate that SOUP-GAN underperforms quantitatively and qualitatively compared to both SMORE and our framework using the fine-tuned EDSR and RealESRGAN models. The qualitative examples in Figures 3, 4 suggest that SOUP-GAN suffers from artificial textures, artifacts that are likely introduced when promoting only the perceptual

quality of images (via perceptual loss and adversarial training) without accounting for reconstruction accuracy.

White matter lesion segmentation is a highly relevant task when processing brain MRI data of PwMS. Assessing the impact of our SR framework on this task is crucial for validating its practical use. For this evaluation, we applied two different automated methods for lesion segmentation, LST-lpa (Schmidt, 2017) and SAMSEG (Cerri et al., 2021), on T2-W FLAIR images. We compared the segmentation performance when using the LR images versus the SR reconstructed images, and also include the segmentation performance on HR images as reference. The results demonstrate that, compared to segmentation on LR images, our SR reconstruction approach improves the Dice score, sensitivity, precision, and lesion volume estimation, bringing them closer to what is achievable with HR images (Table 4 and Figure 5). Consistently, we observe that LST-lpa exhibits higher sensitivity but lower precision than SAMSEG when applied to LR and HR T2-W FLAIR images. Notably, our SR approach improves the precision of LST-lpa without compromising sensitivity, suggesting it effectively refines lesion boundaries, as illustrated in Figure 6. Conversely, SR enhances the low sensitivity of SAMSEG while
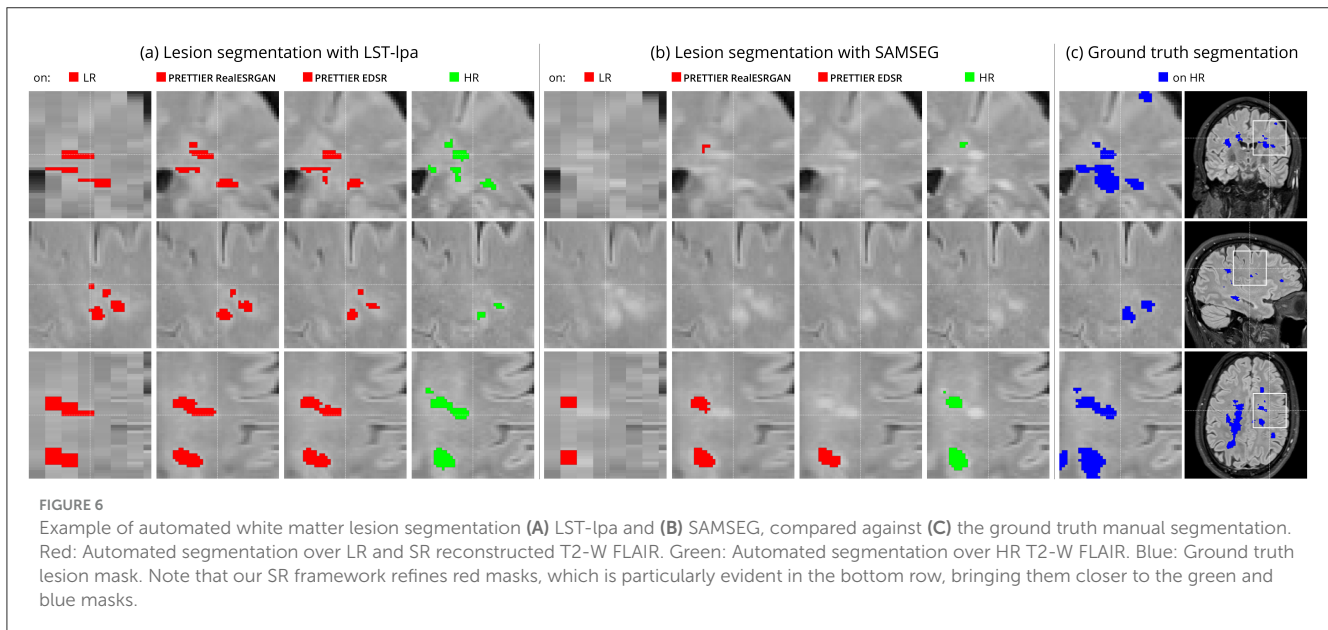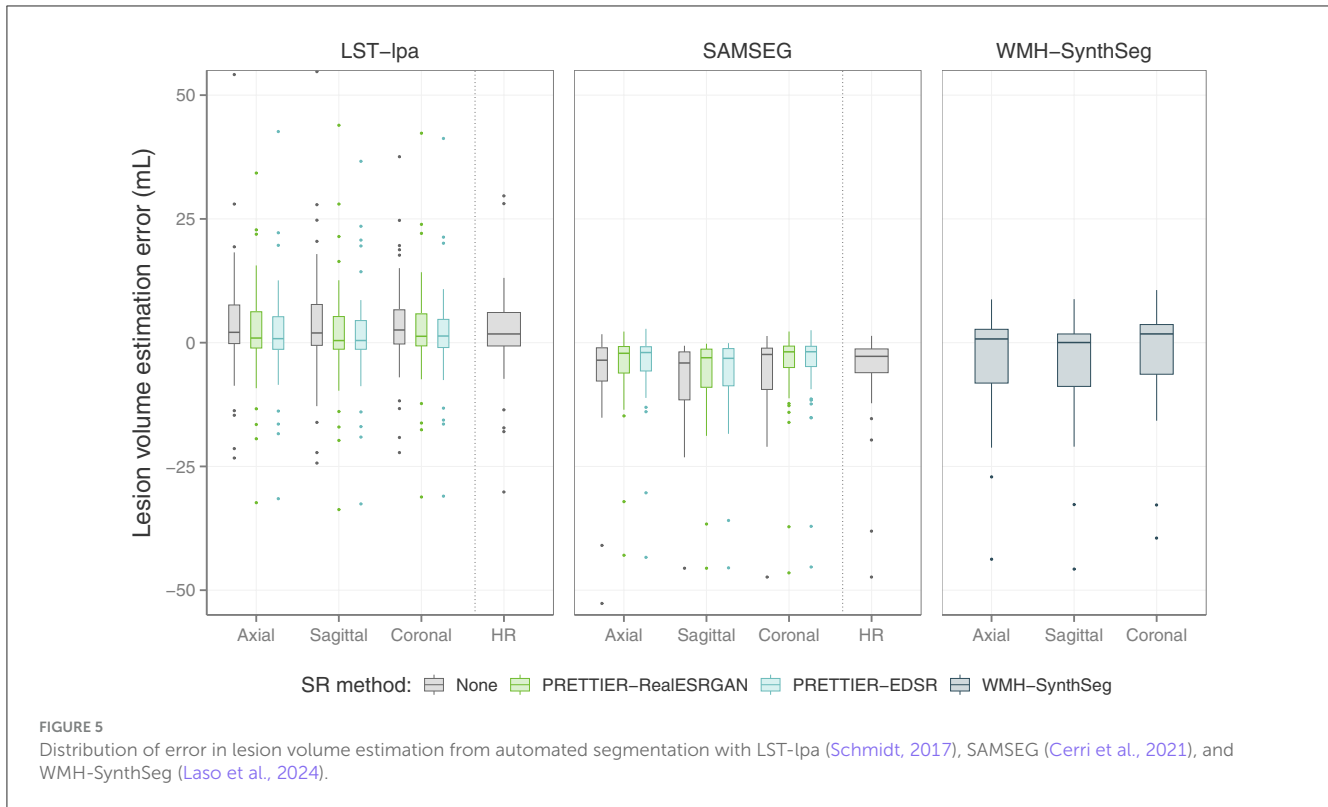
FIGURE 4
Qualitative result for simulated low-resolution LR T1-W MRI with axial slice orientation. Coronal **(top row)** and sagittal **(bottom row)** views of LR input, volumes reconstructed using SMORE (Remedios et al., 2023; Zhao et al., 2021), SOUP-GAN (Zhang et al., 2022), and out SR framework (PRETTIER) with the fine-tuned RealESRGAN and EDSR, and the HR reference volume. PSNR and SSIM values are calculated within a brain mask. The arrow indicates the WM-ventricle boundary. HR imaging reveals periventricular lesions, but tissue interfaces are unclear in LR views. PRETTIER recovers sharper, more accurate tissue boundaries and periventricular lesions than SMORE and SOUP-GAN.

also improving its precision. Furthermore, we also include a comparison against the recently proposed WMH-SynthSeg (Laso et al., 2024), a method aiming to produce a HR segmentations of WM hyperintensities (and 36 brain regions) from scans of any resolution and contrast. Our evaluation shows that, given a LR T2-W FLAIR image (acquired in a 1.5 T or 3 T scanner), applying our SR approach followed by LST-lpa or SAMSEG yields superior lesion segmentation compared to directly applying WMH-SynthSeg on the LR images.

The work presented herein has some limitations. First, LR MRIs used for model fine-tuning and evaluation are simulated using only one slice profile: slice thickness of 5 mm and slice spacing of 6 mm (i.e., 1 mm of slice gap). While our evaluation results demonstrate the capabilities of our SR framework in this challenging and common clinical scenario, and preliminary results show good performance on images with different slice profiles (see Supplementary Figure 5), future work should evaluate its performance across a broader range of acquisition settings. Second, the computational requirements of using deep CNN models can pose barriers to their implementation, especially in resource-limited settings. To address this, future research will explore the

capabilities of more efficient SR models (Li et al., 2023; Ren et al., 2024). Lastly, we evaluated the impact of our SR framework on only one downstream task, the automated segmentation of WM lesions. Expanding this evaluation to other downstream tasks in MS neuroimaging analyzes, such as regional volumetry and radiomic feature extraction, would provide a more comprehensive assessment of SR potential benefits and limitations in MS research.

5 Conclusion

We have presented PRETTIER, a single-image SR framework for multi-slice structural MRI of PwMS that leverages existing CNN architectures for image SR. Our framework demonstrates superior image quality results than existing methods for MRI SR, and improves the automated lesion segmentation on LR T2-W FLAIR. By effectively addressing the limitations of routinely acquired multi-slice MRI with low through-plane resolution, our approach facilitates the use of retrospective MRI datasets already acquired in the clinics to conduct 3D analyzes and investigate image-based biomarkers of MS outcomes.

FIGURE 5
Distribution of error in lesion volume estimation from automated segmentation with LST-lpa (Schmidt, 2017), SAMSEG (Cerri et al., 2021), and WMH-SynthSeg (Laso et al., 2024).

FIGURE 6
Example of automated white matter lesion segmentation **(A)** LST-lpa and **(B)** SAMSEG, compared against **(C)** the ground truth manual segmentation. Red: Automated segmentation over LR and SR reconstructed T2-W FLAIR. Green: Automated segmentation over HR T2-W FLAIR. Blue: Ground truth lesion mask. Note that our SR framework refines red masks, which is particularly evident in the bottom row, bringing them closer to the green and blue masks.

Data availability statement

Publicly available datasets were analyzed in this study. Data used in this work included four publicly available datasets: ISBI2015, https://smart-stats-tools.org/lesion-challenge; Lesjak-3D, https://lit.fe.uni-lj.si/en/research/resources/3D-MR-MS/; MSSEG1, https://portal.fli-iam.irisa.fr/msseg-challenge/; MSSEG2, https://portal.fli-iam.irisa.fr/msseg-2/.

Ethics statement

The studies involving humans were approved by the ethical commission of the University of Hasselt (CME2019/046) and the Ethics Committee of Hospital Universitario Nacional (CEI-HUN-2019-12). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin

in accordance with the national legislation and institutional requirements.

Author contributions

DG: Conceptualization, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. HK: Data curation, Investigation, Validation, Writing – review & editing. GP: Data curation, Investigation, Validation, Writing – review & editing. ZL: Methodology, Software, Writing – review & editing. AL-C: Data curation, Investigation, Validation, Writing – review & editing, Resources. BV: Data curation, Resources, Writing – review & editing. HW: Resources, Supervision, Writing – review & editing. PL: Resources, Supervision, Writing – review & editing. ER: Resources, Supervision, Writing – review & editing. LP: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. JS: Funding acquisition, Resources, Supervision, Writing – review & editing.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2024. 1473132/full#supplementary-material

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). "Optuna: a next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: Association for Computing Machinery), 2623–2631. doi: 10.1145/3292500.3330701

Beirinckx, Q., Jeurissen, B., Nicastro, M., Poot, D. H., Verhoye, M., den Dekker, A. J., et al. (2022). Model-based super-resolution reconstruction with joint motion estimation for improved quantitative MRI parameter mapping. *Comput. Med. Imag. Graph.* 100:102071. doi: 10.1016/j.compmedimag.2022.102071

Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B. (2011). "Algorithms for hyper-parameter optimization," in *Advances in Neural Information Processing Systems, Vol. 24*, eds. J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger (Red Hook, NY: Curran Associates, Inc.),1–9.

Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., and Zelnik-Manor, L. (2019). "The 2018 PIRM challenge on perceptual image super-resolution," in *Computer Vision—ECCV 2018 Workshops*, eds. L. Leal-Taixé and S. Roth (Cham: Springer International Publishing), 334–355.

Carass, A., Roy, S., Gherman, A., Reinhold, J. C., Jesson, A., Arbel, T., et al. (2020). Evaluating white matter lesion segmentations with refined sørensen-dice analysis. *Sci. Rep.* 10:64803. doi: 10.1038/s41598-020-64803-w

Carass, A., Roy, S., Jog, A., Cuzzocreo, J. L., Magrath, E., Gherman, A., et al. (2017). Longitudinal multiple sclerosis lesion segmentation: resource

and challenge. *NeuroImage* 148, 77–102. doi: 10.1016/j.neuroimage.2016. 12.064

Cerri, S., Puonti, O., Meier, D. S., Wuerfel, J., Mühlau, M., Siebner, H. R., et al. (2021). A contrast-adaptive method for simultaneous whole-brain and lesion segmentation in multiple sclerosis. *NeuroImage* 225:117471. doi: 10.1016/j.neuroimage.2020.117471

Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M. (2021a). "MSSEG-2 challenge proceedings: Multiple sclerosis new lesions segmentation challenge using a data management and processing infrastructure," in *MICCAI 2021—24th International Conference on Medical Image Computing and Computer Assisted Intervention* (Strasbourg), 12.

Commowick, O., Kain, M., Casey, R., Ameli, R., Ferré, J.-C., Kerbrat, A., et al. (2021b). Multiple sclerosis lesions segmentation from multiple experts: the MICCAI 2016 challenge dataset. *NeuroImage* 244:118589. doi: 10.1016/j.neuroimage.2021.118589

Danelakis, A., Theoharis, T., and Verganelakis, D. A. (2018). Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging. *Comput. Med. Imag. Graph.* 70, 83–100. doi: 10.1016/j.compmedimag.2018. 10.002

Ding, K., Ma, K., Wang, S., and Simoncelli, E. P. (2022). Image quality assessment: unifying structure and texture similarity. *IEEE Trans. Pat. Anal. Machine Intell.* 44, 2567–2581. doi: 10.1109/TPAMI.2020.3045810

Dong, C., Loy, C. C., He, K., and Tang, X. (2014). "Learning a deep convolutional network for image super-resolution," in *Computer Vision—ECCV 2014*, eds. D. Fleet, T. Pajdla, B. Schiele and T.Tuytelaars (Cham: Springer International Publishing), 184–199.

Du, J., He, Z., Wang, L., Gholipour, A., Zhou, Z., Chen, D., et al. (2020). Super-resolution reconstruction of single anisotropic 3D MR images using residual convolutional neural network. *Neurocomputing* 392, 209–220. doi: 10.1016/j.neucom.2018.10.102

Fiscone, C., Curti, N., Ceccarelli, M., Remondini, D., Testa, C., Lodi, R., et al. (2024). Generalizing the enhanced-deep-super-resolution neural network to brain MR images: a retrospective study on the cam-can dataset. *eNeuro* 11:2023. doi: 10.1523/ENEURO.0458-22.2023

Gholipour, A., Afacan, O., Aganj, I., Scherrer, B., Prabhu, S. P., Sahin, M., et al. (2015). Super-resolution reconstruction in frequency, image, and wavelet domains to reduce through-plane partial voluming in MRI. *Med. Phys.* 42, 6919–6932. doi: 10.1118/1.4935149

Giraldo, D. L., Beirinckx, Q., Den Dekker, A. J., Jeurissen, B., and Sijbers, J. (2023). "Super-resolution reconstruction of multi-slice T2-W FLAIR MRI improves multiple sclerosis lesion segmentation," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Sydney, NSW), 1–4. doi: 10.1109/EMBC40787.2023.10341047

Guerreiro, J., Tomás, P., Garcia, N., and Aidos, H. (2023). Super-resolution of magnetic resonance images using generative adversarial networks. *Comput. Med. Imag. Graph.* 108:102280. doi: 10.1016/j.compmedimag.2023.102280

Han, S., Remedios, S. W., Schär, M., Carass, A., and Prince, J. L. (2023). ESPRESO: an algorithm to estimate the slice profile of a single magnetic resonance image. *Magnet. Reson. Imag.* 98, 155–163. doi: 10.1016/j.mri.2023.01.012

Iglesias, J. E., Billot, B., Balbastre, Y., Magdamo, C., Arnold, S. E., Das, S., et al. (2023). Synthsr: A public AI tool to turn heterogeneous clinical brain scans into high-resolution T1-weighted images for 3D morphometry. *Sci. Adv.* 9:eadd3607. doi: 10.1126/sciadv.add3607

Iglesias, J. E., Billot, B., Balbastre, Y., Tabari, A., Conklin, J., Gilberto González, R., et al. (2021). Joint super-resolution and synthesis of 1 mm isotropic MP-RAGE volumes from clinical MRI exams with scans of different orientation, resolution and contrast. *NeuroImage* 237:118206. doi: 10.1016/j.neuroimage.2021.118206

Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., et al. (2019). Automated brain extraction of multisequence MRI using artificial neural networks. *Hum. Brain Map.* 40, 4952–4964. doi: 10.1002/hbm.24750

Johnson, J., Alahi, A., and Fei-Fei, L. (2016). "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision—ECCV 2016*, eds. B. Leibe, J. Matas, N. Sebe, and M. Welling (Cham: Springer International Publishing), 694–711.

Kastryulin, S., Zakirov, J., Pezzotti, N., and Dylov, D. V. (2023). Image quality assessment for magnetic resonance imaging. *IEEE Access* 11, 14154–14168. doi: 10.1109/ACCESS.2023.3243466

Kastryulin, S., Zakirov, J., Prokopenko, D., and Dylov, D. V. (2022). *Pytorch Image Quality: Metrics for Image Quality Assessment*. doi: 10.48550/arXiv.2208.14818

Kingma, D. P., and Ba, J. (2015). *Adam: A Method for Stochastic Optimization*. doi: 10.48550/arXiv.1412.6980

Kolb, H., Al-Louzi, O., Beck, E. S., Sati, P., Absinta, M., and Reich, D. S. (2022). From pathology to MRI and back: clinically relevant biomarkers of multiple sclerosis lesions. *NeuroImage* 36:103194. doi: 10.1016/j.nicl.2022.103194

Laso, P., Cerri, S., Sorby-Adams, A., Guo, J., Mateen, F., Goebl, P., et al. (2024). "Quantifying white matter hyperintensity and brain volumes in heterogeneous clinical and low-field portable MRI," in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)* (Athens), 1–5. doi: 10.1109/ISBI56570.2024.10635502

Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., et al. (2017). "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Los Alamitos, CA: IEEE Computer Society), 105–114.

Lesjak, Ž., Galimzianova, A., Koren, A., Lukin, M., Pernuš, F., Likar, B., et al. (2017). A novel public MR image dataset of multiple sclerosis patients with lesion segmentations based on multi-rater consensus. *Neuroinformatics* 16, 51–63. doi: 10.1007/s12021-017-9348-2

Li, Y., Zhang, Y., Timofte, R., Van Gool, L., Yu, L., Li, Y., et al. (2023). "NTIRE 2023 challenge on efficient super-resolution: methods and results," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (Vancouver, BC), 1922–1960.

Lim, B., Son, S., Kim, H., Nah, S., and Lee, K. M. (2017). "Enhanced deep residual networks for single image super-resolution," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (Honolulu, HI), 1132–1140.

Loshchilov, I., and Hutter, F. (2017). *SGDR: Stochastic Gradient Descent With Warm Restarts*. doi: 10.48550/arXiv.1608.03983

Manjón, J. V., Coupé, P., Martí-Bonmatí, L., Collins, D. L., and Robles, M. (2010). Adaptive non-local means denoising of MR images with spatially varying noise levels. *J. Magnet. Reson. Imag.* 31, 192–203. doi: 10.1002/jmri.22003

Mason, A., Rioux, J., Clarke, S. E., Costa, A., Schmidt, M., Keough, V., et al. (2020). Comparison of objective image quality metrics to expert radiologists' scoring of diagnostic quality of MR images. *IEEE Trans. Med. Imag.* 39, 1064–1072. doi: 10.1109/TMI.2019.2930338

Mendelsohn, Z., Pemberton, H. G., Gray, J., Goodkin, O., Carrasco, F. P., Scheel, M., et al. (2023). Commercial volumetric MRI reporting tools in multiple sclerosis: a systematic review of the evidence. *Neuroradiology* 65, 5–24. doi: 10.1007/s00234-022-03074-w

Pham, C.-H., Ducournau, A., Fablet, R., and Rousseau, F. (2017). "Brain MRI super-resolution using deep 3D convolutional networks," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (Melbourne, VIC), 197–200.

Pham, C.-H., Tor-Díez, C., Meunier, H., Bednarek, N., Fablet, R., Passat, N., et al. (2019). Multiscale brain MRI super-resolution using deep 3D convolutional networks. *Comput. Med. Imag. Graph.* 77:101647. doi: 10.1016/j.compmedimag.2019.101647

Poot, D. H. J., Van Meir, V., and Sijbers, J. (2010). "General and efficient super-resolution method for multi-slice MRI," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*, eds. T. Jiang, N. Navab, J. P. W. Pluim, and M. A. Viergever (Berlin; Heidelberg: Springer Berlin Heidelberg), 615–622.

Reisenhofer, R., Bosse, S., Kutyniok, G., and Wiegand, T. (2018). A Haar wavelet-based perceptual similarity index for image quality assessment. *Sign. Process.* 61, 33–43. doi: 10.1016/j.image.2017.11.001

Remedios, S. W., Han, S., Zuo, L., Carass, A., Pham, D. L., Prince, J. L., et al. (2023). "Self-supervised super-resolution for anisotropic MR images with and without slice gap," in *Simulation and Synthesis in Medical Imaging*, eds. J. M. Wolterink, D. Svoboda, C.Zhao, and V. Fernandez (Cham: Springer Nature Switzerland), 118–128.

Ren, B., Li, Y., Mehta, N., Timofte, R., Yu, H., Wan, C., et al. (2024). *The Ninth NTIRE 2024 Efficient Super-Resolution Challenge Report*. doi: 10.48550/arXiv.2404.10343

Sanchez, I., and Vilaplana, V. (2018). Brain MRI super-resolution using 3D generative adversarial networks. *Med. Imag. Deep Learn.* 2018, 1–8. doi: 10.48550/arXiv.1812.11440

Schmidt, P. (2017). *Bayesian Inference for Structured Additive Regression Models for Large-Scale Problems With Applications to Medical Imaging* (Ph. D. thesis). Ludwig-Maximilians-Universität München, München, Germany.

Schonfeld, E., Schiele, B., and Khoreva, A. (2020). "A U-Net based discriminator for generative adversarial networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Seattle, WA: IEEE Computer Society), 8207–8216. doi: 10.1109/CVPR42600.2020.00823

Sheikh, H., and Bovik, A. (2006). Image information and visual quality. *IEEE Trans. Image Process.* 15, 430–444. doi: 10.1109/TIP.2005.859378

Shilling, R. Z., Robbie, T. Q., Bailloeul, T., Mewes, K., Mersereau, R. M., and Brummer, M. E. (2009). A super-resolution framework for 3-D high-resolution and high-contrast imaging using 2-D multislice MRI. *IEEE Trans. Med. Imag.* 28, 633–644. doi: 10.1109/TMI.2008.2007348

Simonyan, K., and Zisserman, A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. doi: 10.48550/arXiv.1409.1556

Timofte, R., Agustsson, E., Van Gool, L., Yang, M.-H., and Zhang, L. (2017). "NTIRE 2017 challenge on single image super-resolution: methods and results," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (Honolulu, HI: IEEE Computer Society), 1110–1121. doi: 10.1109/CVPRW.2017.149

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., et al. (2010). N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imag.* 29, 1310–1320. doi: 10.1109/tmi.2010.2046908

Wang, X., Xie, L., Dong, C., and Shan, Y. (2021). "Real-ESRGAN: training real-world blind super-resolution with pure synthetic data," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* (Montreal, BC), 1905–1914.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., et al. (2019). "ESRGAN: enhanced super-resolution generative adversarial networks," in *Computer Vision—ECCV 2018 Workshops*, eds. L. Leal-Taixé and S. Roth (Cham: Springer International Publishing), 63–79.

Wattjes, M. P., Ciccarelli, O., Reich, D. S., Banwell, B., de Stefano, N., Enzinger, C., et al. (2021). 2021 MAGNIMS–CMSC–NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. *Lancet Neurol.* 20, 653–670. doi: 10.1016/S1474-4422(21)00095-8

Zhang, K., Hu, H., Philbrick, K., Conte, G. M., Sobek, J. D., Rouzrokh, P., et al. (2022). SOUP-GAN: super-resolution MRI using generative adversarial networks. *Tomography* 8, 905–919. doi: 10.3390/tomography8020073

Zhang, L., Shen, Y., and Li, H. (2014). VSI: a visual saliency-induced index for perceptual image quality assessment. *IEEE Trans. Image Process.* 23, 4270–4281. doi: 10.1109/TIP.2014.2346028

Zhang, L., Zhang, L., Mou, X., and Zhang, D. (2011). FSIM: a feature similarity index for image quality assessment. *IEEE Trans. Image Process.* 20, 2378–2386. doi: 10.1109/TIP.2011.2109730

Zhao, C., Dewey, B. E., Pham, D. L., Calabresi, P. A., Reich, D. S., and Prince, J. L. (2021). SMORE: a self-supervised anti-aliasing and super-resolution algorithm for MRI using deep learning. *IEEE Trans. Med. Imag.* 40, 805–817. doi: 10.1109/TMI.2020.3037187

Zhao, C., Shao, M., Carass, A., Li, H., Dewey, B. E., Ellingsen, L. M., et al. (2019). Applications of a deep learning method for anti-aliasing and super-resolution in MRI. *Magnet. Reson. Imag.* 64, 132–141. doi: 10.1016/j.mri.2019.05.038