



MR-self Noise2Noise: self-supervised deep learning-based image quality improvement of submillimeter resolution 3D MR images

Woojin Jung¹ · Hyun-Soo Lee² · Minkook Seo³ · Yoonho Nam⁴ · Yangsean Choi³ · Na-Young Shin³ · Kook-Jin Ahn³ · Bum-soo Kim³ · Jinhee Jang³

Received: 11 July 2022 / Revised: 28 September 2022 / Accepted: 14 October 2022
© The Author(s), under exclusive licence to European Society of Radiology 2022

Abstract

Objectives The study aimed to develop a deep neural network (DNN)-based noise reduction and image quality improvement by only using routine clinical scans and evaluate its performance in 3D high-resolution MRI.

Methods This retrospective study included T1-weighted magnetization-prepared rapid gradient-echo (MP-RAGE) images from 185 clinical scans: 135 for DNN training, 11 for DNN validation, 20 for qualitative evaluation, and 19 for quantitative evaluation. Additionally, 18 vessel wall imaging (VWI) data were included to evaluate generalization. In each scan of the DNN training set, two noise-independent images were generated from the k-space data, resulting in an input-label pair. 2.5D U-net architecture was utilized for the DNN model. Qualitative evaluation between conventional MP-RAGE and DNN-based MP-RAGE was performed by two radiologists in image quality, fine structure delineation, and lesion conspicuity. Quantitative evaluation was performed with full sampled data as a reference by measuring quantitative error metrics and volumetry at 7 different simulated noise levels. DNN application on VWI was evaluated by two radiologists in image quality.

Results Our DNN-based MP-RAGE outperformed conventional MP-RAGE in all image quality parameters (average scores = 3.7 vs. 4.9, $p < 0.001$). In the quantitative evaluation, DNN showed better error metrics ($p < 0.001$) and comparable ($p > 0.09$) or better ($p < 0.02$) volumetry results than conventional MP-RAGE. DNN application to VWI also revealed improved image quality (3.5 vs. 4.6, $p < 0.001$).

Conclusion The proposed DNN model successfully denoises 3D MR image and improves its image quality by using routine clinical scans only.

Key Points

- Our deep learning framework successfully improved conventional 3D high-resolution MRI in all image quality parameters, fine structure delineation, and lesion conspicuity.
- Compared to conventional MRI, the proposed deep neural network-based MRI revealed better quantitative error metrics and comparable or better volumetry results.
- Deep neural network application to 3D MRI whose pulse sequences and parameters were different from the training data showed improvement in image quality, revealing the potential to generalize on various clinical MRI.

Keywords Magnetic resonance imaging · Quality improvement · High-resolution · Deep learning

Abbreviations

✉ Jinhee Jang znee@catholic.ac.kr	DNN	Deep neural network
¹ AIRS Medical, Seoul, Republic of Korea	DSC	Dice similarity coefficient
² Siemens Healthineers Ltd, Seoul, Republic of Korea	FOV	Field of view
³ Department of Radiology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul 06591, Republic of Korea	MP-RAGE	Magnetization-prepared rapid gradient-echo
⁴ Division of Biomedical Engineering, Hankuk University of Foreign Studies, Yongin, Gyeonggi-do, Republic of Korea	MRI	Magnetic resonance imaging
	NRMSE	Normalized root mean square error
	PSNR	Peak signal-to-noise ratio
	ROI	Region of interest
	sN2N	Self Noise2Noise

SNR	Signal-to-noise ratio
SSIM	Structural similarity index measure
VWI	Vessel wall imaging

Introduction

Thanks to recent advances, submillimeter-resolution brain images can be achieved using modern clinical magnetic resonance imaging (MRI) [1–3]. However, to achieve a sufficient signal-to-noise ratio (SNR) for submillimeter MR images, a long scan time is required, which is barely possible in most clinical situations. Even if extended scans are possible, some degree of additional scan time may not improve the image adequately because the relationship between scan time and SNR is not linear. Additionally, aggravated motion artifacts during longer scans are another nontrivial issue for this approach. Acquisition of a high-resolution image with sufficient SNR in reasonable scan time is desirable but a challenging goal.

Another approach is to improve image quality, such as noise reduction. Several previous studies have been proposed to denoise the images using mathematical priors or well-designed filtering [4]. However, their limited performance has led to difficulty in direct application in clinical practice. Recently, several studies have shown that deep neural network (DNN)-based methods may outperform conventional denoising algorithms [5, 6]. One caveat is that they are mostly based on supervised training with clean-label images [5]. However, noise is always incorporated in signal measurements, revealing the difficulty in achieving clean targets. Additionally, collecting clean-like MR images was based on scans with longer data sampling, and it has the same problem of achieving images with a high SNR.

In this study, we proposed a self-supervised deep learning framework for noise reduction and image quality improvement in MR images without additional scans or reference data. The framework was demonstrated in 3D magnetization-prepared rapid gradient-echo (MP-RAGE) [7], one of the most widely used and requiring high resolution, with quantitative and qualitative evaluations. In addition, we applied this model to vessel wall imaging (VWI) [1], another 3D MR imaging demanding a high SNR, to explore the clinical potential of this approach.

Materials and methods

Data acquisition

This single-center study (Seoul St. Mary's Hospital, Seoul, Republic of Korea) was approved by the internal review board, and informed consent was waived due to the

retrospective nature of this study. A total of 185 participants were scanned by 3D T1-weighted MP-RAGE imaging, and 18 participants were scanned by 3D VWI from January 2021 to November 2021 using a 3-T scanner (MAGNETOM Vida, Siemens Healthcare) with a 64-channel head/neck coil. All data were from clinical scans, and participants were referred to with various clinical indications. After scanning, MP-RAGE and VWI data were collected in k-space and image format, respectively. Demographic information was collected through a review of the medical records.

The 3D sagittal T1-weighted MP-RAGE images were acquired by the following scan parameters: TR ranging from 2100 to 2300 ms, TE = 3.2 ms, TI = 1040 ms, field of view (FOV) = $230 \times 230 \times 168\text{--}180 \text{ mm}^3$, voxel size = $0.7 \times 0.7 \times 0.7 \text{ mm}^3$, and autocalibrating signal (ACS) lines = 24 with acceleration factor (R) = 2×1 . The scan time ranged from 4 min 52 s to 5 min 20 s. For quantitative evaluation, full-sampled 3D MP-RAGE imaging of 19 participants was included (scan time of 9 min 49 s).

For 3D VWI, coronal T1-weighted turbo spin-echo images (sampling perfection with application-optimized contrasts using different flip angle evolutions; SPACE) [8] were acquired by the following scan parameters: TR = 800 ms, TE = 23 ms, echo spacing = 5.84 ms, echo train length = 40, FOV = $180 \times 180 \times 80 \text{ mm}^3$, voxel size = $0.56 \times 0.56 \times 0.5 \text{ mm}^3$, $R = 2 \times 2$ with controlled aliasing in parallel imaging results in higher acceleration (CAIPIRINHA) reconstruction [9]. The scan time was 5 min 25 s.

MR-self Noise2Noise

Recently, the framework named Noise2Noise has demonstrated that DNN-based denoising can be achieved without using clean labels [10]. When the noise of inputs and labels has independent distributions with zero-mean, the network can be trained to have a similar performance to the network trained with clean labels using noisy input-label pairs only. Expanding this framework, several studies have proposed self-supervised learning-based models [11–14]. These models are based on splitting a single (i.e., unpaired) noisy image into a pair of two noisy independent images. Such a self-supervised framework can be extended to MR image reconstruction by splitting the k-space into two disjoint sets. A few studies have utilized this disjoint k-space splitting for DNN-based MR reconstruction [10, 14]. However, their splitting algorithms have led to different undersampling patterns between inputs and labels, limiting the application to the Noise2Noise framework in MRI.

The proposed MR-self Noise2Noise (sN2N) framework leverages the noise independence in MR k-space data to generate training input-label pairs as follows (Fig. 1a). First, the acquired (original) multi-channel k-space data were retrospectively split into two disjoint sets by the alternate sampling

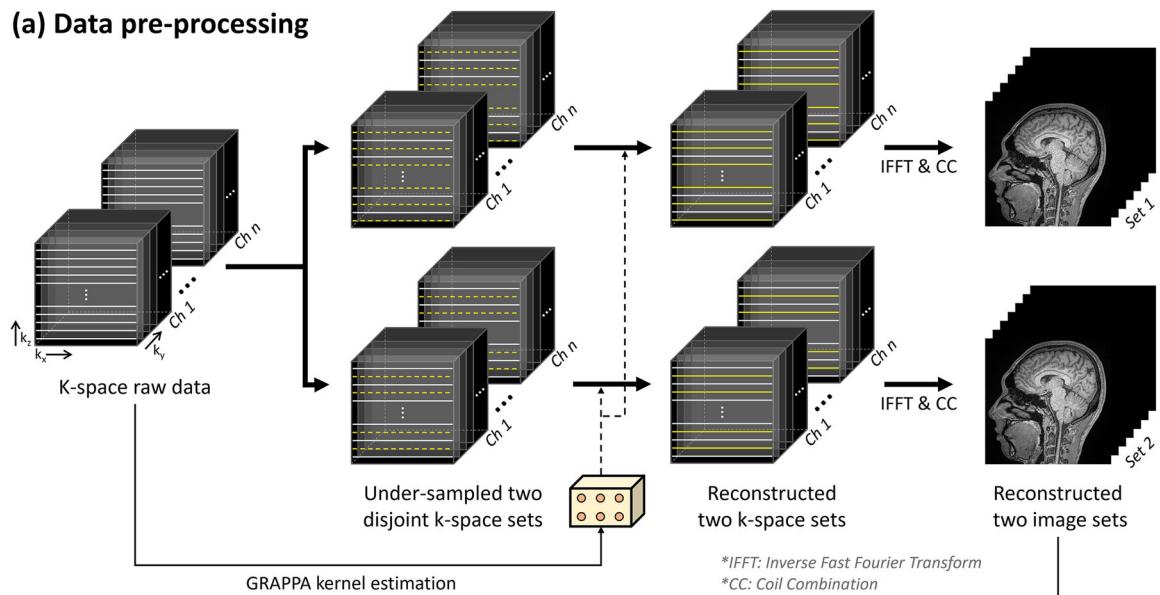
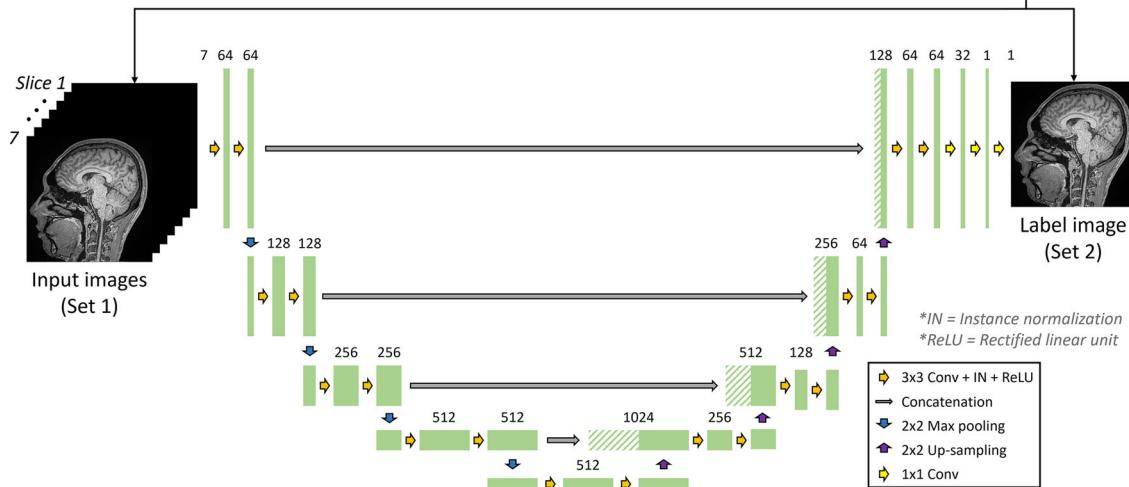
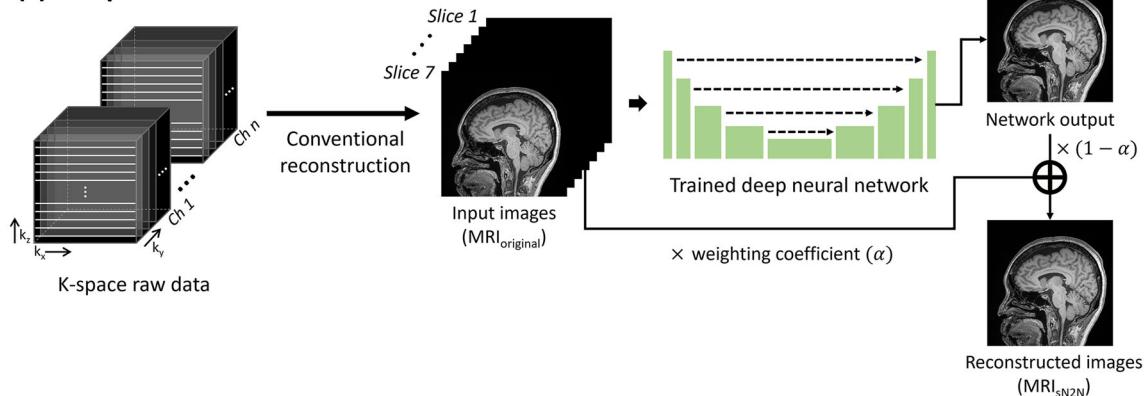
(a) Data pre-processing**(b) Deep neural network training****(c) Deep neural network test**

Fig. 1 **a** Schematic diagram of data preprocessing for the proposed self-supervised deep neural network-based image improvement. Data preprocessing generates a two-image pair with independent noise distributions. **b** The deep neural network architecture of the proposed model. The model was designed with a 2.5D U-Net as a foundational architecture. **c** After

the training process, the conventional MRI images ($\text{MRI}_{\text{original}}$) were utilized as inputs of the network, generating high-quality images (MRI_{sN2N}) by a weighted sum of $\text{MRI}_{\text{original}}$ and corresponding network outputs with coefficient α

scheme along the k_z -axis, generating two noisy independent k -spaces. To compensate for different undersampling patterns between the two multi-channel k -space sets, a $3 \times 2 \times 2$ GRAPPA kernel was estimated from the ACS lines of the original k -space data [15]. In each multi-channel k -space set, the missing data points were estimated by the convolution of the sampled data points with the GRAPPA kernel, followed by inverse Fourier transform and adaptive coil combination [16] (Appendix 1). The procedure resulted in two independent noisy MR images. This image pair was utilized as the input and label for training the DNN.

Deep neural network development

The MR-self Noise2Noise model was constructed by U-Net [17] as a baseline architecture (Fig. 1b). To utilize the spatial information of anatomical brain structures, DNN architecture is designed in a 2.5D fashion (slice concatenation along the channel dimension for inputs). Seven slices were concatenated for the network inputs, and the corresponding outputs were targeted to reconstruct the center slice of the inputs. When training the network, the images of 135 and 11 participants were utilized for training and validation, respectively (33,632 images for training; 2,688 images for validation). The details of the training process are summarized in the Supplementary material (Appendix 2).

In the test phase of the proposed self-supervised trained network, conventional MR images ($\text{MRI}_{\text{original}}$) were utilized as inputs (Fig. 1c). Note that the data preprocessing of the sN2N framework was applied only to the training data, not the test data. For image quality optimization, the final reconstructed images (MRI_{sN2N}) were generated by the weighted sum of $\text{MRI}_{\text{original}}$ and the network output with the coefficient α :

$$\text{MRI}_{\text{sN2N}} = \alpha \times \text{MRI}_{\text{original}} + (1-\alpha) \times \text{Network output} \quad (1)$$

The α value was determined to have the highest peak signal-to-noise-ratio (PSNR) values compared to the labels in the validation dataset. A total of 11 α values ranging from 0 to 1 with a step size of 0.1 were tested. Based on the PSNR error metric, the optimal α was determined to be 0.3, as described in the Supplementary material (Appendix 3).

Image quality evaluation

Qualitative evaluation

For the assessment of image quality, the conventional MP-RAGE images ($\text{MP-RAGE}_{\text{original}}$) and corresponding deep learning-reconstructed images ($\text{MP-RAGE}_{\text{sN2N}}$) were compared in the test set of 20 participants. Visual appropriateness and identification of clinically relevant anatomical structures were assessed by two experienced neuroradiologists (J.J. and

N.-Y.S., each with more than 15 years of experience in clinical neuroradiology). Overall image quality, perceived SNR, contrast, sharpness, and artifacts were scored according to a five-point Likert scale (Table S1). For identification of anatomical structures, fine structure delineation was scored in nine subregions, including the hippocampus, mammillothalamic tract, anterior thalamic nuclei, putamen tail, periaqueductal gray, claustrum, substantia nigra, and entorhinal cortex. Those regions were chosen for the analysis because the image qualities of those structures are clinically important, but visualization was frequently insufficient in clinical scans. Note that high scores indicate better image quality. To demonstrate our DNN reconstruction performance on the lesions, the patient images in the test dataset were additionally evaluated by the two neuroradiologists. Lesion conspicuity was graded on a five-point Likert scale (Table S1).

Quantitative evaluation

In the quantitative evaluation, the additional test data of the 19 participants who scanned with full-sampled MP-RAGE were utilized. Using the fully sampled MP-RAGE images, retrospective undersampling was performed to obtain the same k -space sampling patterns as $\text{MP-RAGE}_{\text{original}}$ ($R = 2 \times 1$). From these retrospectively undersampled k -space data, simulated Gaussian noise was added in both real and imaginary channels, followed by $\text{MP-RAGE}_{\text{original}}$ and $\text{MP-RAGE}_{\text{sN2N}}$ reconstructions. Given the noise standard deviation in the acquired k -space as σ , the additional simulated noise was generated 6 times with different noise standard deviations from 0.5σ to 3σ . The resulting 7 different $\text{MP-RAGE}_{\text{original}}$ and $\text{MP-RAGE}_{\text{sN2N}}$ images (including no simulated noise) were compared by PSNR, structural similarity index (SSIM), and normalized root mean square error (NRMSE) measurements with the full-sampled MP-RAGE images as a reference. Additionally, the volumetric analysis was performed by two different publicly available algorithms (FSL-FAST [18], segmentation of gray matter, white matter, and cerebrospinal fluid; FSL-FIRST [19], segmentation of subcortical structures). To measure the quantitative similarity compared with the full-sampled MP-RAGE volumetry results, dice similarity coefficient (DSC) was estimated in each region of interest (ROI):

$$\text{Dice similarity coefficient } (A, B) = \frac{2(A \cap B)}{(A + B)} \quad (2)$$

where A and B are the two segmentation maps and \cap is the intersection [20].

Generalization on different images

To explore the generalization capability of DNN on different images, the DNN model trained by MP-RAGE data was

applied to VWI data (VWI_{sN2N}). Note that VWI data were acquired with a different sequence, parallel imaging, and scan parameters, including a resolution to MP-RAGE data. The qualitative evaluation of original VWI ($\text{VWI}_{\text{original}}$) and VWI_{sN2N} was performed on the 18 participants by qualitative assessment of two experienced neuroradiologists (J.J. and M.S., 5-year experiences for clinical neuroradiology). Overall image quality, perceived SNR, contrast, sharpness, and artifacts were scored according to a five-point Likert scale (Table S2).

Statistical analysis

In qualitative evaluation, a Wilcoxon signed-rank test was performed in all image quality parameters between $\text{MP-RAGE}_{\text{original}}$ and $\text{MP-RAGE}_{\text{sN2N}}$ (or $\text{VWI}_{\text{original}}$ and VWI_{sN2N}) images. To demonstrate interobserver agreement, weighted Cohen's kappa (κ) values were calculated. Interobserver agreement was defined as poor (0.00~0.20), fair (0.21~0.40), moderate (0.41~0.60), substantial (0.61~0.80), and almost perfect (0.81~1.00). In quantitative evaluation, a paired t-test was performed between the PSNR, SSIM, NRMSE, and DSC of $\text{MP-RAGE}_{\text{original}}$ and $\text{MP-RAGE}_{\text{sN2N}}$. For DSC comparison, averaged DSC values across all ROIs were calculated in each volumetry algorithm. All statistical measurements were estimated by SPSS, Version 28.0 (IBM Corporation), and statistical significance was determined with $p < 0.05$.

Results

Demographic characteristics

The demographic characteristics of the included participants are presented in Table 1. There were no significant differences in participant age or sex between the training, validation, and test sets. Clinical indications of training and validation datasets ($n = 146$) were summarized as follows: work-up for metastasis ($n = 75$), workup for neurodegenerative diseases ($n = 39$), brain tumor ($n = 16$), hemorrhages ($n = 10$), and others including multiple sclerosis ($n = 6$). Among the 20 test set scans, 7 participants had pathological findings, including metastatic tumors ($n = 1$), infarcts ($n = 2$), demyelinating disease ($n = 2$), postsurgical changes ($n = 1$), and brain stem atrophy ($n = 1$).

Qualitative evaluation

Figure 2 shows the qualitative evaluation results. $\text{MP-RAGE}_{\text{sN2N}}$ significantly outperformed $\text{MP-RAGE}_{\text{original}}$ in all image quality parameters, such as overall image quality, perceived SNR, contrast, sharpness, and artifacts ($p < 0.002$). Fine structure delineation was also significantly improved in MP-

Table 1 Demographic characteristics

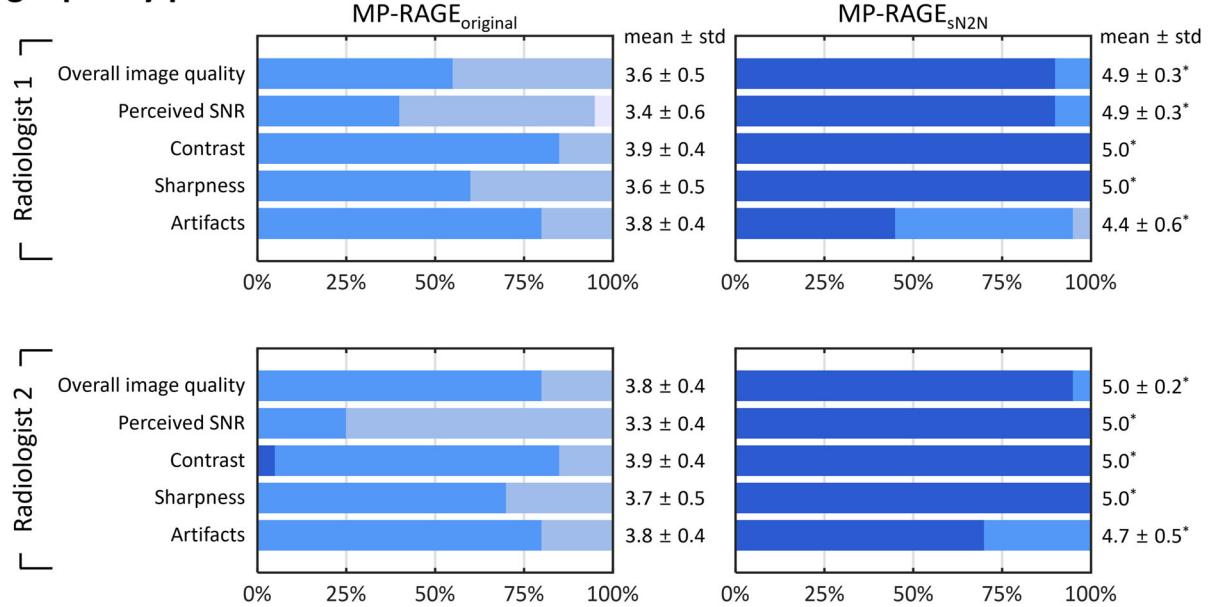
Characteristic	Value
Training/validation dataset	
Sequence	MP-RAGE
No. of participants	135 / 11
Age (years)*	57 ± 20 / 52 ± 17
No. of [female/male] participants	[87/48] / [6/5]
Clinical indications (Number of patients)	
Workup for metastasis	75
Work up for neurodegenerative diseases	39
Brain tumor	16
Hemorrhage	10
Others including multiple sclerosis	6
Qualitative evaluation dataset	
Sequence	MP-RAGE
No. of participants	20
Age (years)*	55 ± 20
No. of female/male participants	13/7
Pathological conditions (Number of patients)	
Cerebellar tumor	1
Infarct	2
Multiple sclerosis	2
Postsurgical changes	1
Brain stem atrophy	1
No relevant abnormal lesions	13
Quantitative evaluation dataset	
Sequence	MP-RAGE
No. of participants	19
Age (years)*	58 ± 15
No. of female/male participants	14/5
Generalization on different images	
Sequence	T1 weighted VWI
No. of participants	18
Age (years)*	53 ± 16
No. of female/male participants	8/10

Note: *Except for age, data are the number of participants. Ages are presented as means ± standard deviations. *MP-RAGE* magnetization-prepared rapid gradient-echo, *VWI* vessel wall imaging

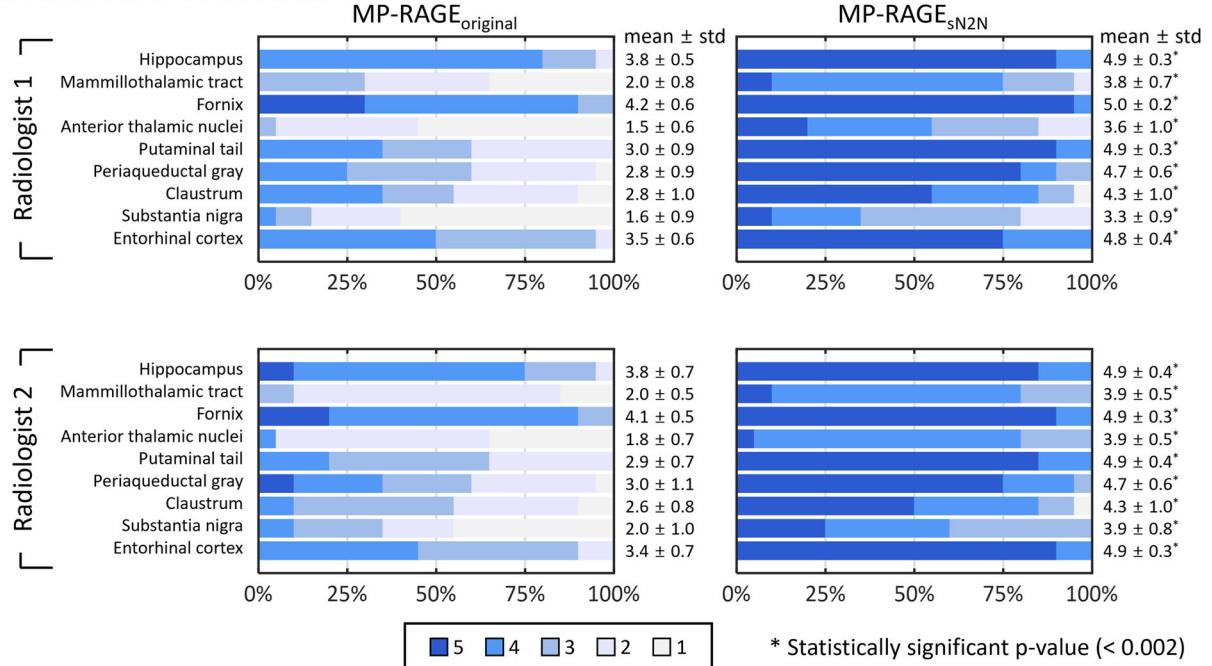
$\text{RAGE}_{\text{sN2N}}$ at all nine subregions ($p < 0.001$). The interobserver reliability of the qualitative scores was substantial to almost perfect for all qualitative scores (Table S3, weighted $\kappa = 0.619 \sim 0.937$, $p < 0.001$).

Figure 3 and Figure S2 are examples of comparisons between $\text{MP-RAGE}_{\text{original}}$ and $\text{MP-RAGE}_{\text{sN2N}}$ for anatomical details. Overall axial images from the lower brain to the upper brain of one representative participant are shown in Figure S3, revealing the performance of $\text{MP-RAGE}_{\text{sN2N}}$ in whole-brain imaging.

(a) Image quality parameter



(b) Fine structure delineation



* Statistically significant p-value (< 0.002)

Fig. 2 Qualitative evaluation results of the two neuroradiologists are displayed. The comparison was performed between conventional MP-RAGE images (MP-RAGE_{original}) and the proposed deep neural network-based images (MP-RAGE_{SN2N}) at five image quality parameters (a) and fine structure delineation (b). The mean and standard deviation of

each parameter across the test participants are displayed on the right side of each score bar. Note that MP-RAGE_{SN2N} outperformed MP-RAGE_{original} in all qualitative parameters with statistically significant differences (all, $p < 0.002$)

Lesion analysis

Among seven patients who had brain lesions, lesion conspicuity scores were significantly higher in MP-RAGE_{SN2N} (5.0 ± 0.0) than those in MP-RAGE_{original} (4.1 ± 0.7 , $p < 0.03$) by

both radiologists. Figure 4 shows two representative images of MP-RAGE_{original} and MP-RAGE_{SN2N} for brain lesions. MP-RAGE_{SN2N} more distinctly demarcates the boundary between cerebellar tumors (white arrows) and edema (white asterisks) than MP-RAGE_{original}, revealing good agreement with T2-

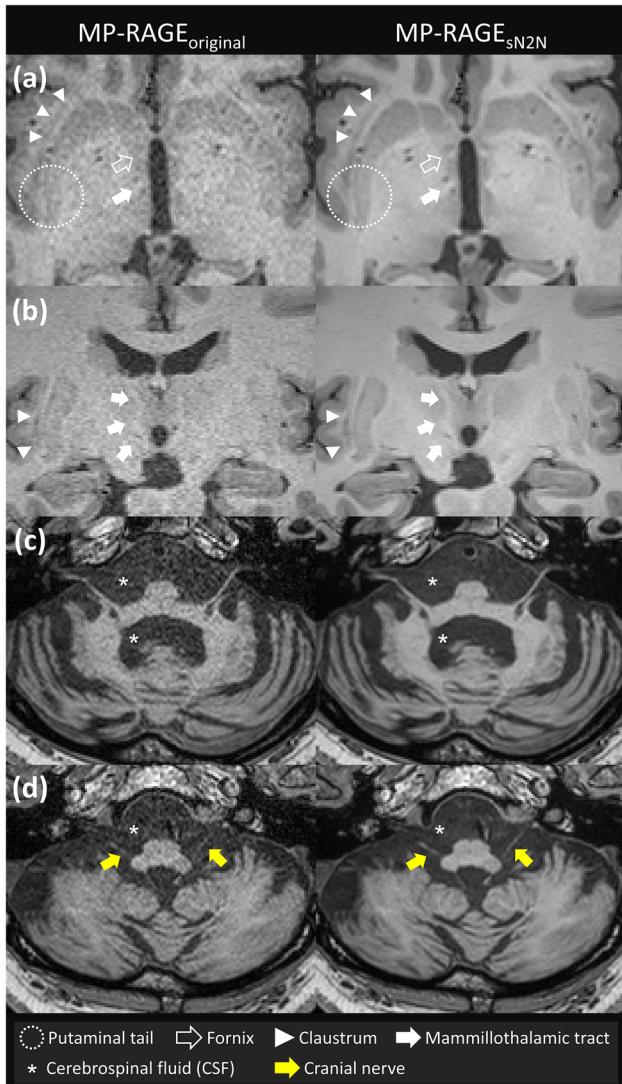


Fig. 3 Representative comparison of conventional MP-RAGE images ($\text{MP-RAGE}_{\text{original}}$) and proposed deep neural network-based images ($\text{MP-RAGE}_{\text{sN2N}}$). $\text{MP-RAGE}_{\text{sN2N}}$ from a 63-year-old woman shows better image quality with less noise than $\text{MP-RAGE}_{\text{original}}$ while preserving details of fine structures (dashed line circles = putaminal tail; blank arrows = fornix; white triangles = claustrum; white arrows = mammillothalamic tract). In another example of a 28-year-old man (c and d), cerebrospinal fluid (CSF) regions were successfully denoised in $\text{MP-RAGE}_{\text{sN2N}}$, as indicated by white asterisks (*), resulting in better delineation of the cranial nerve (yellow arrows) compared to $\text{MP-RAGE}_{\text{original}}$

weighted TSE images (Fig. 4a). In another case, $\text{MP-RAGE}_{\text{sN2N}}$ showed improved contrast of white matter lesions and deep gray matter compared to $\text{MP-RAGE}_{\text{original}}$ (Fig. 4b), which was well matched to lesions on T2 FLAIR images.

Quantitative evaluation

The PSNR, SSIM, NRMSE, and DSC values of $\text{MP-RAGE}_{\text{original}}$ and $\text{MP-RAGE}_{\text{sN2N}}$ were compared in 19

participants, each with 7 different simulated noise levels with fully sampled MP-RAGE images as a reference. Each value was calculated with mean and standard deviation across the participants (Table 2). Significantly higher PSNR, higher SSIM, and lower NRMSE were measured in $\text{MP-RAGE}_{\text{sN2N}}$ than in $\text{MP-RAGE}_{\text{original}}$ in all cases ($p < 0.001$). In volumetry results, $\text{MP-RAGE}_{\text{sN2N}}$ revealed comparable ($p > 0.09$) or higher ($p < 0.02$) DSC values than $\text{MP-RAGE}_{\text{original}}$. Visual assessment of $\text{MP-RAGE}_{\text{sN2N}}$ showed excellent image quality and small structure details, even comparable to fully sampled data (Fig. 5 and Figure S4-6).

Generalization on different images

All image quality parameters of VWI_{sN2N} were significantly higher than those of $\text{VWI}_{\text{original}}$ except artifact scores (artifact scores $p > 0.15$, others < 0.001 , Figure S7). Interobserver agreements between both radiologists were moderate to substantial ($\kappa = 0.455 \sim 0.645$, $p < 0.001$; Table S4). Figure 6 shows the representative images of $\text{VWI}_{\text{original}}$ and VWI_{sN2N} reconstructions. Since VWI_{sN2N} successfully denoised both CSF regions outside of the vessel wall and black blood regions inside of the vessel wall, the boundaries of the inner-outer wall of vessels were clearly demarcated. Additionally, the quality of other structures, such as brain parenchyma regions, was also improved.

Discussion

In this study, we proposed a self-supervised deep neural network-based image reconstruction to improve MR image quality without additional scans or reference data for labels. The quality improvement of our framework was demonstrated in two different 3D high-resolution MR images, where sufficient SNR is needed: MP-RAGE and VWI. Compared to conventional MP-RAGE reconstruction ($\text{MP-RAGE}_{\text{original}}$), the proposed reconstruction ($\text{MP-RAGE}_{\text{sN2N}}$) outperformed in image quality (averaged Likert scores = 3.7 vs. 4.9, $p < 0.001$), fine structure delineation (2.8 vs. 4.4, $p < 0.001$), and lesion conspicuity (4.1 vs. 5.0, $p < 0.001$). In addition, $\text{MP-RAGE}_{\text{sN2N}}$ showed higher PSNR (0.5~2.3 dB, $p < 0.001$), higher SSIM (0.002~0.047, $p < 0.001$), lower NRMSE (0.2~5.6 %, $p < 0.001$), and comparable ($p > 0.09$) or higher DSC (0.005~0.084, $p < 0.02$) than $\text{MP-RAGE}_{\text{original}}$, supporting the results of qualitative evaluation. When the DNN model was applied to VWI images, the image quality was also improved (average Likert scores = 3.5 vs. 4.6, $p < 0.001$), although the VWI test set was scanned with a different sequence, parallel imaging, and scan parameters than the DNN training set.

To generate training data, the k-space splitting was applied to $R = 2 \times 1$ data acquired by 3D MP-RAGE sequence. Note

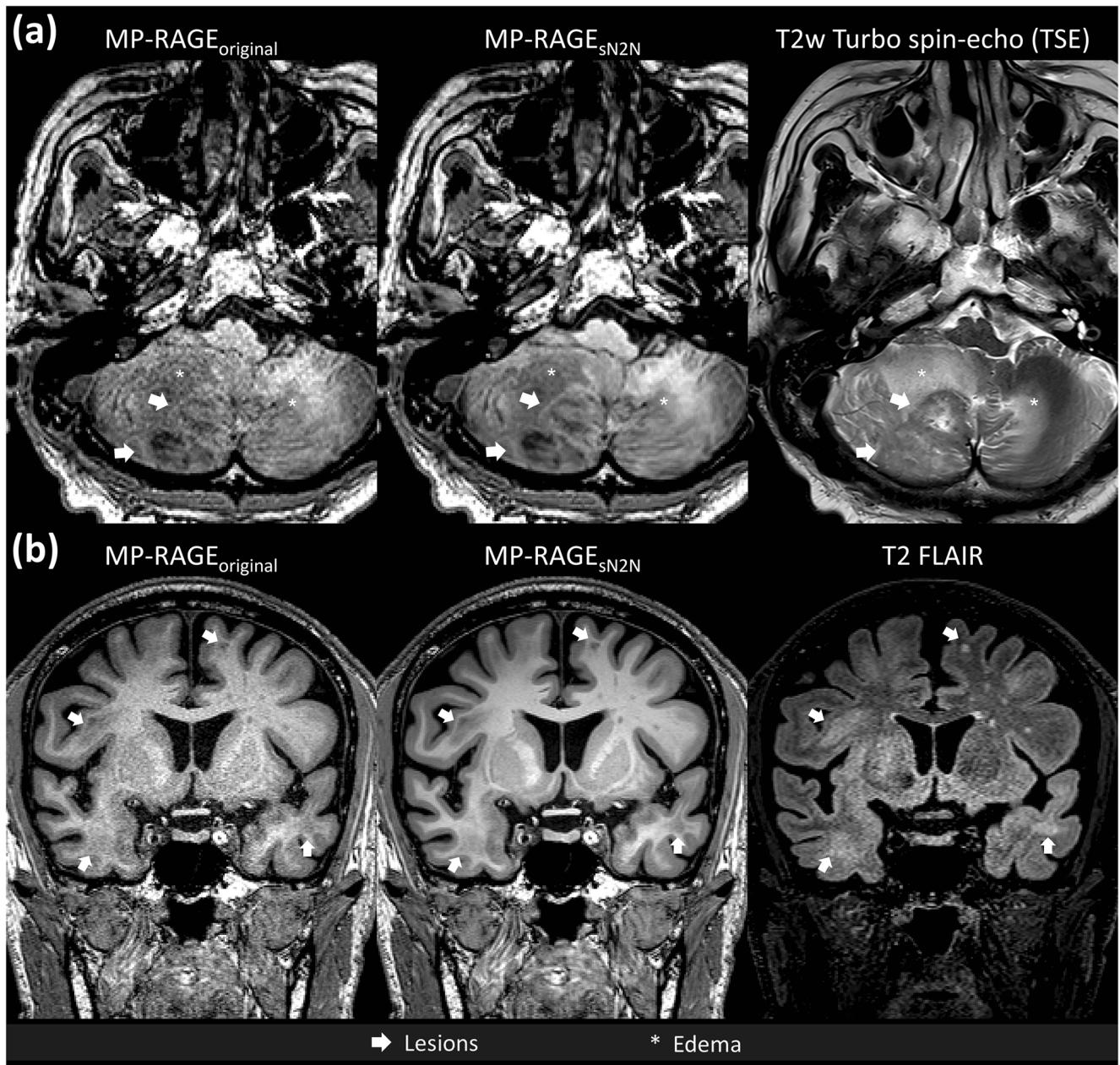


Fig. 4 MR images of the two patients are displayed. **a** The patient in the upper row is a 59-year-old man who has cerebellar tumors, and **b** the patient in the lower row is a 66-year-old man who has multiple sclerosis. Conventional MP-RAGE images (MP-RAGE_{original}), the proposed deep

neural network-based images (MP-RAGE_{SN2N}), and clinical reference images (T2 weighted TSE for **a**; T2 FLAIR for **b**) are shown. Both results indicate that MP-RAGE_{SN2N} shows better agreement with clinical reference images than MP-RAGE_{original}

that $R = 2 \times 2$ data with 2 averages were not utilized for the training data because the scan time of $R = 2 \times 2$ data is the same as that of $R = 2 \times 1$ at conventional 3D MP-RAGE sequence [21]. Therefore, $R = 2 \times 2$ data with 2 averages require additional scan time to construct the training input-label pairs.

Most previous approaches for improving MRI image quality utilized study-specific protocols for training labels with sufficient quality [22–24], requiring additional or separate scans. However, these approaches highly limit their

applicability to diverse clinical situations. For example, additional protocols are not commonly allowed in clinical practice. Moreover, the extended scan time might increase the occurrence of motion, leading to training label image quality degradation. The proposed reconstruction demonstrated that improvement in MR image quality can be achieved by using data from routine clinical protocols only.

As demonstrated in qualitative and quantitative evaluations, our DNN model improved the MR image quality without blurring the fine structures ($p < 0.002$). We believe two

Table 2 Quantitative measurements of MP-RAGE_{original} and MP-RAGE_{sN2N} at 7 different simulated noise levels with fully sampled images as a reference

	Additional noise							
	0	+ 0.5σ	+ 1σ	+ 1.5σ	+ 2σ	+ 2.5σ	+ 3σ	
PSNR (dB)	MP-RAGE _{original}	41.2 ± 0.8	39.5 ± 1.0	37.1 ± 1.0	34.8 ± 1.0	32.7 ± 1.0	31.0 ± 1.0	29.5 ± 1.0
	MP-RAGE _{sN2N}	41.7 ± 0.9	40.5 ± 1.0	38.8 ± 1.0	36.8 ± 1.0	34.9 ± 1.0	33.2 ± 1.0	31.8 ± 1.0
	p value	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*
SSIM (0-1)	MP-RAGE _{original}	0.984 ± 0.003	0.981 ± 0.003	0.966 ± 0.004	0.939 ± 0.007	0.901 ± 0.012	0.857 ± 0.016	0.810 ± 0.020
	MP-RAGE _{sN2N}	0.986 ± 0.003	0.985 ± 0.003	0.977 ± 0.003	0.959 ± 0.005	0.931 ± 0.009	0.900 ± 0.013	0.857 ± 0.017
	p value	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*
NRMSE (%)	MP-RAGE _{original}	4.4 ± 0.6	7.3 ± 1.7	9.7 ± 2.4	13.2 ± 3.5	17.2 ± 4.8	21.6 ± 5.9	26.1 ± 7.7
	MP-RAGE _{sN2N}	4.2 ± 0.6	6.4 ± 1.5	7.7 ± 1.8	10.1 ± 2.6	13.3 ± 3.7	16.8 ± 4.8	20.6 ± 6.4
	p value	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*
FSL-FAST DSC (0-1)	MP-RAGE _{original}	0.955 ± 0.006	0.946 ± 0.007	0.923 ± 0.008	0.895 ± 0.011	0.866 ± 0.013	0.836 ± 0.015	0.807 ± 0.017
	MP-RAGE _{sN2N}	0.956 ± 0.006	0.954 ± 0.006	0.947 ± 0.007	0.936 ± 0.009	0.922 ± 0.010	0.907 ± 0.011	0.891 ± 0.012
	p value	0.10	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*
FSL-FIRST DSC (0-1)	MP-RAGE _{original}	0.929 ± 0.014	0.923 ± 0.016	0.906 ± 0.020	0.892 ± 0.021	0.873 ± 0.025	0.857 ± 0.026	0.839 ± 0.029
	MP-RAGE _{sN2N}	0.934 ± 0.016	0.932 ± 0.012	0.928 ± 0.016	0.922 ± 0.016	0.911 ± 0.017	0.902 ± 0.018	0.892 ± 0.016
	p value	0.01*	0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*	< 0.001*

Note: Mean and standard deviation were calculated across the participants. σ, the noise standard deviation of original k-space data. PSNR, peak signal-to-noise ratio; SSIM, structural similarity index; NRMSE, normalized mean squared error; FSL-FAST, segmentation protocol for gray matter, white matter, and cerebrospinal fluid; FSL-FIRST, segmentation protocol for subcortical structures; DSC, dice similarity coefficient; MP-RAGE_{original}, conventional MP-RAGE reconstruction; MP-RAGE_{sN2N}, the proposed self-supervised deep neural network-based reconstruction

*With bold letters indicate statistically significant differences between MP-RAGE_{original} and MP-RAGE_{sN2N} ($p < 0.05$)

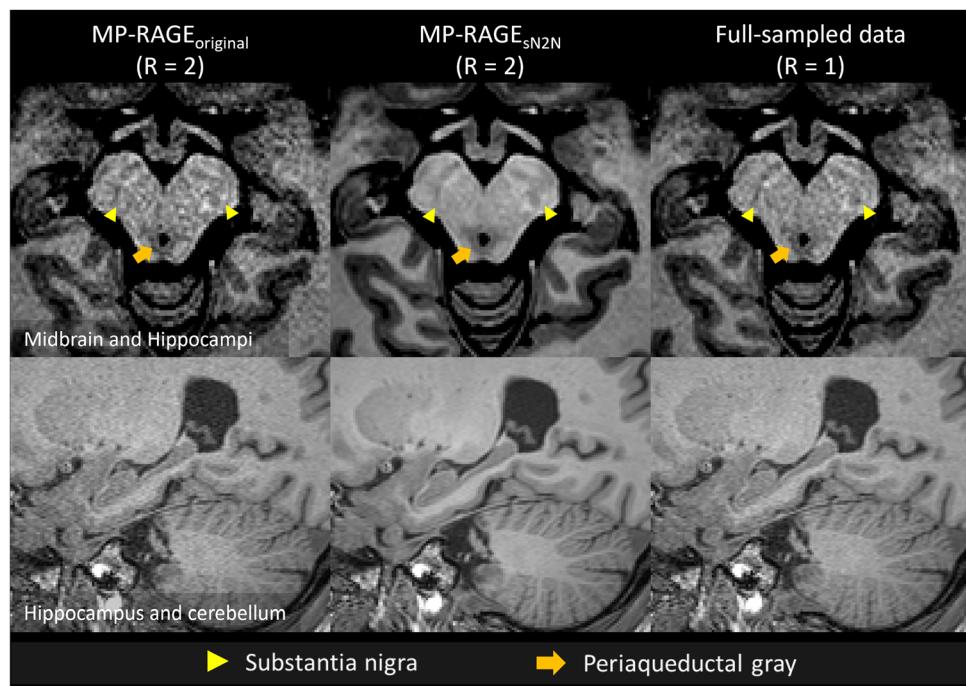


Fig. 5 The reconstructed images from a 72-year-old woman (upper row) and a 53-year-old woman (lower row) are displayed with full-sampled MP-RAGE images. From the full-sampled MP-RAGE data, retrospectively undersampling with $R = 2$ was performed, followed by conventional MP-RAGE (MP-RAGE_{original}) and the proposed deep neural network-

based reconstruction (MP-RAGE_{sN2N}). Compared to MP-RAGE_{original}, MP-RAGE_{sN2N} showed better agreement with full-sampled MP-RAGE in the structural delineation of the substantia nigra (yellow triangles), periaqueductal gray (orange arrows), hippocampus, and cerebellum

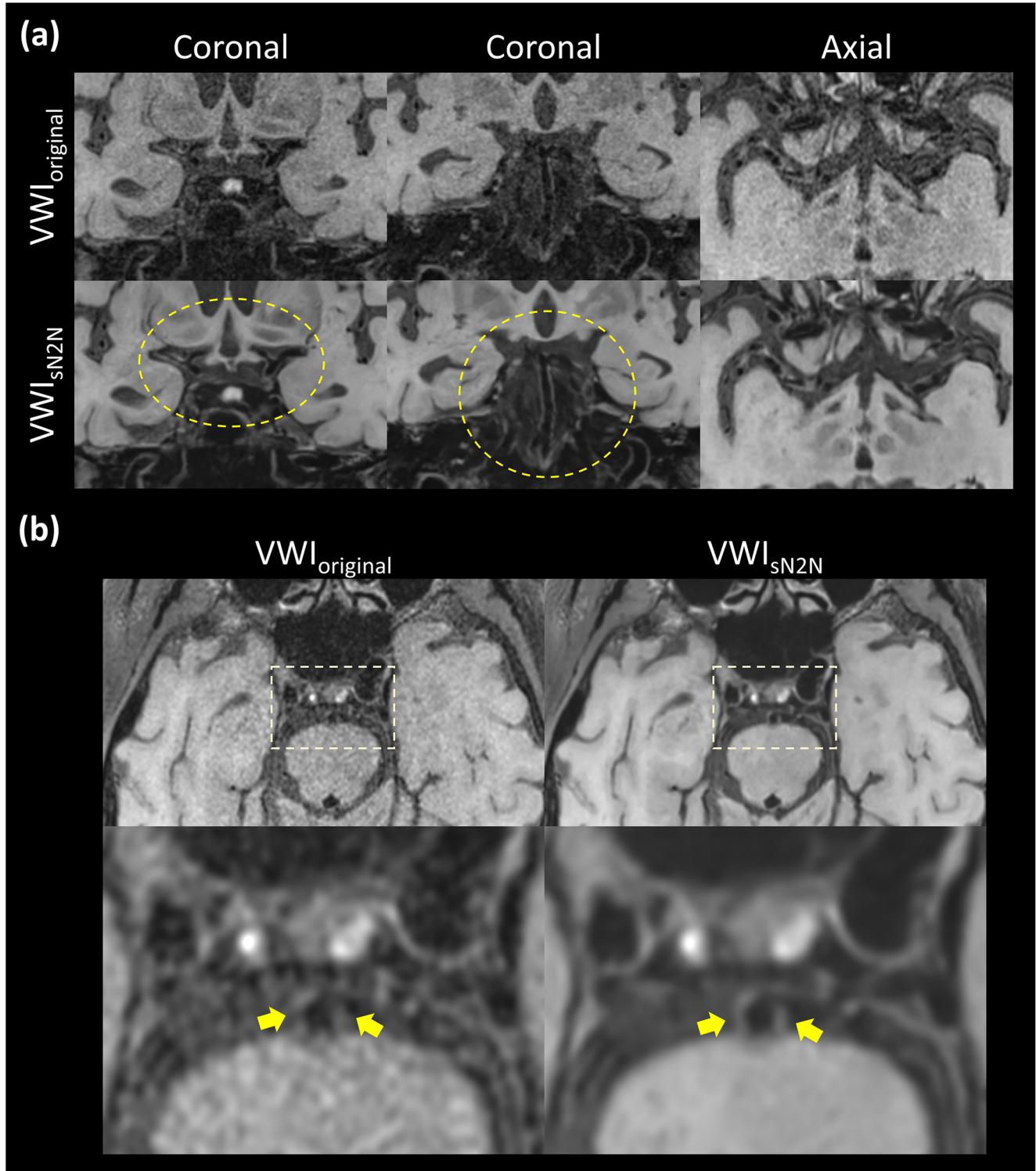


Fig. 6 Conventional vessel wall images (VWI_{original}) and proposed deep neural network-based images (VWI_{sN2N}) from a 68-year-old man with coronal and axial views are displayed (a). VWI_{sN2N} successfully denoised the regions around the circle of Willis (yellow dashed circles), leading to

better vessel wall delineation. On magnified view of another axial slice (b), improved image with better vessel wall delineation in small vessels (yellow arrows)

rationales are behind this improvement. First, our MR physics-based processing generated the training images to have noise characteristics similar to those of the images from

the original MR scans, and the network successfully improved the conventional MR image quality. Second, the 2.5D U-Net in our network architecture may utilize anatomical structure

continuity by slice concatenation. Such priors can improve the performance of the network by differentiating the structure and noise. The 2.5D implementation was demonstrated in previous studies [25, 26], supporting our results.

Interestingly, our model enhanced the VWI image qualities. Considering the significant differences between MP-RAGE training data and VWI data, the performance of our model for VWI is unexpected. This may be explained by the convolutional layer-based DNN in our model. Since the convolutional filters in DNN learned the local noise features rather than voxel-wise features, the influence of different characteristics between MP-RAGE training data and VWI test data was reduced [27]. Nevertheless, the different noise distributions between MP-RAGE training data and VWI data pose potential risks to our DNN-based reconstruction. The risks can be reduced by traditional approaches in supervised learning, such as training with a dataset containing various noise characteristics (i.e., data augmentation) or fine-tuning the network with small amounts of VWI training data [27].

Currently, despite the clinical need for higher-resolution images, there is a practical limit of the SNR and resolution of clinical MR scans associated with the strength of the magnet, coil systems, pulse sequences, and scan time [28, 29]. As a result, higher-resolution images were selectively acquired for limited clinical situations [30, 31]. However, our approach enables the acquisition of submillimeter high-resolution images in clinically acceptable scan time. It can provide better delineation of fine structures, such as the hippocampus, entorhinal cortex, mammillothalamic tract, anterior thalamic nuclei, or posterior aspect of the putamina, which can help to target deep brain stimulation or detect fine cortical or hippocampal abnormalities for treatment planning [32–36]. Moreover, high-resolution images with improved SNR can be helpful for segmenting brain structures and volumetric assessment. The generalization potential can extend the usefulness of our model to other clinical applications requiring a high SNR for imaging tiny structures, such as VWI, nigrosome imaging, and MR neurography [1, 3, 37].

A few limitations need to be addressed in this study. First, while the proposed model showed excellent performances, we did not compare our results to the current standard supervised approach. However, as we discussed, the acquisition of clean data for supervised learning is a challenging and expensive task. Although the performances of our results using only a clinically available dataset exceeded expectations, it is worth exploring the performance of other models that require clean (or clean-like) data for training labels. Second, the generalization performance of our method on various scan parameters or vendors needs to be conducted. One promising result is that the proposed model trained only with MP-RAGE images also improved VWI images despite different pulse sequences and parameters. Third, the performance of the proposed N2N model can be degraded when the SNR of training data is too

low. Since the zero-mean noise assumption of the N2N framework is not valid in severe SNR degradation, the reconstructed images from the network trained with low SNR data may have low image quality [38]. The boundary conditions of trainable SNR may be worth exploring further. Fourth, our framework was only developed in the MR magnitude images. However, a few sequences utilize the phase information of MRI, such as susceptibility-weighted imaging [39] and quantitative susceptibility mapping [40]. Therefore, a framework for improving the MR phase is necessary. Last, the qualitative evaluation was only performed with a relatively small number of participants. Although the number of participants was not sufficient, the test group included six different types of clinical conditions, revealing the potential of our reconstruction. Further evaluations need to be performed in large datasets and their clinical performances.

In conclusion, we proposed a framework for MR image quality improvement using self-supervised learning without additional scan data. Our model not only performs MR image denoising but also reveals better delineation of the anatomical fine structures. The proposed framework can be generally applied to MR reconstruction and may become a promising tool for a wide range of MR imaging modalities.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00330-022-09243-y>.

Funding This work was supported by the Korea Medical Device Development Fund grant funded by the Korean government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: KMDF_PR_20200901_0062, 9991006735) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (Grant Number: NRF-2020R1C1C1012320).

Declarations

Guarantor The scientific guarantor of this publication is Jinhee Jang.

Conflict of interest Woojin Jung is an employee of AIRS Medical. Hyunsoo Lee is an employee of Siemens Healthineers Ltd. The other authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology

- retrospective
- cross-sectional study
- performed at one institution

References

- Obusez EC, Hui F, Hajj-ali RA et al (2014) High-resolution MRI vessel wall imaging: spatial and temporal patterns of reversible cerebral vasoconstriction syndrome and central nervous system vasculitis. *AJNR Am J Neuroradiol* 35:1527–1532. <https://doi.org/10.3174/ajnr.a3909>
- Pinker K, Noebauer-Huhmann IM, Stavrou I et al (2007) High-resolution contrast-enhanced, susceptibility-weighted MR imaging at 3T in patients with brain tumors: correlation with positron-emission tomography and histopathologic findings. *AJNR Am J Neuroradiol* 28:1280–1286. <https://doi.org/10.3174/ajnr.a0540>
- Noh Y, Sung YH, Lee J, Kim EY (2015) Nigrosome 1 detection at 3T MRI for the diagnosis of early-stage idiopathic parkinson disease: assessment of diagnostic accuracy and agreement on imaging asymmetry and clinical laterality. *AJNR Am J Neuroradiol* 36: 2010–2016. <https://doi.org/10.3174/ajnr.a4412>
- Buades A, Coll B, Morel JM (2005) A review of image denoising algorithms, with a new one. *Multiscale Model Simul* 4:490–530. <https://doi.org/10.1137/040616024>
- Zhang K, Zuo W, Chen Y et al (2017) Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans Image Process* 26:3142–3155. <https://doi.org/10.1109/tip.2017.2662206>
- Gondara L (2016) Medical image denoising using convolutional denoising autoencoders. 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW) 241–246. <https://doi.org/10.1109/icdmw.2016.0041>
- Mugler JP, Brookeman JR (1990) Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). *Magn Reson Med* 15:152–157. <https://doi.org/10.1002/mrm.1910150117>
- Kato Y, Higano S, Tamura H et al (2009) Usefulness of contrast-enhanced T1-weighted sampling perfection with application-optimized contrasts by using different flip angle evolutions in detection of small brain metastasis at 3T MR imaging: comparison with magnetization-prepared rapid acquisition of gradient echo imaging. *AJNR Am J Neuroradiol* 30:923–929. <https://doi.org/10.3174/ajnr.a1506>
- Breuer FA, Blaimer M, Heidemann RM et al (2005) Controlled aliasing in parallel imaging results in higher acceleration (CAPIRINHA) for multi-slice imaging. *Magn Reson Med* 53: 684–691. <https://doi.org/10.1002/mrm.20401>
- Lehtinen J, Munkberg J, Hasselgren J et al (2018) Noise2Noise: learning image restoration without clean data. *arXiv preprint arXiv:180304189*. <https://doi.org/10.48550/arXiv.1803.04189>
- Batson J, Royer L (2019) Noise2self: blind denoising by self-supervision. *International Conference on Machine Learning*, PMLR 97:524–533
- Krull A, Buchholz T-O, Jug F (2019) Noise2Void - learning denoising from single noisy images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 2129–2137
- Fadnavis S, Batson J, Garyfallidis E (2020) Patch2Self: denoising diffusion MRI with self-supervised learning. *arXiv preprint arXiv:201101355*. <https://doi.org/10.48550/arXiv.2011.01355>
- Yaman B, Hosseini SAH, Moeller S et al (2020) Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magn Reson Med* 84:3172–3191. <https://doi.org/10.1002/mrm.28378>
- Griswold MA, Jakob PM, Heidemann RM et al (2002) Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med* 47:1202–1210. <https://doi.org/10.1002/mrm.10171>
- Walsh DO, Gmitro AF, Marcellin MW (2000) Adaptive reconstruction of phased array MR imagery. *Magn Reson Med* 43: 682–690. [https://doi.org/10.1002/\(sici\)1522-2594\(200005\)43:5<682::aid-mrm10>3.0.co;2-g](https://doi.org/10.1002/(sici)1522-2594(200005)43:5<682::aid-mrm10>3.0.co;2-g)
- Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: Int Conf Med Image Comput Comput Interv 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
- Zhang Y, Brady M, Smith S (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* 20:45. <https://doi.org/10.1109/42.906424>
- Patenaude B, Smith SM, Kennedy DN, Jenkinson M (2011) A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* 56:907–922. <https://doi.org/10.1016/j.neuroimage.2011.02.046>
- Fenster A, Chiu B (2005) Evaluation of segmentation algorithms for medical Imaging. In: Conf Proc IEEE Eng Med Biol Soc. Shanghai, pp 7186–7189
- Brenner D, Stirnberg R, Pracht ED, Stöcker T (2014) Two-dimensional accelerated MP-RAGE imaging with flexible linear reordering. *MAGMA* 27:455–462. <https://doi.org/10.1007/s10334-014-0430-y>
- Kaye EA, Aherne EA, Duzgol C et al (2020) Accelerating prostate diffusion-weighted MRI using a guided denoising convolutional neural network: retrospective feasibility study. *Radiol Artif Intell* 2:e200007. <https://doi.org/10.1148/ryai.2020200007>
- Gassenmaier S, Afat S, Nickel MD et al (2021) Accelerated T2-weighted TSE imaging of the prostate using deep learning image reconstruction: a prospective comparison with standard T2-weighted TSE imaging. *Cancers* 13:3593. <https://doi.org/10.3390/cancers13143593>
- Kidoh M, Shinoda K, Kitajima M et al (2020) Deep learning based noise reduction for brain MR imaging: tests on phantoms and healthy volunteers. *Magn Reson Med Sci* 19:195–206. <https://doi.org/10.2463/mrms.mp.2019-0018>
- Kittrungrotsakul T, Han X-H, Iwamoto Y et al (2018) A 2.5D cascaded convolutional neural network with temporal information for automatic mitotic cell detection in 4D microscopic images. 2018 14th Int Conf Nat Comput Fuzzy Syst Knowl Discov (ICNC-FSKD) 00:202–205. <https://doi.org/10.1109/fskd.2018.8687125>
- Ziabari A, Ye DH, Srivastava S, et al (2018) 2.5D deep learning For CT image reconstruction using a multi-gPU implementation. 2018 52nd Asilomar Conf Signals Syst Comput 00:2044–2049. <https://doi.org/10.1109/acssc.2018.8645364>
- Knoll F, Hammerik K, Kobler E et al (2018) Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magn Reson Med* 81:116–128. <https://doi.org/10.1002/mrm.27355>
- Keil B, Blau JN, Biber S et al (2013) A 64-channel 3T array coil for accelerated brain MRI. *Magn Reson Med* 70:248–258. <https://doi.org/10.1002/mrm.24427>
- Bilgic B, Gagoski BA, Cauley SF et al (2014) Wave-CAIPI for highly accelerated 3D imaging. *Magn Reson Med* 73:2152–2162. <https://doi.org/10.1002/mrm.25347>
- Kang N, Qiao Y, Wasserman BA (2021) Essentials for interpreting intracranial vessel wall MRI results: state of the art. *Radiology* 300: 492–505. <https://doi.org/10.1148/radiol.2021204096>
- Bae YJ, Kim J-M, Sohn C-H et al (2021) Imaging the substantia nigra in Parkinson disease and other parkinsonian syndromes. *Radiology* 300:203341. <https://doi.org/10.1148/radiol.2021203341>
- Guo W, Koo B-B, Kim J-H et al (2021) Defining the optimal target for anterior thalamic deep brain stimulation in patients with drug-

- refractory epilepsy. *J Neurosurg* 134:1054–1063. <https://doi.org/10.3171/2020.2.jns193226>
- 33. Wong-Kisiel LC, Quiroga DFT, Kenney-Jung DL et al (2018) Morphometric analysis on T1-weighted MRI complements visual MRI review in focal cortical dysplasia. *Epilepsy Res* 140:184–191. <https://doi.org/10.1016/j.eplepsyres.2018.01.018>
 - 34. Niri SG, Khalaf AM, Massoud TF (2020) The mammillothalamic tracts: age-related conspicuity and normative morphometry on brain magnetic resonance imaging. *Clin Anat* 33:911–919. <https://doi.org/10.1002/ca.23595>
 - 35. Farid N, Girard HM, Kemmotsu N et al (2012) Temporal lobe epilepsy: quantitative MR volumetry in detection of hippocampal atrophy. *Radiology* 264:542–550. <https://doi.org/10.1148/radiol.12112638>
 - 36. Kimura Y, Shioya A, Saito Y et al (2019) Radiologic and pathologic features of the transmantle sign in focal cortical dysplasia: the T1 signal is useful for differentiating subtypes. *AJNR Am J Neuroradiol* 40:1060–1066. <https://doi.org/10.3174/ajnr.a6067>
 - 37. Chhabra A, Lee PP, Bizzell C, Soldatos T (2011) 3 Tesla MR neurography—technique, interpretation, and pitfalls. *Skeletal Radiol* 40:1249. <https://doi.org/10.1007/s00256-011-1183-6>
 - 38. Zhussip M, Soltanayev S, Chun SY (2019) Extending Stein's unbiased risk estimator to train deep denoisers with correlated pairs of noisy images. *arXiv preprint arXiv:190202452*. <https://doi.org/10.48550/arXiv.1902.02452>
 - 39. Haacke EM, Xu Y, Cheng YN, Reichenbach JR (2004) Susceptibility weighted imaging (SWI). *Magn Reson Med* 52: 612–618. <https://doi.org/10.1002/mrm.20198>
 - 40. de Rochefort L, Brown R, Prince MR, Wang Y (2008) Quantitative MR susceptibility mapping using piece-wise constant regularized inversion of the magnetic field. *Magn Reson Med* 60:1003–1009. <https://doi.org/10.1002/mrm.21710>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.