Platform

SunCHECK™

One Database for Complete Quality Management

Learn more >

Patient

- Plan Quality Checks
- Secondary Dose Calculations
- Pre-Treatment QA
- In-Vivo Monitoring

Machine

- Standardized Routine QA
- Direct Device Control
- Automated Imaging, MLC & VMAT QA
- Protocol-Based QA

Received: 16 June 2021 | Revised: 22 November 2021 | Accepted: 12 December 2021

DOI: 10.1002/mp.15427

RESEARCH ARTICLE

MEDICAL PHYSICS

High-fidelity fast volumetric brain MRI using synergistic wave-controlled aliasing in parallel imaging and a hybrid denoising generative adversarial network (HDnGAN)

Ziyu Li[1] | **Qiyuan Tian**[2,3] | **Chanon Ngamsombat**[2,4] | **Samuel Cartmell**[5] | **John Conklin**[2,3,5] | **Augusto Lio M. Gonçalves Filho**[2,5] | **Wei-Ching Lo**[6] | **Guangzhi Wang**[1] | **Kui Ying**[7] | **Kawin Setsompop**[2,3,8] | **Qiuyun Fan**[2,3] | **Berkin Bilgic**[2,3,8] | **Stephen Cauley**[2,3] | **Susie Y. Huang**[2,3,8]

[1] Department of Biomedical Engineering, Tsinghua University, Beijing, P.R. China

[2] Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts, USA

[3] Harvard Medical School, Boston, Massachusetts, USA

[4] Department of Radiology, Faculty of Medicine, Siriraj HospitalMahidol University, Mahidol, Thailand

[5] Department of Radiology, Massachusetts General Hospital, Boston, Massachusetts, USA

[6] Siemens Medical Solutions, Boston, Massachusetts, USA

[7] Department of Engineering Physics, Tsinghua University, Beijing, P.R. China

[8] Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Correspondence
Qiyuan Tian, Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, 149 13th Street, Charlestown, MA 02129, USA.
Email: qtian@mgh.harvard.edu

Funding information
National Institutes of Health (NIH), Grant/Award Numbers: P41-EB030006, K23-NS096056, R01-NS118187, R01-EB020613, R01-EB028797, R03-EB031175, K99-AG073506; Siemens Healthineers (research grant); Athinoula A. Martinos Center for Biomedical Imaging; Massachusetts General Hospital Claflin Distinguished Scholar Award; Tsinghua University Initiative Scientific Research Program

Abstract

Purpose: The goal of this study is to leverage an advanced fast imaging technique, wave-controlled aliasing in parallel imaging (Wave-CAIPI), and a generative adversarial network (GAN) for denoising to achieve accelerated high-quality high-signal-to-noise-ratio (SNR) volumetric magnetic resonance imaging (MRI).

Methods: Three-dimensional (3D) T_2-weighted fluid-attenuated inversion recovery (FLAIR) image data were acquired on 33 multiple sclerosis (MS) patients using a prototype Wave-CAIPI sequence (acceleration factor $R = 3 \times 2$, 2.75 min) and a standard T_2-sampling perfection with application-optimized contrasts by using flip angle evolution (SPACE) FLAIR sequence ($R = 2$, 7.25 min). A hybrid denoising GAN entitled "HDnGAN" consisting of a 3D generator and a 2D discriminator was proposed to denoise highly accelerated Wave-CAIPI images. HDnGAN benefits from the improved image synthesis performance provided by the 3D generator and increased training samples from a limited number of patients for training the 2D discriminator. HDnGAN was trained and validated on data from 25 MS patients with the standard FLAIR images as the target and evaluated on data from eight MS patients not seen during training. HDnGAN was compared to other denoising methods including adaptive optimized nonlocal means (AONLM), block matching with 4D filtering (BM4D), modified U-Net (MU-Net), and 3D GAN in qualitative and quantitative analysis of output images using the mean squared error (MSE) and Visual Geometry Group (VGG) perceptual loss compared to standard FLAIR images, and a reader assessment by two neuroradiologists regarding sharpness, SNR, lesion conspicuity, and overall quality. Finally, the performance of these denoising methods was compared at higher noise levels using simulated data with added Rician noise.

Results: HDnGAN effectively denoised low-SNR Wave-CAIPI images with sharpness and rich textural details, which could be adjusted by controlling the contribution of the adversarial loss to the total loss when training the generator. Quantitatively, HDnGAN ($\lambda = 10^{-3}$) achieved low MSE and the lowest VGG perceptual loss. The reader study showed that HDnGAN ($\lambda = 10^{-3}$) significantly improved the SNR of Wave-CAIPI images ($p < 0.001$), outperformed AONLM ($p = 0.015$), BM4D ($p < 0.001$), MU-Net ($p < 0.001$), and 3D GAN ($\lambda = 10^{-3}$) ($p < 0.001$) regarding image sharpness, and outperformed MU-Net ($p < 0.001$) and 3D GAN ($\lambda = 10^{-3}$) ($p = 0.001$) regarding lesion conspicuity. The overall

quality score of HDnGAN ($\lambda = 10^{-3}$) (4.25 ± 0.43) was significantly higher than those from Wave-CAIPI ($3.69 \pm 0.46, p = 0.003$), BM4D ($3.50 \pm 0.71, p = 0.001$), MU-Net ($3.25 \pm 0.75, p < 0.001$), and 3D GAN ($\lambda = 10^{-3}$) ($3.50 \pm 0.50, p < 0.001$), with no significant difference compared to standard FLAIR images (4.38 ± 0.48, $p = 0.333$). The advantages of HDnGAN over other methods were more obvious at higher noise levels.

Conclusion: HDnGAN provides robust and feasible denoising while preserving rich textural detail in empirical volumetric MRI data. Our study using empirical patient data and systematic evaluation supports the use of HDnGAN in combination with modern fast imaging techniques such as Wave-CAIPI to achieve high-fidelity fast volumetric MRI and represents an important step to the clinical translation of GANs.

KEYWORDS
adversarial loss, convolutional neural network, fast imaging, multiple sclerosis, T_2-weighted FLAIR, VGG perceptual loss

1 | INTRODUCTION

High-resolution volumetric brain magnetic resonance imaging (MRI) is widely used in clinical and research applications to provide rich and detailed anatomical information and delineation of structural pathology. For example, three-dimensional (3D) T_1-weighted structural MRI exhibits high tissue contrast and is therefore routinely used to reconstruct cerebral cortical surfaces for cortical morphometry, analysis, and visualization.[1–4] Furthermore, T_2-weighted fluid-attenuated inversion recovery (FLAIR) imaging is highly sensitive to white matter abnormalities due to its excellent suppression of cerebrospinal fluid signal and is therefore routinely used to characterize lesion pathology in a wide range of neurological disorders.[5] A major barrier to the greater adoption of volumetric MRI in research and clinical protocols is the long acquisition time (typically ~5–7 min for 1 mm isotropic resolution and even longer for submillimeter resolution[6]), which may lead to subject anxiety and motion artifacts that compromise image quality, especially for children, elderly subjects, and some patient populations who cannot tolerate long scans.

Fast imaging techniques have been increasingly adopted to accelerate volumetric brain MRI. Parallel imaging methods such as sensitivity encoding (SENSE)[7] and generalized autocalibrating partial parallel acquisition (GRAPPA)[8] have been adopted by vendors and are used for routine volumetric MRI acquisition. However, these conventional methods can only achieve moderate acceleration factors (i.e., R = threefold along one phase encoding dimension and $R = 2 \times 2$ along both k_y and k_z axes in SENSE and GRAPPA) before suffering from severe image artifacts and noise amplification (i.e., high g-factor). Modern image acquisition and reconstruction methods such as compressed sensing,[9] low-rank modeling of local k-space neighborhoods (LORAKS),[10] bunched phase encoding (BPE),[11] 2D controlled aliasing in parallel imaging (CAIPIRINHA),[12] and wave-controlled aliasing in parallel imaging (Wave-CAIPI)[13] have been proposed to achieve even higher acceleration factors.

Among these fast imaging methods, Wave-CAIPI is a state-of-the-art technique that employs a corkscrew trajectory with CAIPI shifts in the k_y and k_z directions to efficiently encode k-space.[13] It better utilizes the available degrees of freedom of acquired information in different coils and uniformly spreads the voxel aliasing. Therefore, it can accelerate volumetric MRI by an order of magnitude with negligible g-factor and image artifact penalties.[13–15] Unfortunately, like images from other fast imaging techniques, the signal-to-noise ratio (SNR) ($\propto \sqrt{1/R}$) of highly accelerated Wave-CAIPI images is intrinsically lower compared to those acquired without acceleration or with mild acceleration factors due to substantially fewer acquired k-space signals.

Supervised deep learning using convolutional neural networks (CNNs) has superior performance to conventional methods for image restoration tasks such as super-resolution[16] and denoising[17] in digital and biomedical imaging. CNNs are remarkably effective in learning the underlying image priors and resolving the complex high-dimensional mapping from low-quality (e.g., low-SNR or low-resolution) images to high-quality images,[16–20] which is an ill-posed inverse problem. Even the simplest CNN for denoising (i.e., DnCNN[17]) achieves superior performance compared to the state-of-the-art block matching and 3D filtering (BM3D)[21] denoising method. Recently, it has been shown that DnCNN is also more advantageous in removing noise from submillimeter resolution T_1-weighted structural MR image volumes[22] compared to the state-of-the-art block matching with 4D filtering (BM4D)[23] and adaptive optimized nonlocal means (AONLM).[24] However, a well-known problem of CNNs trained using the voxel-wise error as the loss function is the tendency to generate blurry images that lack realistic textural details in both

super-resolution and denoising tasks, even though the mean squared error (MSE), mean absolute error (MAE) or related metrics such as peak SNR (PSNR) can be minimized or maximized,[25–27] respectively.

Generative adversarial networks (GANs)[28] have been shown to be effective in reducing the blurring effects and recovering realistic textures for digital photography,[25,29] optical coherence tomography,[30] microscopy,[31] X-ray computed tomography,[27,32] and MRI.[33–36] A GAN is composed of two subnetworks, including a generator and a discriminator, which are trained in synchrony to compete against each other. The generator generates high-quality images from the input low-quality images by minimizing the MSE (or the mean of other voxel-wise error measures) between the synthesized and target images and maximizing the probability that the synthesized image can be classified as the actual image (i.e., one probability value for one image slice/volume determined by the discriminator), while the discriminator tries to distinguish the synthesized images from the actual target images by minimizing the classification errors. Essentially, the discriminator loss, also known as adversarial loss, serves as a regularizer that forces the generator to synthesize images in the manifold of target images such that they are indistinguishable from target images with similar image features such as image contrast and textural details.

Nonetheless, several barriers limit the performance and feasibility of GANs for denoising empirical volumetric MRI data. Thus far, most studies have demonstrated the efficacy of 2D GANs consisting of a 2D generator and a 2D discriminator on image slices.[25,33,37] The advantage of 2D GANs is that their parameters can be optimized on only a few subjects, because each image volume from a subject provides millions of voxels as training samples for calculating the voxel-wise loss for the generator and hundreds of image slices as training samples for calculating the image-wise loss for the discriminator. However, the image synthesis performance of 2D generators is often limited compared to 3D generators, which can incorporate complementary information from an additional spatial dimension[22,38,39] (Figure S1). Moreover, there may be boundary artifacts across synthesized 2D image slices along the cross-sectional directions.

Consequently, 3D CNNs are more often adopted for volumetric data[20,22,38–40] and 3D GANs consisting of a 3D generator and a 3D discriminator have been also proposed.[34–36] Unfortunately, training a 3D GAN requires data from a large number of subjects since the data of each subject can only provide one image volume as a training sample for the 3D discriminator (or several blocks with smaller size as training samples), which impedes the use of 3D GANs in practice. For example, Chen and coworkers[34,36] needed data from 1113 subjects to train a 3D GAN for brain MRI super-resolution. Ran et al.[35] had to not only split the whole-brain volumes into small blocks (i.e., $32 \times 32 \times 6$ voxels) to increase the training sample number while restricting the receptive field of the adopted CNN, but also use a suboptimal network (i.e., a very shallow three-layer discriminator) with compromised classification performance comparing to that of very deep discriminators, which allowed them to train a 3D GAN for MRI denoising with data from 100 subjects. Because of the intensive data requirement, the training of 3D GANs in these studies had to be performed using simulated training data in which the degraded images were generated by adding noise to high-SNR images or down-sampling high-resolution images from public databases.[34–36] However, it is unclear how well 3D GANs trained on simulated data generalize to empirical data and whether the resultant images are as good as those from 3D GANs trained on actually acquired data, especially for super-resolution.

To improve the performance and feasibility of GANs for empirical volumetric MRI data, we propose a hybrid denoising GAN architecture (entitled "HDnGAN") consisting of a 3D generator and a 2D discriminator, which concurrently benefits from the improved image synthesis performance provided by the 3D generator and increased training samples from a limited number of subjects for training the 2D discriminator. Benefiting from this new architecture, we demonstrate the efficacy of HDnGAN on empirical Wave-CAIPI T_2-SPACE FLAIR volumetric data acquired in 33 multiple sclerosis (MS) patients, in distinction to existing studies that only used simulated data from healthy subjects.[27,34–36] The systematic and comprehensive comparison and assessment using different quantitative metrics and by two neuroradiologists show that HDnGAN results are similar to standard FLAIR data acquired in longer scan time, and outperform the raw Wave-CAIPI images and those from conventional denoising methods including BM4D and AONLM and a 3D GAN trained on the same training data. In addition, we characterize and elucidate the effects of adversarial loss on resultant images and propose to utilize this mechanism to control the sharpness and texture richness of resultant images from GANs by adjusting the contribution of the adversarial loss. In summary, our study with empirical patient data and systematic evaluation demonstrates the feasibility of using GANs to denoise empirical volumetric MRI data and supports their use in combination with modern fast imaging techniques such as Wave-CAIPI adopted in this study to achieve high-fidelity, fast volumetric MRI, which represents an important step toward the clinical translation of GANs.

2 | MATERIALS AND METHODS

2.1 | Generative adversarial network

HDnGAN consists of a 3D generator and a 2D discriminator (Figure 1). The generator synthesizes image volumes similar to target high-SNR image volumes

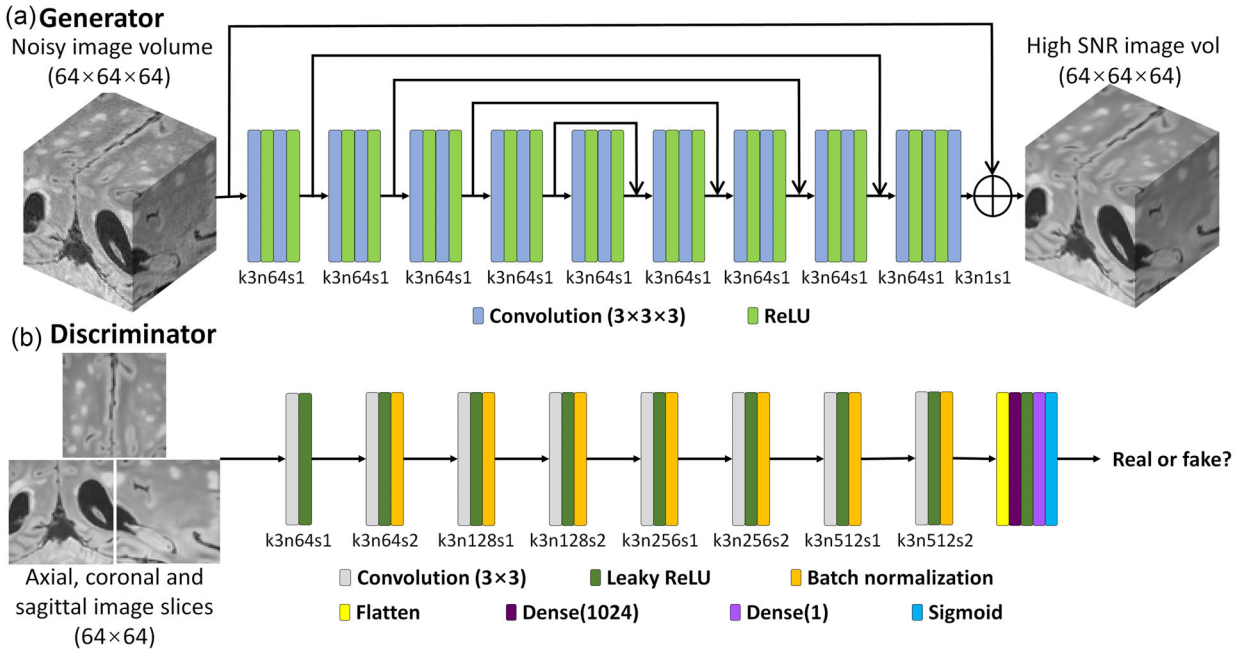

FIGURE 1 Hybrid denoising generative adversarial network (HDnGAN) architecture. HDnGAN consists of a 3D generator (a) and a 2D discriminator (b). The 3D generator is modified from 3D U-Net by removing all max-pooling, up-sampling, and batch normalization layers and keeping the number of kernels constant ($n = 64$) across all layers (denoted as MU-Net). The input of the generator is a noisy image volume ($64 \times 64 \times 64$ voxels). The output of the generator is a synthesized image volume with high signal-to-noise ratio (SNR). The 2D discriminator adopts the discriminator of Super-Resolution Generative Adversarial Network (SRGAN), with spectral normalization incorporated in each layer to stabilize training. The input of the discriminator is an axial, coronal, or sagittal image slice (64×64 pixels) from the image volume synthesized by the generator or the ground-truth high-SNR image volume. The output of the discriminator is a probability value of the input image slice being classified as a real high-SNR image slice. The abbreviation k3n64s1 stands for a kernel size equal to $3 \times 3 \times 3$ for the generator or 3×3 for the discriminator, a kernel number equal to 64 and stride equal to 1, and so on.

from input noisy image volumes, while the discriminator tries to distinguish synthesized image slices from the acquired target high-SNR image slices. The generator and the discriminator are trained in synchrony to compete against each other.

2.1.1 | Network architecture

A modified 3D U-Net[41] (denoted as MU-Net) is used as the generator (Figure 1a), which predicts the residuals between noisy input image volumes and target high-SNR image volumes (i.e., residual learning). The input and output image volumes consist of $d \times d \times d$ voxels ($d = 64$ in this study due to the limited memory of the graphics processing unit (GPU)). Specifically, all max pooling, up-sampling, and batch normalization layers are removed, and the number of kernels in each layer is kept constant ($n = 64$ in this study). MU-Net creates several short paths from early layers to later layers to alleviate the vanishing-gradient problem and strengthen feature propagation with a moderate number of parameters (~2.3 million parameters). It represents an intermediate network between a plain network (e.g., Very Deep Super-Resolution (VDSR)[42] and DnCNN[17]) without any short paths and a densely connected network (e.g., DenseNet[43]) that comprehensively connects each layer to every other layer.

The discriminator of Super-resolution generative adversarial Network (SRGAN)[25] is adopted (~13.1 million parameters) (Figure 1b) with spectral normalization incorporated in each layer to stabilize training.[44] The discriminator is designed to classify 2D image slices rather than 3D blocks to increase the number of training samples and ensure that the HDnGAN could be optimized on data from a limited number of subjects. The discriminator uses all axial, coronal, and sagittal image slices of the resultant 3D image blocks from the generator as separate samples during the training. In this way, one image volume could provide $3 \times d$ image slices as training samples to optimize the discriminator parameters, where d is the size of the input image volume ($d = 64$ in this study).

2.1.2 | Loss functions

The generator and the discriminator are trained alternately. When the generator is trained, the parameters of the discriminator are fixed. The training process tries to optimize the parameters of the generator (θ_G) such that the synthesized image volumes from the generator are

similar to the target high-SNR image volumes. Specifically, the optimizer tries to solve:

$$\underset{\theta_G}{\text{minimize}}\ \mathcal{L}^G_{\text{content}} + \lambda \mathcal{L}^G_{\text{adversarial}}, \quad (1)$$

where $\mathcal{L}^G_{\text{content}}$ is the voxel-wise MSE between the synthesized image volumes and the target high-SNR image volumes, $\mathcal{L}^G_{\text{adversarial}}$ relates to the probability that a synthesized image slice could be classified as a real acquired high-SNR image slice, and λ determines the contribution of the adversarial loss $\mathcal{L}^G_{\text{adversarial}}$ to the total loss $\mathcal{L}^G$, which is a weighted summation of the content loss $\mathcal{L}^G_{\text{content}}$ and the adversarial loss $\mathcal{L}^G_{\text{adversarial}}$. When $\lambda = 0$, $\mathcal{L}^G = \mathcal{L}^G_{\text{content}}$, and the HDnGAN is essentially the generator (i.e., MU-Net) without the discriminator. When $\lambda = \infty$, $\mathcal{L}^G = \mathcal{L}^G_{\text{adversarial}}$. Intermediate λ values achieve mixed content loss and adversarial loss.

The content loss from the generator is defined as the voxel-wise MSE between the synthesized image volumes and the target high-SNR image volumes, calculated as:

$$\mathcal{L}^G_{\text{content}} = \frac{1}{N_{\text{vol}}} \sum_{n=1}^{N_{\text{vol}}} \left\{ \frac{1}{N_{\text{vox}}} \sum_{i,j,k \in \mathcal{M}} \left[I^{\text{target}}_{i,j,k} - G^{3D}_{\theta_G}(I^{\text{noisy}})_{i,j,k} \right]^2 \right\}, \quad (2)$$

where I^{target} and I^{noisy} denote target high-SNR and input noisy image volumes; $G^{3D}_{\theta_G}$ is the 3D generator parametrized by θ_G; i, j, k are coordinates within the mask $\mathcal{M}$ of interest (herein referred to as the "loss mask") for calculating the loss; N_{vox} is the number of voxels within the loss mask for each volume; N_{vol} is the number of volumes for training.

The adversarial loss is defined as:

$$\mathcal{L}^G_{\text{adversarial}} = \sum_{n=1}^{N_{\text{vol}}} \sum_{m=1}^{N_{\text{acs}}} -\log D^{2D}_{\theta_D}\{S_m[G^{3D}_{\theta_G}(I^{\text{noisy}})]\}, \quad (3)$$

where $D^{2D}_{\theta_D}$ is the 2D discriminator parametrized by θ_D (fixed during the training of the generator); $S_m(:)$ denotes the operation of selecting the mth slice out of all N_{acs} axial, coronal and sagittal slices from the synthesized image block $G^{3D}_{\theta_G}(I^{\text{noisy}})$; $D^{2D}_{\theta_D}\{S_m[G^{3D}_{\theta_G}(I^{\text{noisy}})]\}$ is the probability that an image slice is classified as a real acquired target high-SNR image slice.

When the discriminator is trained, the parameters of the generator are fixed. The training process tries to optimize the parameters of the discriminator (θ_D) such that it could accurately discriminate the synthesized and actual images. Specifically, each of the axial, coronal, and sagittal slices from the denoised image volumes synthesized by the generator are labeled as 0 and the corresponding real high-SNR image slices are labeled as 1. The optimizer tries to solve:

$$\underset{\theta_D}{\text{minimize}}\ \mathcal{L}^D, \quad (4)$$

where $\mathcal{L}^D$is the binary cross-entropy loss between true labels and predicted labels from the discriminator, defined as:

$$\mathcal{L}^D = -\sum_{n=1}^{N_{\text{vol}}} \sum_{m=1}^{N_{\text{acs}}} \log D^{2D}_{\theta_D}[S_m(I^{\text{target}})] - \log D^{2D}_{\theta_D}\{S_m[G^{3D}_{\theta_G}(I^{\text{noisy}})]\}. \quad (5)$$

For better gradient behavior, $-\log D^{2D}_{\theta_D}\{S_m[G^{3D}_{\theta_G}(I^{\text{noisy}})]\}$ was minimized rather than $\log\{1 - D^{2D}_{\theta_D}\{S_m[G^{3D}_{\theta_G}(I^{\text{noisy}})]\}\}$.[25]

2.2 | Experiments

2.2.1 | Data acquisition

This study was approved by the Mass General Brigham institutional review board and was HIPAA compliant. With written informed consent, data were acquired in 33 patients (20 for training, five for validation, and eight for evaluation) undergoing clinical evaluation for demyelinating disease at the Massachusetts General Hospital as part of a separate clinical validation study of Wave-CAIPI FLAIR compared to standard 3D T_2-SPACE FLAIR.[45] All patients were scanned on a whole-body 3-T MAGNETOM Prisma MRI scanner (Siemens Healthcare) equipped with a 20-channel head coil.

Standard FLAIR data were acquired using a 3D T_2-SPACE FLAIR sequence with the following parameters: repetition time = 5000 ms, echo time = 390 ms, inversion time = 1800 ms, flip angle = 120°, 176 sagittal slices, slice thickness = 1 mm, field of view = 256 × 256 mm^2, resolution = 1 mm isotropic, bandwidth = 650 Hz/pixel, GRAPPA factor = 2, and acquisition time = 7.25 min. Wave-CAIPI FLAIR data were acquired using a prototype sequence with matched parameter values except for the following parameters: echo time = 392 ms, bandwidth = 750 Hz/pixel, acceleration factor = 3 × 2, and acquisition time = 2.75 min.

2.2.2 | Image processing

Wave-CAIPI and standard FLAIR images were processed for HDnGAN training, validation, and evaluation. Standard FLAIR images were nonlinearly co-registered to Wave-CAIPI images using the "*reg_f3d*" function from the NiftyReg software,[46,47] which was initialized with an affine transformation derived from NiftyReg's "*reg_aladin*" function. The nonlinear co-registration

slightly adjusted the image alignment locally, which was employed to account for the subtle nonlinear shifts of tissue in the images due to factors such as subject bulk motion and the associated changes in distortions due to changing B0 inhomogeneity and gradient nonlinearity distortion during the acquisition of each repetition of data. The resampling used cubic spline interpolation.

Brain masks were created from Wave-CAIPI images using the unified segmentation algorithm in the Statistical Parametric Mapping software.[48] The Wave-CAIPI and standard FLAIR image volumes image volumes were brain masked to exclude regions such as the skull and the background, which are irrelevant to the image content within the brain. To account for subject-to-subject variations in image intensity, the image intensities of the Wave-CAIPI and standard FLAIR image volumes were standardized by subtracting the mean intensity and then dividing by the standard deviation of the image intensities of the voxels within the brain mask in the Wave-CAIPI data.

For each subject, a binary mask was created to exclude parts of the frontal lobe, temporal lobe, cerebellum, and brainstem where the residuals between the Wave-CAIPI and standard FLAIR images were dominated by large image artifacts and geometric distortions rather than noise. These artifactual regions could not be well aligned by image co-registration and would slightly decrease the performance of the CNNs if included since the noise characteristics within these regions are different from those in other brain regions. To create such a mask for each subject, the absolute difference between the standardized Wave-CAIPI and co-registered standard FLAIR images was blurred using a Gaussian kernel with a standard deviation of 2 mm and then binarized using a threshold of 0.04. These parameters were empirically selected and confirmed by visual inspection of all subjects. Only the MSE within this mask was used to optimize the generator (i.e., the "loss mask" in Equation (2)).

2.2.3 | HDnGAN training

HDnGAN was used to perform image quality transfer from Wave-CAIPI to standard FLAIR images. HDnGAN was implemented using the Keras application program interface (https://keras.io) with a Tensorflow backend (https://www.tensorflow.org) in Python, trained on data from 20 patients, and validated on data from additional five patients using an NVIDIA GeForce RTX 2080 Ti GPU. To facilitate further studies, we made the source codes of HDnGAN freely and publicly available on GitHub at https://github.com/liziyu0929/HDnGAN.

Because of the limited GPU memory, the generator was trained on input and output image blocks consisting of $64 \times 64 \times 64$ voxels, and the discriminator was trained on input image slices consisting of 64×64 voxels (64×3 axial, coronal, and sagittal image slices from each block). For each subject, 18–27 overlapped blocks were selected, which were evenly distributed within the brain mask with overlap in a sliding window fashion. All training data were flipped along the anatomical left–right direction for augmentation.

Network parameters were optimized as described in Equations (1) and (4) using an Adam optimizer with default parameter values except for the learning rates which were set to 5×10^{-5} and 2×10^{-4} for the generator and discriminator, respectively. The generator and the discriminator were trained alternately for each batch (one time for the generator and three times for the discriminator). HDnGAN with 10 different λ values (i.e., 0, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, ∞) were trained and validated for determining the optimal λ value, each for 21 epochs and ~24 hours. For $\lambda = \infty$, the content loss was simply excluded. For $\lambda = 0$, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10, ∞, HDnGAN approached convergence after 21 epochs (Figure S2) while for $\lambda = 10^{-1}$, 1, 10^2, HDnGAN approached convergence after 15 epochs.

2.2.4 | Image denoising using other methods

For comparison, Wave-CAIPI image volumes were also denoised by other state-of-the-art denoising methods, including AONLM and BM4D, and a 3D GAN. The AONLM denoising was performed using the publicly available MATLAB-based software (https://sites.google.com/site/pierrickcoupe/softwares/denoising/mri-denoising/mri-denoising-software) assuming Rician noise with $3 \times 3 \times 3$ block and $7 \times 7 \times 7$ search volume.[24,49] The BM4D denoising was performed using the publicly available MATLAB-based software (https://www.cs.tut.fi/~foi/GCF-BM3D) assuming Rician noise with the collaborative Wiener filtering and the "modified profile" option with default parameters. The parameters of the 3D GAN were the same as those from HDnGAN, except for that the 3D discriminator classified $64 \times 64 \times 64$ generated and ground-truth blocks rather than 64×64 image slices. The 3D GAN with $\lambda = 10^{-3}$ was trained in the same way as the HDnGAN training (e.g., identical training and validation data, 21 epochs for ~24 h using an Adam optimizer), expect for that the learning rate of the generator was set to 10^{-4}. To demonstrate the advantages of 3D generators over 2D generators, a 2D MU-Net for denoising 64×64 image slices was also trained to compare to our 3D MU-Net. The parameters and the training for the 2D MU-Net were identical to those of the 3D MU-Net, except for that the convolutional kernels were 2D.

2.2.5 | Image quality evaluation

Data from eight separate patients not included in the training were used for evaluation. For each evaluation

subject, the optimized HDnGAN and 3D GAN was applied to the extracted image blocks. The resultant denoised image blocks were assembled to create the denoised brain image volume. The overlapping regions of the denoised blocks were averaged.

The group mean and the group standard deviation of the MSE and VGG perceptual loss[25] across the eight evaluation subjects were used to quantify the similarity between Wave-CAIPI, AONLM-denoised, BM4D-denoised, MU-Net-denoised, 3D GAN-denoised, and HDnGAN-denoised images and standard FLAIR images. For each subject, the MSE was calculated as the mean of the squared differences of the standardized image intensities of all voxels within the brain mask between the denoised and the standard FLAIR image volumes. VGG perceptual loss was calculated as the Euclidean distance between the 4096×1 feature vectors extracted from the denoised and the standard FLAIR image slices using the VGG16 network pretrained on ImageNet natural image dataset for object classification.[50] For each subject, the VGG loss was calculated for each axial, coronal, and sagittal slice from the brain volume and then averaged. The VGG loss closely reflects human perception, and a lower VGG loss value indicates the two images for comparison are visually more similar. For HDnGAN, MSE, and VGG perceptual loss were calculated for each weight of adversarial loss.

Images from different methods were also evaluated by two neuroradiologists (Chanon Ngamsombat and Samuel Cartmell) based on four image quality metrics, including image sharpness, SNR, lesion conspicuity, and overall quality. For HDnGAN and 3D GAN, results generated using $\lambda = 10^{-3}$ were used for comparison since they had the lowest VGG perceptual loss compared to the standard FLAIR images. The two readers used the standard FLAIR images as the ground truth for evaluation but were blinded to images from the other methods. Images were scored using a five-point scale: 1 nondiagnostic, 2 limited, 3 diagnostic, 4 good, 5 excellent. The group mean and group standard deviation of the scores from each type of images were computed and compared. Moreover, *t*-tests were performed to assess whether there were significant differences between the image quality scores for different methods. The Benjamini–Hochberg procedure was performed to account for multiple comparisons assuming a 5% false discovery rate.

2.3 | Simulation study

A simulation study was performed to demonstrate the efficacy of HDnGAN on noisier images (I_N), which were generated by adding simulated Rician noise to Wave-CAIPI data as:

$$I_N(x_i) = \sqrt{I_r(x_i)^2 + I_i(x_i)^2}, \quad (6)$$

where $I_r(x_i) = I_0(x_i) + n_1(x_i)$; $I_i(x_i) = n_2(x_i)$; the noise n followed a Gaussian distribution with zero mean and standard deviation σ; I_0 is Wave-CAIPI image volumes and x_i represents the coordinates of each voxel. The noise level was described in terms of the mean image intensity ($\bar{S}$) within the brain mask. Noisier data with two different noise levels (i.e., $\sigma = 3\%\ \bar{S}, 12\%\ \bar{S}$) were simulated, which were denoised by AONLM, BM4D, MU-Net, and HDnGAN ($\lambda = 10^{-3}$) as described in previous sections. The denoised results were compared in terms of MSE and VGG perceptual loss.

3 | RESULTS

The effect of the contribution of adversarial loss on the resultant images is shown in Figure 2. For HDnGAN with low λ values (i.e., 0, 10^{-5}, 10^{-4}) when MSE dominated the loss function, the denoised images were smooth with reduced richness of high-frequency textures (Figure 2, rows a and b, columns iii–v), as expected. When λ gradually increased, more textural details became apparent, and the denoised images became visually more similar to the standard FLAIR images. Images from HDnGAN around $\lambda = 10^{-3}$ (Figure 2, rows a and b, column vi) were visually most similar to the standard FLAIR images.

The optimal λ varied according to different image similarity metrics (Figure 3). In terms of MSE, HDnGAN ($\lambda = 0$) (i.e., MU-Net) achieved the best performance ($5.92 \times 10^{-4} \pm 0.56 \times 10^{-4}$), substantially better than the Wave-CAIPI input images ($1.08 \times 10^{-3} \pm 0.12 \times 10^{-3}$) and those generated with adversarial losses ($\lambda > 0$) (Figure 3a), which was consistent with the knowledge that denoising CNNs optimized with voxel-wise MSE could minimize MSE but tend to generate blurry images that lack realistic textures (Figure 2, rows a and b, column iii). When λ gradually increased from zero, MSE increased first, reached a local maximum at $\lambda = 10^{-1}$ and then stayed similar after that. In other words, the adversarial loss served to introduce more textural details into resultant images (Figure 2), at the cost of increased MSE.

HDnGAN ($\lambda = 10^{-3}$) achieved the best performance in terms of VGG loss ($1.09 \times 10^{-2} \pm 0.18 \times 10^{-2}$) (Figure 3b), improving significantly upon Wave-CAIPI inputs ($2.03 \times 10^{-2} \pm 0.33 \times 10^{-2}$) and MU-Net ($2.45 \times 10^{-2} \pm 0.44 \times 10^{-2}$). As a result, HDnGAN ($\lambda = 10^{-3}$) was used for comparison with other methods. When λ gradually increased from zero, VGG perceptual loss first decreased (along with more textural details visible in resultant images in Figure 2), reached the minimum at $\lambda = 10^{-3}$, then increased slightly and stayed similar after that, which was consistent with visual inspection of denoised images (Figure 2).

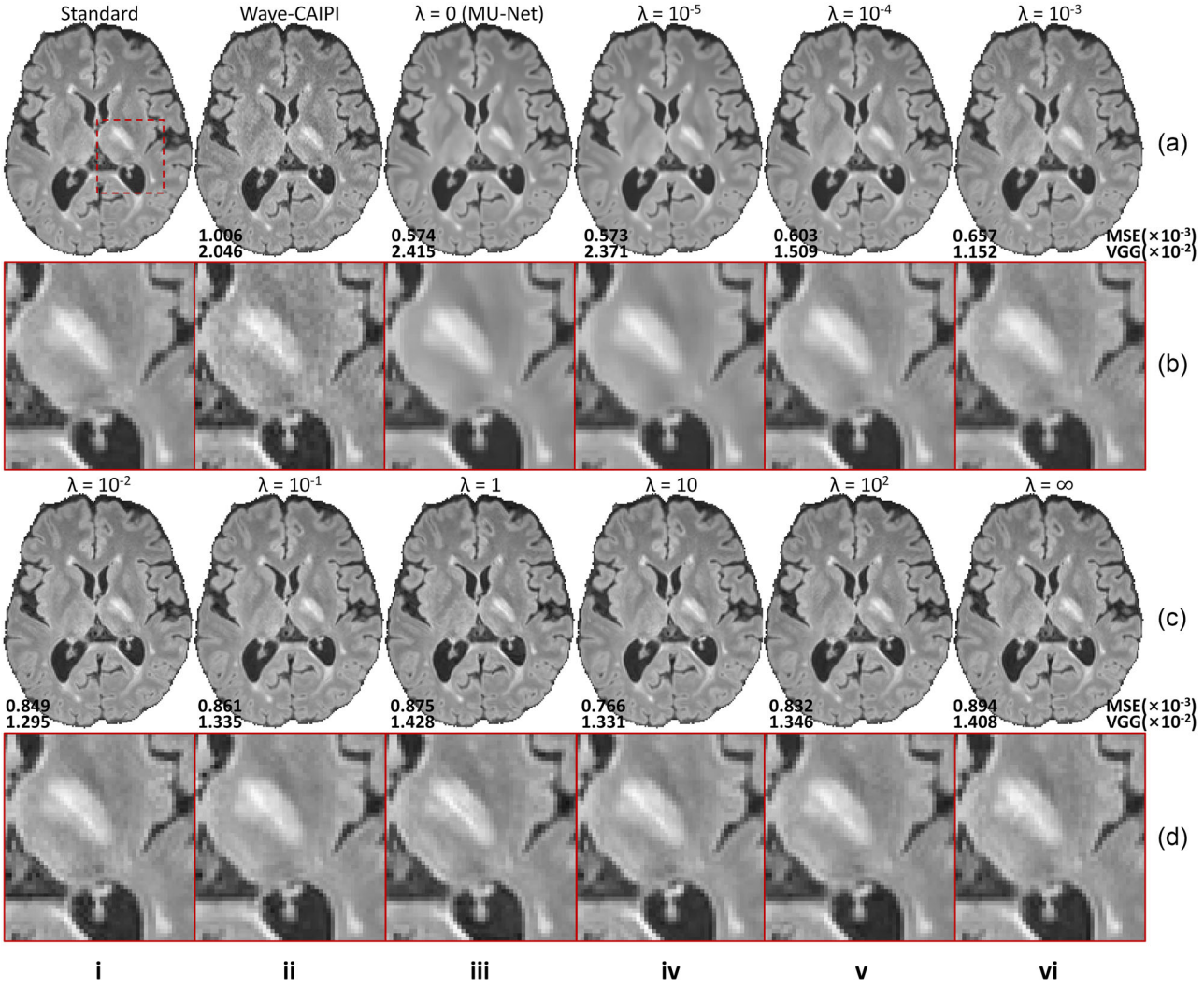

FIGURE 2 Effects of the adversarial loss on image quality. Representative axial image slices (rows a and c) and enlarged views of the left basal ganglia and left thalamus (rows b and d) from different methods and weights (λ) of the adversarial loss in a multiple sclerosis patient for evaluation. For $\lambda = 0$ (rows a, b, column iii), the training only minimizes the content loss (i.e., voxel-wise mean squared error). In this case, the hybrid denoising generative adversarial network (HDnGAN) is effectively the generator (i.e., modified U-Net (MU-Net)). For $\lambda = \infty$, the training only minimizes the adversarial loss (rows c, d, column vi). Image similarity metrics including the mean squared error (MSE) and VGG perceptual loss (VGG) are listed to quantify the similarity between images from different methods and the standard fluid-attenuated inversion recovery (FLAIR) image.

Denoised images from different methods of a representative subject are shown in Figure 4 for comparison (results from axial and coronal view are available in Figures S3 and S4). The residual maps between the denoised and standard FLAIR images did not contain anatomical structures (Figure S5). All denoised images exhibited higher SNR than the Wave-CAIPI input images. MU-Net-denoised images were the smoothest with reduced amount of texture. The results from the two conventional methods AONLM and BM4D were visually similar and slightly sharper than the MU-Net results. HDnGAN ($\lambda = 10^{-3}$) generated high-quality image volumes with rich and realistic textural details, which were slightly sharper than those from the 3D GAN ($\lambda = 10^{-3}$). This is potentially because the current training and validation data from 25 subjects were not sufficient to optimize the 3D GAN parameters.

The group mean and the group standard deviation of the MSE and VGG loss across the eight evaluation subjects were listed in Table 1. Among all methods for comparison, MU-Net-denoised images achieved the best MSE ($5.92 \times 10^{-4} \pm 0.56 \times 10^{-4}$) but also the worst VGG loss ($2.45 \times 10^{-2} \pm 0.44 \times 10^{-2}$). HDnGAN ($\lambda = 10^{-3}$) generated images with low MSE ($7.43 \times 10^{-4} \pm 0.94 \times 10^{-4}$), which was comparable to those from BM4D ($7.18 \times 10^{-4} \pm 0.75 \times 10^{-4}$) and AONLM ($7.49 \times 10^{-4} \pm 0.76 \times 10^{-4}$) and lower than that from 3D GAN ($\lambda = 10^{-3}$)

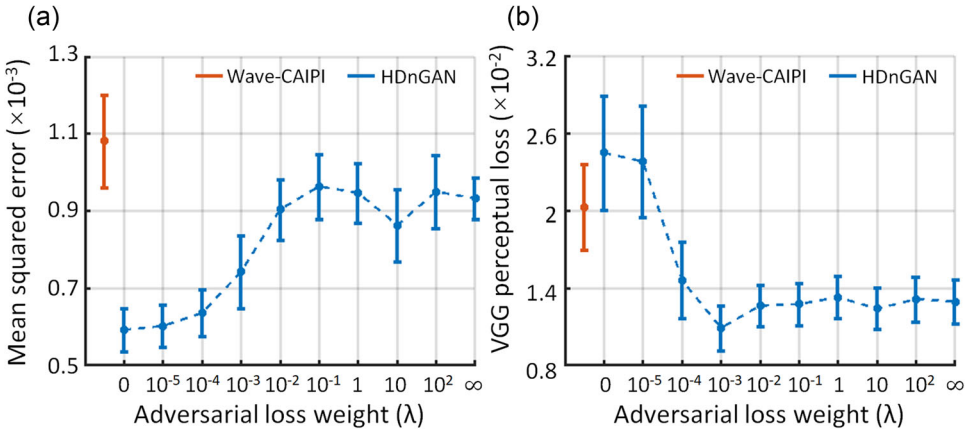

FIGURE 3 Quantification of the effects of the adversarial loss. The similarity between images derived from different methods and λ values and the standard fluid-attenuated inversion recovery (FLAIR) images is quantified using the mean squared error (a) and VGG perceptual loss (b). The red and blue dots and error bars represent the group mean and group standard deviation of different metrics for wave-controlled aliasing in parallel imaging (Wave-CAIPI) images and results of hybrid denoising generative adversarial network (HDnGAN) trained with different weights for the adversarial loss. The metrics were calculated from eight patients for evaluation.

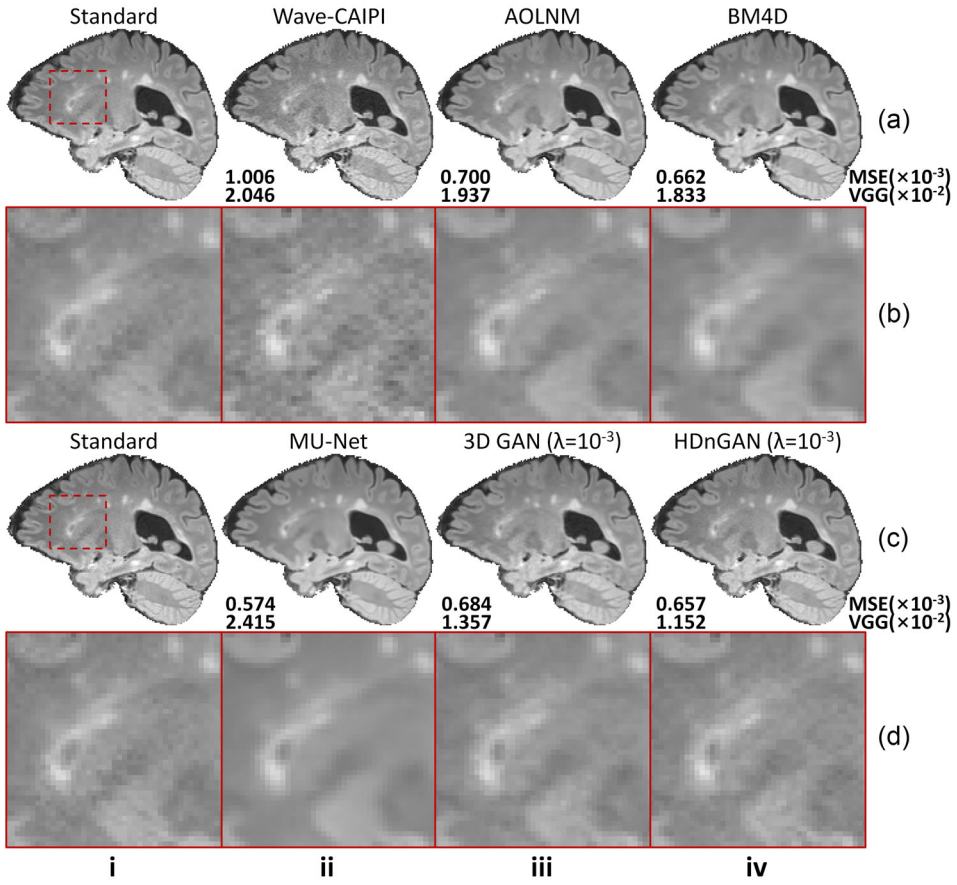

FIGURE 4 Visual comparison of results from different methods. Representative sagittal image slices (rows a and c) and enlarged regions (rows b and d) from standard fluid-attenuated inversion recovery (FLAIR) data (rows a, b, c, d, column i), wave-controlled aliasing in parallel imaging (Wave-CAIPI) data (rows a, b, column ii), adaptive optimized nonlocal means (AONLM)-denoised results (rows a, b, column iii), block matching with 4D filtering (BM4D)-denoised results (rows a, b, column iv), modified U-Net (MU-Net) results (rows c, d, column ii), 3D generative adversarial network (GAN) ($\lambda = 10^{-3}$) results (rows c, d, column iii), and hybrid denoising GAN (HDnGAN) ($\lambda = 10^{-3}$) results (rows c, d, column iv) from one evaluation subject. Image similarity metrics including the mean squared error (MSE) and VGG perceptual loss (VGG) are listed to quantify the similarity between images from different methods and the standard FLAIR image.

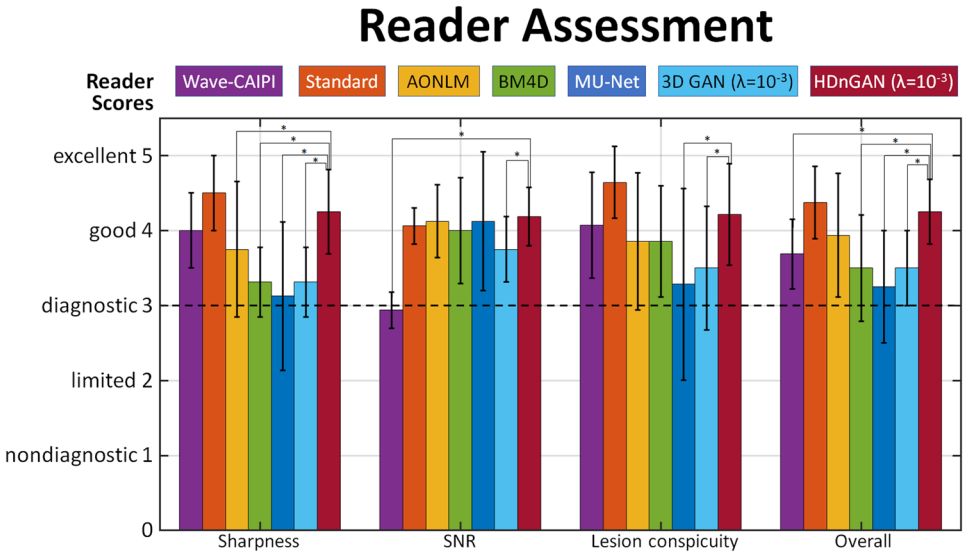

FIGURE 5 Multi-reader assessment of results from different methods. The group mean and group standard deviation of image quality (sharpness, signal-to-noise ratio (SNR), lesion conspicuity, and overall quality) scores (1 nondiagnostic, 2 limited, 3 diagnostic, 4 good, 5 excellent) from two radiologists for the standard, wave-controlled aliasing in parallel imaging (Wave-CAIPI), adaptive optimized nonlocal means (AONLM), block matching with 4D filtering (BM4D), modified U-Net (MU-Net), 3D generative adversarial network (GAN) ($\lambda = 10^{-3}$), and hybrid denoising GAN (HDnGAN) ($\lambda = 10^{-3}$) images of eight evaluation subjects for evaluation. Comparisons between scores of HDnGAN ($\lambda = 10^{-3}$) and scores of other methods surviving multiple comparisons correction at a false discovery rate threshold of 0.05 are denoted with an asterisk.

TABLE 1 Quantitative comparison of results from different methods

Method	Mean squared error ($\times 10^{-3}$)	VGG perceptual loss ($\times 10^{-2}$)
Wave-CAIPI	1.082 ± 0.120	2.031 ± 0.334
AONLM	0.749 ± 0.076	1.687 ± 0.271
BM4D	0.718 ± 0.075	1.716 ± 0.230
MU-Net	**0.592 ± 0.056**	2.453 ± 0.444
3D GAN ($\lambda = 10^{-3}$)	0.775 ± 0.095	1.349 ± 0.179
HDnGAN ($\lambda = 10^{-3}$)	0.743 ± 0.094	**1.091 ± 0.175**

Note: The group mean and group standard deviation of image similarity metrics (mean ± std) including the mean squared error and VGG perceptual loss are listed to quantify the similarity between images from different methods and the standard fluid-attenuated inversion recovery (FLAIR) image. The methods with the best performance for each metrics are marked in bold. The metrics were calculated from eight patients for evaluation.

Abbreviations: AONLM, adaptive optimized nonlocal means; BM4D, block matching with 4D filtering; GAN, generative adversarial network; HDnGAN, hybrid denoising generative adversarial network; MU-Net, modified U-Net; Wave-CAIPI, wave-controlled aliasing in parallel imaging.

($7.75 \times 10^{-4} \pm 0.95 \times 10^{-4}$). HDnGAN ($\lambda = 10^{-3}$) achieved the best VGG loss ($1.09 \times 10^{-2} \pm 0.18 \times 10^{-2}$), which was substantially lower than those from BM4D ($1.72 \times 10^{-2} \pm 0.23 \times 10^{-2}$), AONLM ($1.69 \times 10^{-2} \pm 0.27 \times 10^{-2}$), and 3D GAN ($\lambda = 10^{-3}$) ($1.35 \times 10^{-2} \pm 0.18 \times 10^{-2}$). Finally, HDnGAN ($\lambda = 10^{-3}$) outperformed 3D GAN ($\lambda = 10^{-3}$) in terms of both MSE ($7.43 \times 10^{-4} \pm 0.94 \times 10^{-4}$ vs. $7.75 \times 10^{-4} \pm 0.95 \times 10^{-4}$) and VGG loss ($1.09 \times 10^{-2} \pm 0.18 \times 10^{-2}$ vs. $1.35 \times 10^{-2} \pm 0.18 \times 10^{-2}$).

The results of the reader study (Figure 5, Table S1) showed that HDnGAN ($\lambda = 10^{-3}$) significantly outperformed AONLM (4.25 ± 0.56 vs. 3.75 ± 0.90, $p = 0.015$), BM4D (4.25 ± 0.56 vs. 3.31 ± 0.46, $p < 0.001$), MU-Net (4.25 ± 0.56 vs. 3.13 ± 0.99, $p < 0.001$), and 3D GAN ($\lambda = 10^{-3}$) (4.25 ± 0.56 vs. 3.31 ± 0.46, $p < 0.001$) in terms of image sharpness. HDnGAN ($\lambda = 10^{-3}$) also significantly improved the input Wave-CAIPI images' SNR (4.19 ± 0.39 vs. 2.94 ± 0.24, $p < 0.001$) and outperformed 3D GAN ($\lambda = 10^{-3}$) (4.19 ± 0.39 vs. 3.75 ± 0.43, $p = 0.004$). The lesion conspicuity of the generated images from HDnGAN ($\lambda = 10^{-3}$) was significantly better than those from MU-Net (4.21 ± 0.67 vs. 3.29 ± 1.28, $p < 0.001$) and 3D GAN ($\lambda = 10^{-3}$) (4.21 ± 0.67 vs. 3.5 ± 0.82, $p = 0.001$). The overall scores from HDnGAN ($\lambda = 10^{-3}$) were significantly higher than those from Wave-CAIPI (4.25 ± 0.43 vs. 3.69 ± 0.46, $p = 0.003$), BM4D (4.25 ± 0.43 vs. 3.50 ± 0.71, $p = 0.001$), MU-Net (4.25 ± 0.43 vs. 3.25 ± 0.75, $p < 0.001$), and 3D GAN ($\lambda = 10^{-3}$) (4.25 ± 0.43 vs. 3.50 ± 0.50, $p < 0.001$), with no significant difference compared to the standard FLAIR images (4.25 ± 0.43 vs. 4.38 ± 0.48, $p = 0.333$).

The advantages of HDnGAN became more obvious for noisier data in the simulation study (Figure 6, Table 2). With moderate noise ($\sigma = 3\%$ $\bar{S}$) added, the blurring effects of AONLM, BM4D, and MU-Net became more obvious (Figure 6, rows a and b, columns iii-v), while HDnGAN ($\lambda = 10^{-3}$) improved the image SNR and preserved realistic textures (Figure 6, rows a and b, columns vi), achieving the lowest VGG perceptual loss ($1.336 \times 10^{-2} \pm 0.186 \times 10^{-2}$). MU-Net achieved the lowest MSE ($0.703 \times 10^{-3} \pm 0.065 \times 10^{-3}$). At a very high

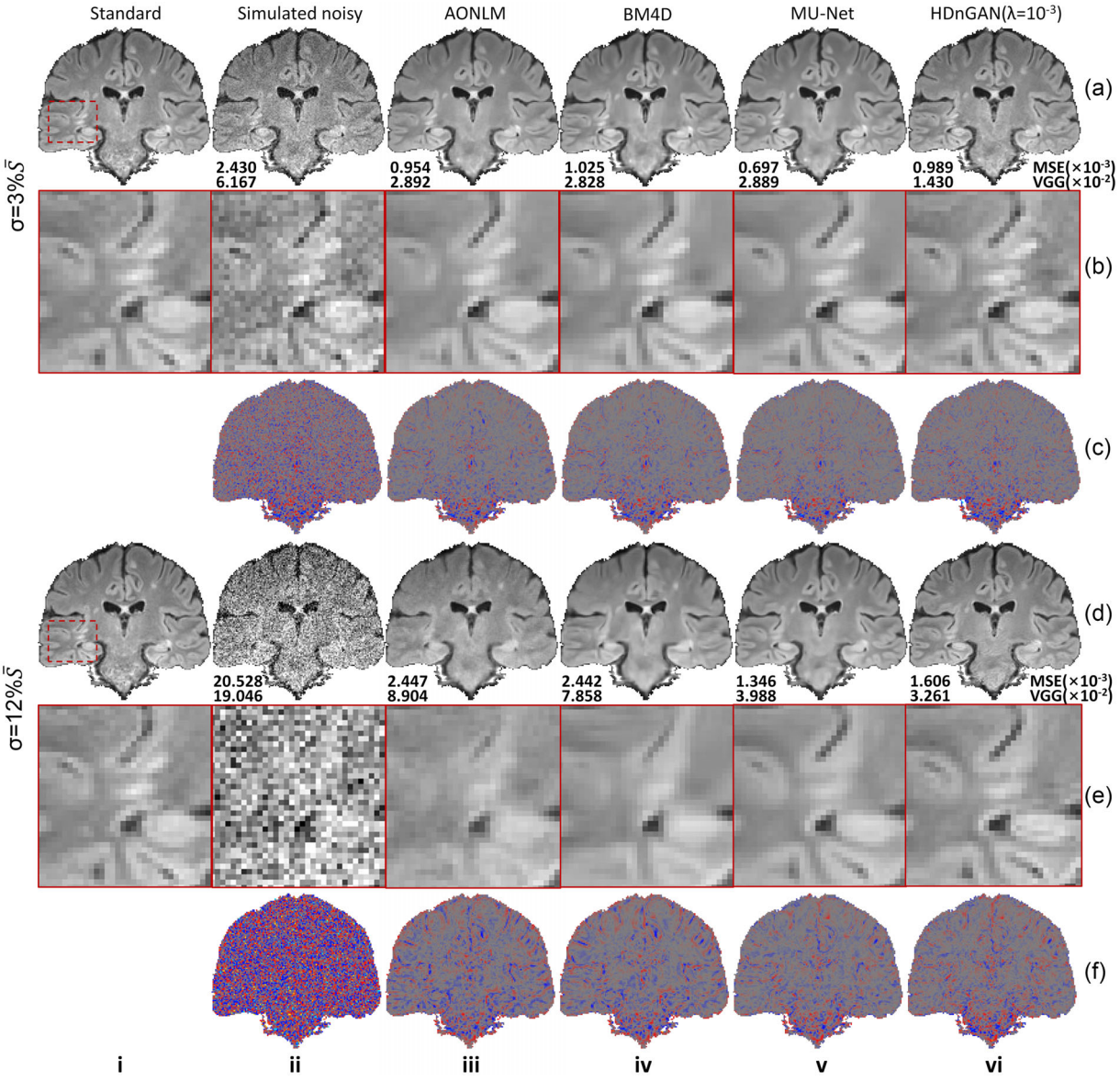

FIGURE 6 Visual comparison of results from different methods at higher noise levels. Representative coronal image slices (rows a and d), enlarged regions (rows b and e) and residual maps to standard fluid-attenuated inversion recovery (FLAIR) data (rows c and f) from the standard FLAIR data (column i), simulated noisy data (column ii), adaptive optimized nonlocal means (AONLM)-denoised data (column iii), block matching with 4D filtering (BM4D)-denoised data (column iv), modified U-Net (MU-Net)-denoised data (column v), and hybrid denoising generative adversarial network (HDnGAN)-denoised ($\lambda = 10^{-3}$) data (column vi) for different levels of simulated noise from one evaluation subject. The noise level is described in terms of the standard deviation (σ) of the Gaussian noise components of the Rician noise added to wave-controlled aliasing in parallel imaging (Wave-CAIPI) data. $\bar{S}$ denotes the mean image intensity within the brain mask. Metrics including the mean squared error (MSE) and VGG perceptual loss (VGG) are listed to quantify the similarity between images from different methods and the standard FLAIR images of the patient.

noise level ($\sigma = 12\%\ \bar{S}$), AONLM-denoised and BM4D-denoised images exhibited obvious errors in anatomical structures, for example, with obscured contrast between the gray matter and white matter (Figure 6, rows d and e, columns iii and iv). MU-Net still achieved the lowest MSE ($1.324 \times 10^{-3} \pm 0.039 \times 10^{-3}$) but the output images were very blurry (Figure 6, rows d and e, column v). HDnGAN ($\lambda = 10^{-3}$) removed noise and recovered most signals with acceptable visual quality (Figure 6, rows d and e, column vi) and achieved the best VGG perceptual loss ($2.834 \times 10^{-2} \pm 0.425 \times 10^{-2}$).

4 | DISCUSSION

This study combines data acquired with the state-of-the-art fast imaging method Wave-CAIPI and machine learning technology (GAN) to achieve fast volumetric

TABLE 2 Quantitative comparison of results from different denoising methods at higher noise levels

Noise level	Method	Mean squared error ($\times 10^{-3}$)	VGG perceptual loss ($\times 10^{-2}$)
$\sigma = 3\%\bar{S}$	Simulated noisy	2.306 ± 0.124	5.701 ± 0.391
	AONLM	0.940 ± 0.072	2.700 ± 0.399
	BM4D	0.960 ± 0.072	2.779 ± 0.388
	MU-Net	**0.703 ± 0.065**	2.785 ± 0.483
	HDnGAN ($\lambda = 10^{-3}$)	0.917 ± 0.050	**1.336 ± 0.186**
$\sigma = 12\%\ \bar{S}$	Simulated noisy	20.544 ± 0.111	18.523 ± 1.071
	AONLM	2.228 ± 0.144	8.031 ± 0.510
	BM4D	2.126 ± 0.183	6.495 ± 0.913
	MU-Net	**1.324 ± 0.039**	3.637 ± 0.486
	HDnGAN ($\lambda = 10^{-3}$)	1.550 ± 0.043	**2.834 ± 0.425**

Note: The group mean and group standard deviation of image similarity metrics (mean ± SD) including the mean squared error and VGG perceptual loss are listed to quantify the similarity between images from different methods and the standard fluid-attenuated inversion recovery (FLAIR) image at higher noise levels. The noise level is described in terms of the standard deviation (σ) of the Gaussian noise components of the Rician noise added to wave-controlled aliasing in parallel imaging (Wave-CAIPI) data. $\bar{S}$ denotes the mean image intensity within the brain mask. The methods with the best performance for each metrics of each noise level are marked in bold. The metrics were calculated from eight patients for evaluation.

Abbreviations: AONLM, adaptive optimized nonlocal means; BM4D, block matching with 4D filtering; HDnGAN, hybrid denoising generative adversarial network; MU-Net, modified U-Net.

MRI with high-fidelity image quality similar to the standard images acquired with more than doubled scan time. The proposed HDnGAN not only improves SNR but also recovers realistic textures, the richness of which can be controlled by adjusting adversarial loss contributions. Quantitatively, HDnGAN ($\lambda = 0$), that is, MU-Net, achieves the lowest MSE while HDnGAN ($\lambda = 10^{-3}$) achieves the lowest VGG perceptual loss compared to the standard FLAIR images among all denoising methods, including AONLM, BM4D, and the 3D GAN trained on limited training data. Reader assessment demonstrates radiologists' preference for images from HDnGAN ($\lambda = 10^{-3}$) over raw Wave-CAIPI images and denoised images from AONLM, BM4D, MU-Net, and the 3D GAN ($\lambda = 10^{-3}$).

Image denoising provides a complimentary approach for improving the intrinsically low SNR of modern fast MRI techniques. Even though the state-of-the-art fast imaging methods such as the Wave-CAIPI adopted in our study can achieve high acceleration rates with negligible noise amplification and structural artifacts, the SNR of highly accelerated images suffers from the inherent $\sqrt{R}$ penalty, which substantially benefits from image denoising using frameworks such as compressed sensing[9,33] and LORAKS[10] as well as CNNs. In our study, the denoising step by CNNs is cast as a stand-alone post-processing step with several advantages. First, the sole input consists of reconstructed images from the MRI scanner as opposed to raw k-space data, which may be large in size and require further image reconstruction to render the data useable for such pipelines. Therefore, our proposed method can be readily incorporated into existing software packages for visualization and analysis, thereby facilitating the real-time visualization of the resulting images on the clinical console. Second, the denoising process is independent of the imaging process, such that HDnGAN can be applied to noisy images from any MRI contrast, sequence or reconstruction method as well as different imaging modalities (e.g., X-ray computed tomography, positron emission tomography). Third, denoising images also facilitates the acquisition of the training data as reconstructed images from a wider range of MRI scanners or the use of legacy image data as the training data, which is especially critical for GANs that require much more training data than normal CNNs. This also greatly facilitates the use of this method by clinicians and neuroscientists who most often use images that are directly reconstructed from the MRI scanner.

Thus far, only 2D GANs (consisting of a 2D generator and a 2D discriminator) or 3D GANs (consisting of a 3D generator and a 3D discriminator) have been proposed, which have their own benefits and disadvantages. Two-dimensional GANs can be trained on data from a few subjects, since each data volume provides millions of voxels and hundreds of image slices as training samples for the generator and the discriminator, respectively. However, the image synthesis performance of 2D generators is limited compared to 3D generators, which is clearly demonstrated using a 2D MU-Net (trained on sagittal image slices) and a 3D MU-Net in Figure S1. In the images from the 2D MU-Net, the lesion contrast is reduced in the sagittal plane, and some subtle lesions become undetectable (Figure S1 red arrows). Furthermore, the transition between 2D imaging slices is not smooth in the axial (Figure S1 yellow arrows) and coronal views (Figure S1 blue arrows), such that the lesion geometry is obscured. Quantitatively, the 2D MU-Net exhibits higher MSE compared to the 3D MU-Net ($6.64 \times 10^{-4} \pm 0.67 \times 10^{-4}$ vs. $5.92 \times 10^{-4} \pm 0.56 \times 10^{-4}$ for the eight evaluation patients).

On the other hand, the 3D discriminator in a 3D GAN requires substantially more data for training because its loss is computed for a single probability value that classifies the input image volume. The data from a single subject can only provide one image volume or several image blocks (e.g., 18–27 blocks of $64 \times 64 \times 64$ voxels). In theory, 3D GANs are superior, if trained and validated on sufficient data, since the 3D discriminator has more information to distinguish the 3D synthesized and real image blocks. In previous simulation studies that used 1113 subjects or 100 subjects for training and validation, 3D GANs were indeed shown to be effective for image super-resolution and denoising.[34–36] However, our study shows that the 3D GAN cannot compare to the proposed HDnGAN (Figures 4 and 5, Table 1) if trained and validated on the data from only 25 MS patients. Presumably, this is because the 10-layer 3D discriminator (13.1 million parameters) cannot be optimized using only ~650

blocks from 25 MS patients. Acquiring training data on much more subjects, for example, 100+ MS patients, would be challenging in practice. This is an important reason that previous studies have only proven the principle of 3D GANs using simulation data but have never demonstrated their efficacy on empirical data.

The hybrid architecture of HDnGAN with a 3D generator and a 2D discriminator facilitates robust training of HDnGAN on limited empirical data (e.g., 25 patients in our study), which makes it possible to use GANs in practice. The use of 2D discriminator in HDnGAN substantially increases the training samples by 64 × 3 times by classifying all axial, coronal, and sagittal slices (i.e., ~650 blocks × 64 × 3 image slices), which provides a practical solution for concurrent benefits from a 3D generator and a 2D discriminator and also mimics the visual inspection performed by radiologists who read 2D images from different views. The results confirm HDnGAN's superiority to state-of-the-art 3D GAN in objective evaluation including data consistency (i.e., MSE), visual quality (i.e., VGG), and subjective evaluation including sharpness, SNR, lesion conspicuity, and overall quality as it alleviates the overfitting problem for the 3D discriminator when training data are limited.

We also systematically characterized the effects of adversarial loss on resultant image sharpness. Overall, larger adversarial loss weight leads to images with more textural details (Figure 2) with lower VGG perceptual loss but higher MSE (Figure 3). In order to demonstrate this effect, we did not use the VGG perceptual loss in the optimization as in many previous studies,[25,27,35,37] which confounds the origin of the resultant textural details. Clarifying this effect not only facilitates the selection of optimal weight of the adversarial loss (i.e., λ), but also enables an elegant way to control the output image sharpness, which could address the needs of radiologists who may have different preferences for imaging texture. In our study, we selected $\lambda = 10^{-3}$ which generates output images that are most similar to the standard FLAIR images as quantified by the VGG perceptual loss (even though the VGG network was trained on natural images from the ImageNet database) and by visual inspection of radiologists (Figure 2). In practice, λ could be selected or adjusted by radiologists according to their preferences.

Previous studies often perform a linear blending of the GAN generator network parameters[29] or output images of models optimized using only the MSE loss and the adversarial loss[29,36] to achieve different sharpness levels for image super-resolution. However, these two methods are suboptimal as shown in Figures S6 and S7. Visually, the network parameter blending approach fails to denoise Wave-CAIPI images effectively, while the textures from the image blending method are less realistic than those from HDnGAN ($\lambda = 10^{-3}$) (Figure S6). The VGG perceptual losses of both methods are higher than HDnGAN ($\lambda = 10^{-3}$) results (Figure S7), demonstrating the superiority of our method on controlling resultant image sharpness.

The results of the reader study confirm the superiority of HDnGAN as HDnGAN ($\lambda = 10^{-3}$) achieves the highest mean score among all denoising methods provided for comparison. The mean overall score of HDnGAN ($\lambda = 10^{-3}$) is significantly higher than those of Wave-CAIPI, BM4D, MU-Net, and 3D GAN ($\lambda = 10^{-3}$), with no significant difference compared to standard FLAIR images. The mean overall score from HDnGAN ($\lambda = 10^{-3}$) was also higher than that of AONLM, but the difference was not statistically significant. The large standard deviation of overall scores for images from AONLM contributes to the lack of statistical significance, which indicates the two radiologists have varied opinions on the resultant images from AONLM.

HDnGAN and the approach to control the output image sharpness can be simply extended for different noise levels, MRI sequences and contrasts, imaging tasks (e.g., super-resolution), and CNN-based image reconstruction from k-space data, and different biomedical imaging modalities. The acceleration factor of the Wave-CAIPI data used for our experiment was not very high (3 × 2) due to the use of the 20-channel head coil. For 32-channel or 64-channel[51] coils, HDnGAN could be used to denoise Wave-CAIPI images by 10× and achieve even higher acceleration to further reduce the scan time to within a minute. As our simulation experiments suggest, the advantages of HDnGAN are more obvious at a higher noise level (Figure 6, Table 2). The performance could be further improved by optimizing the parameters of Wave-CAIPI[52] and parallel MRI using wave encoding and virtual conjugate coils.[53] Further, HDnGAN can be used for a variety of other Wave-CAIPI-accelerated 3D sequences beyond T_2-weighted FLAIR,[45] including 3D T_1-weighted magnetization prepared rapid gradient echo (MPRAGE),[15,54] susceptibility-weighted imaging,[55] T_1-weighted SPACE,[56,57] and 3D balanced steady-state free precession (bSSFP),[58] as well as images reconstructed using other accelerated methods such as compressed sensing and LORAKS.[10,59] HDnGAN can also benefit other clinical applications of Wave-CAIPI such as whole-brain direct myelin water imaging,[60] 4D lung MRI,[61] and aortic 4D flow MRI.[62] Finally, the generator of HDnGAN can be replaced with any CNN (e.g., variational network[63]) that reconstructs images directly from k-space data in order to recover image sharpness and realistic textural details in the reconstructed images.

5 | CONCLUSION

This study proposes a synergistic Wave-CAIPI and GAN method for high-fidelity fast FLAIR imaging. The proposed hybrid denoising GAN, entitled HDnGAN, improves the SNR of highly accelerated Wave-CAIPI

images while preserving realistic textural detail. HDnGAN benefits from improved image synthesis performance from the 3D generator and increased training samples for training the 2D discriminator on empirical data from 25 MS patients. HDnGAN ($\lambda = 10^{-3}$) generates images most similar to high-quality images acquired in longer scan times, with the lowest VGG perceptual loss and higher preference by neuroradiologists than Wave-CAIPI images and those denoised by AONLM, BM4D, MU-Net, and 3D GAN ($\lambda = 10^{-3}$).

ACKNOWLEDGMENTS

This work was supported by the Tsinghua University Initiative Scientific Research Program, National Institutes of Health (grant numbers: P41-EB030006, K23-NS096056, R01-NS118187, R01-EB020613, R01-EB028797, R03-EB031175, K99-AG073506), a research grant from Siemens Healthineers, a Massachusetts General Hospital Claflin Distinguished Scholar Award, and Athinoula A. Martinos Center for Biomedical Imaging.

CONFLICTS OF INTEREST

Wei-Ching Lo is an employee of Siemens Medical Solutions. Berkin Bilgic has provided consulting services to Subtle Medical.

DATA AVAILABILITY STATEMENT

The image data including 3D T_2-SPACE fluid-attenuated inversion recovery (FLAIR) and wave-controlled aliasing in parallel imaging (Wave-CAIPI) FLAIR are acquired in 33 multiple sclerosis patients undergoing clinical evaluation for demyelinating disease at the Massachusetts General Hospital as part of a separate clinical validation study of Wave-CAIPI FLAIR compared to standard 3D T_2-SPACE FLAIR.[45] Due to privacy and ethical concerns, data cannot be made available.

CODE AVAILABILITY STATEMENT

The source codes of HDnGAN implemented using Keras application programming interface are available on GitHub at https://github.com/liziyu0929/HDnGAN.

REFERENCES

1. Fischl B. FreeSurfer. *Neuroimage*. 2012;62(2):774-781.
2. Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*. 1999;9(2):195-207.
3. Polimeni JR, Fischl B, Greve DN, Wald LL. Laminar analysis of 7T BOLD using an imposed spatial activation pattern in human V1. *Neuroimage*. 2010;52(4):1334-1346.
4. McNab JA, Polimeni JR, Wang R, et al. Surface based analysis of diffusion orientation for identifying architectonic domains in the in vivo human cortex. *Neuroimage*. 2013;69:87-100.
5. De Coene B, Hajnal JV, Gatehouse P, et al. MR of the brain using fluid-attenuated inversion recovery (FLAIR) pulse sequences. *Am J Neuroradiol*. 1992;13(6):1555-1564.
6. Lüsebrink F, Sciarra A, Mattern H, Yakupov R, Speck O. T1-weighted in vivo human whole brain MRI dataset with an ultrahigh isotropic resolution of 250 μm. *Sci Data*. 2017;4(1):12.
7. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med*. 1999;42(5):952-962.
8. Griswold MA, Jakob PM, Heidemann RM, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med*. 2002;47(6):1202-1210.
9. Lustig M, Donoho DL, Santos JM, Pauly JM. Compressed sensing MRI. *IEEE Signal Process Mag*. 2008;25(2):72-82.
10. Haldar JP. Low-rank modeling of local k-space neighborhoods (LORAKS) for constrained MRI. *IEEE Trans Med Imaging*. 2013;33(3):668-681.
11. Moriguchi H, Duerk JL. Bunched phase encoding (BPE): a new fast data acquisition method in MRI. *Magn Reson Med*. 2006;55(3):633-648.
12. Breuer FA, Blaimer M, Mueller MF, et al. Controlled aliasing in volumetric parallel imaging (2D CAIPIRINHA). *Magn Reson Med*. 2006;55(3):549-556.
13. Bilgic B, Gagoski BA, Cauley SF, et al. Wave-CAIPI for highly accelerated 3D imaging. *Magn Reson Med*. 2015;73(6):2152-2162.
14. Polak D, Cauley S, Huang SY, et al. Highly-accelerated volumetric brain examination using optimized wave-CAIPI encoding. *J Magn Reson Imaging*. 2019;50(3):961-974.
15. Polak D, Setsompop K, Cauley SF, et al. Wave-CAIPI for highly accelerated MP-RAGE imaging. *Magn Reson Med*. 2018;79(1):401-406.
16. Kim J, Lee JK, Lee KM. Accurate image super-resolution using very deep convolutional networks. In: *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2016:1646–1654.
17. Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans Image Process*. 2017;26(7):3142-3155.
18. Hu Y, Xu Y, Tian Q, et al. RUN-UP: accelerated multishot diffusion-weighted MRI reconstruction using an unrolled network with U-Net as priors. *Magn Reson Med*. 2020;85(2):709-720.
19. Bilgic B, Chatnuntawech I, Manhard MK, et al. Highly accelerated multishot echo planar imaging through synergistic machine learning and joint reconstruction. *Magn Reson Med*. 2019;82(4):1343-1358.
20. Chaudhari AS, Fang Z, Kogan F, et al. Super-resolution musculoskeletal MRI using deep learning. *Magn Reson Med*. 2018;80(5):2139-2154.
21. Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans Image Process*. 2007;16(8):2080-2095.
22. Tian Q, Zaretskaya N, Fan Q, et al. Improved cortical surface reconstruction using sub-millimeter resolution MPRAGE by image denoising. *Neuroimage*. 2021;233:117946.
23. Maggioni M, Katkovnik V, Egiazarian K, Foi A. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Trans Image Process*. 2012;22(1):119-133.
24. Manjón JV, Coupé P, Martí-Bonmatí L, Collins DL, Robles M. Adaptive non-local means denoising of MR images with spatially varying noise levels. *J Magn Reson Imaging*. 2010;31(1):192-203.
25. Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network. In: *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2017:4681-4690.
26. Yi X, Babyn P. Sharpness-aware low-dose CT denoising using conditional generative adversarial network. *J Digit Imaging*. 2018;31(5):655-669.
27. Yang Q, Yan P, Zhang Y, et al. Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Trans Med Imaging*. 2018;37(6):1348-1357.
28. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *Adv Neural Inform Process Syst*. 2014;27:2672-2680.

29. Wang X, Yu K, Wu S, et al. ESRGAN: enhanced super-resolution generative adversarial networks. In: *European Conference on Computer Vision (ECCV) Workshops*. Springer; 2018.
30. Chen Z, Zeng Z, Shen H, Zheng X, Dai P, Ouyang P. DN-GAN: denoising generative adversarial networks for speckle noise reduction in optical coherence tomography images. *Biomed Signal Process Control*. 2020;55:101632.
31. Wang H, Rivenson Y, Jin Y, et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat Methods*. 2019;16:103-110.
32. Wolterink JM, Leiner T, Viergever MA, Išgum I. Generative adversarial networks for noise reduction in low-dose CT. *IEEE Trans Med Imaging*. 2017;36(12):2536-2545.
33. Mardani M, Gong E, Cheng JY, et al. Deep Generative Adversarial Neural Networks for Compressive Sensing MRI. *IEEE Trans Med Imaging*. 2019;38:(1):167–179.
34. Chen Y, Shi F, Christodoulou AG, Xie Y, Zhou Z, Li D. Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-Level densely connected network. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2018:91-99.
35. Ran M, Hu J, Chen Y, et al. Denoising of 3D magnetic resonance images using a residual encoder–decoder Wasserstein generative adversarial network. *Med Image Anal*. 2019;55:165-180.
36. Wang J, Chen Y, Wu Y, Shi J, Gee J. Enhanced generative adversarial network for 3D brain MRI super-resolution. *IEEE Winter Conference on Applications of Computer Vision*. In: IEEE; 2020:3627-3636.
37. Yang G, Yu S, Dong H, et al. Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Trans Med Imaging*. 2018;37(6):1310-1321.
38. Tian Q, Bilgic B, Fan Q, et al. DeepDTI: high-fidelity six-direction diffusion tensor imaging using deep learning. *Neuroimage*. 2020;219(1):117017.
39. Tian Q, Bilgic B, Fan Q, et al. Improving in vivo human cerebral cortical surface reconstruction using data-driven super-resolution. *Cereb Cortex*. 2020;31(1):463-482.
40. Chen Y, Xie Y, Zhou Z, Shi F, Christodoulou AG, Li D. Brain MRI super resolution using 3D deep densely connected neural networks. In: *IEEE International Symposium on Biomedical Imaging*. IEEE; 2018:739-742.
41. Falk T, Mai D, Bensch R, et al. U-Net: deep learning for cell counting, detection, and morphometry. *Nat Methods*. 2019;16(1):67-70.
42. Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks. In: *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2016:1646-1654.
43. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2017:4700-4708.
44. Miyato T, Kataoka T, Koyama M, Yoshida Y. Spectral normalization for generative adversarial networks. *arXiv preprint*. 2018. arXiv:1802.05957.
45. Ngamsombat C, Goncalves Filho ALM, Longo MG, et al. Evaluation of ultrafast wave-CAIPI 3D FLAIR in the visualization and volumetric estimation of cerebral white matter lesions. *Am J Neuroradiol*. 2021. doi:10.3174/ajnr.A7191
46. Modat M, Ridgway GR, Taylor ZA, et al. Fast free-form deformation using graphics processing units. *Comput Methods Programs Biomed*. 2010;98(3):278-284.
47. Modat M, Cash DM, Daga P, Winston GP, Duncan JS, Ourselin S. Global image registration using a symmetric block-matching approach. *J Med Imaging*. 2014;1(2):024003.
48. Uwano I, Kudo K, Yamashita F, et al. Intensity inhomogeneity correction for magnetic resonance imaging of human brain at 7T. *Med Phys*. 2014;41(2):022302.
49. Coupé P, Yger P, Prima S, Hellier P, Kervrann C, Barillot C. An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. *IEEE Trans Med Imaging*. 2008;27(4):425-441.
50. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint*. 2014. arXiv:1409.1556.
51. Keil B, Blau JN, Biber S, et al. A 64-channel 3T array coil for accelerated brain MRI. *Magn Reson Med*. 2013;70(1):248-258.
52. Wang H, Qiu Z, Su S, et al. Parameter optimization framework on wave gradients of Wave-CAIPI imaging. *Magn Reson Med*. 2020;83(5):1659-1672.
53. Qiu Z, Jia S, Su S, et al. Highly accelerated parallel MRI using wave encoding and virtual conjugate coils. *Magn Reson Med*. 2021;86(3):1345-1359.
54. Longo M, Conklin J, Cauley S, et al. Evaluation of ultrafast wave-CAIPI MPRAGE for visual grading and automated measurement of brain tissue volume. *Am J Neuroradiol*. 2020;41(8):1388-1396.
55. Conklin J, Longo MGF, Cauley SF, et al. Validation of highly-accelerated wave-CAIPI susceptibility-weighted imaging (SWI) compared to conventional SWI and T2*-weighted gradient-echo for routine clinical brain MRI at 3T. *AJNR Am J Neuroradiol*. 2019;40(12):2073.
56. Goncalves Filho ALM, Conklin J, Longo MGF, et al. Accelerated post-contrast wave-CAIPI T1 SPACE achieves equivalent diagnostic performance compared with standard T1 SPACE for the detection of brain metastases in clinical 3T MRI. *Front Neurol*. 2020;11:587327.
57. Goncalves Filho ALM, Longo MGF, Conklin J, et al. MRI highly accelerated wave-CAIPI T1-SPACE versus standard T1-SPACE to detect brain gadolinium-enhancing lesions at 3T. *J Neuroimaging*. 2021;31(5):893-901. doi:https://doi.org/10.1111/jon.12893
58. Su S, Qiu Z, Luo C, et al. Accelerated 3D bSSFP using a modified wave-CAIPI technique with truncated wave gradients. *IEEE Trans Med Imaging*. 2020;40(1):48-58.
59. Lustig M, Donoho D, Pauly JM. Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn Reson Med*. 2007;58(6):1182-1195.
60. Wu Z, Bilgic B, He H, et al. Wave-CAIPI ViSTa: highly accelerated whole-brain direct myelin water imaging with zero-padding reconstruction. *Magn Reson Med*. 2018;80(3):1061-1073.
61. Richter JA, Wech T, Weng AM, et al. Free-breathing self-gated 4D lung MRI using wave-CAIPI. *Magn Reson Med*. 2020;84(6):3223-3233.
62. Richter JA, Wech T, Weng AM, et al. Accelerated aortic 4D flow MRI with wave-CAIPI. *Magn Reson Med*. 2021;85(5):2595–2607.
63. Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med*. 2018;79(6):3055-3071.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Li Z, Tian Q, Ngamsombat C, et al. High-fidelity fast volumetric brain MRI using synergistic wave-controlled aliasing in parallel imaging and a hybrid denoising generative adversarial network (HDnGAN). *Med. Phys.* 2022;49:1000–1014. https://doi.org/10.1002/mp.15427

WILEY

First Clinical Use of SunSCAN 3D Webinar

Initial Impressions and Workflow Improvements

December 13, 2022 - 12:00 PM EST

AMERICAN ASSOCIATION of PHYSICISTS IN MEDICINE

Sponsored by SUN NUCLEAR