High-fidelity fast volumetric brain MRI using 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[8], Kawin Setsompop[2, 3, 7], Qiuyun Fan[2, 3], Berkin Bilgic[2, 3, 7], Stephen Cauley[2, 3], Susie Y. Huang[2, 3, 7]

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

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

[3]Harvard Medical School, Boston, MA, USA;

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

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

[6]Siemens Medical Solutions, Boston, MA, USA;

[7]Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA;

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

[*]Correspondence to: Qiyuan Tian, Ph.D., Athinoula A. Martinos Center for Biomedical Imaging, 149 13th Street, Charlestown, MA, 02129, United States. E-mail: qtian@mgh.harvard.edu.

Running head: High-fidelity fast volumetric brain MRI using HDnGAN

Word count: 2800 (body), 250 (abstract)

Figure and table count: 5

Funding Information: National Institutes of Health (NIH), Grant numbers: P41-EB030006, K23-NS096056, R01-EB020613, R01-EB028797; Siemens Healthineers (research support); Athinoula A. Martinos Center for Biomedical Imaging; Massachusetts General Hospital Claflin Distinguished Scholar Award; Tsinghua University Initiative Scientific Research Program.

Submitted to Magnetic Resonance in Medicine.

Abstract

Purpose

To improve the signal-to-noise ratio (SNR) of highly accelerated volumetric MRI while preserve realistic textures using a generative adversarial network (GAN).

Methods

A hybrid GAN for denoising entitled "HDnGAN" with a 3D generator and a 2D discriminator was proposed to denoise 3D T_2-weighted fluid-attenuated inversion recovery (FLAIR) images acquired in 2.75 minutes (R=3×2) using wave-controlled aliasing in parallel imaging (Wave-CAIPI). HDnGAN was trained on data from 25 multiple sclerosis patients by minimizing a combined mean squared error and adversarial loss with adjustable weight λ. Results were evaluated on eight separate patients by comparing to standard T2-SPACE FLAIR images acquired in 7.25 minutes (R=2×2) using mean absolute error (MAE), peak SNR (PSNR), structural similarity index (SSIM), and VGG perceptual loss, and by two neuroradiologists using a five-point score regarding gray-white matter contrast, sharpness, SNR, lesion conspicuity, and overall quality.

Results

HDnGAN (λ=0) produced the lowest MAE, highest PSNR and SSIM. HDnGAN ($\lambda=10^{-3}$) produced the lowest VGG loss. In the reader study, HDnGAN ($\lambda=10^{-3}$) significantly improved the gray-white contrast and SNR of Wave-CAIPI images, and outperformed BM4D and HDnGAN (λ=0) regarding image sharpness. The overall quality score from HDnGAN ($\lambda=10^{-3}$) was significantly higher than those from Wave-CAIPI, BM4D, and HDnGAN (λ=0), with no significant difference compared to standard images.

Conclusion

HDnGAN concurrently benefits from improved image synthesis performance of 3D convolution and increased training samples for training the 2D discriminator on limited data. HDnGAN generates images with high SNR and realistic textures, similar to those acquired in longer times and preferred by neuroradiologists.

Keywords: denoising, Wave-CAIPI, T_2-weighted FLAIR, multiple sclerosis, generative adversarial network, adversarial loss

Introduction

High-resolution volumetric brain MRI is widely used in clinical and research applications to provide rich and detailed anatomical information and delineation of structural pathology. For example, 3D 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[1]. A main barrier to greater adoption of volumetric MRI in clinical protocols is the long acquisition time (typically ~5-7 minutes), which may lead to patient anxiety and motion artifacts that compromise diagnostic quality, especially for children, elderly subjects, and some patient populations who cannot tolerate long scans.

Modern fast imaging techniques[2-5] can accelerate volumetric brain MRI with an acceleration factor much higher than that can be provided by conventional parallel imaging methods (e.g., 3 for 2D and 2×2 for 3D in SENSE[6] and GRAPPA[7]). For example, wave-controlled aliasing in parallel imaging (Wave-CAIPI)[5] employs a corkscrew gradient trajectory with CAIPI shifts in the k_y and k_z directions to efficiently encode k-space and uniformly spread the voxel aliasing and can achieve 10× acceleration with negligible g-factor and image artifact penalties. However, the signal-to-noise ratio (SNR) ($\propto \sqrt{1/\text{acceleration factor}}$) of highly accelerated images is intrinsically lower compared to those acquired with mild acceleration factors due to substantially less acquired k-space signals.

Convolutional neural networks (CNNs) have superior performance in image restoration tasks such as super-resolution[8,9] and denoising[10] and provide a promising strategy to improve the SNR of highly accelerated images. However, CNNs using the voxel-wise mean squared error (MSE) as the loss function tend to generate blurry images lacking textural details in both super-resolution and denoising tasks even though MSE or related metrics such as peak SNR (PSNR) can be minimized or maximized. Generative adversarial networks (GANs)[11] have been demonstrated effective in reducing the blur effects and recovering realistic textures for digital photography[12,13], optical coherence tomography[14], microscopy[15], x-ray computed tomography[16,17] and MRI[18-21], by pushing the restored images to the manifold of target images using a discriminator network that is trained to differentiate between the input and target images.

For volumetric imaging, 3D convolution in the generator network is advantageous for increasing the data redundancy from an additional spatial dimension for improved image synthesis performance and smooth transition between 2D image slices along all directions[22-24], but 3D discriminators in GANs require a large number of training data, which are challenging to acquire in practice. Chen et al. utilized data from 1113 subjects to train a 3D GAN for brain MRI super-resolution[19]. Ran et al. split the whole-brain volumes into small blocks (i.e., 32×32×6 voxels) to increase the sample number, which allowed them to train a 3D GAN for MRI denoising with data from 110 subjects[20]. Because of the intensive data requirement, the degraded images were simulated by adding noise to high-SNR images or down-sampling high-resolution images from public database[19,20] or only a shallow discriminator was used[20] in these studies.

To address this challenge, we propose a hybrid denoising GAN (entitled HDnGAN) to improve the SNR of highly accelerated empirical images. HDnGAN adopts a 20-layer 3D generator and a 10-layer 2D discriminator to benefit concurrently from improved image synthesis performance provided by the 3D convolution in the generator and increased training samples from a limited number of subjects for the discriminator. In distinction to existing studies using simulated data[17,19-21], the efficacy of HDnGAN is demonstrated on empirical Wave-CAIPI and standard T2-SPACE FLAIR data acquired on 33 multiple sclerosis (MS) patients. The results are comprehensively and quantitatively assessed using different metrics and by neuroradiologists. The effects of adversarial loss's contribution to the resultant image sharpness are also systematically characterized.

Methods

Data acquisition

This study was approved by the Institutional review board and was HIPAA compliant. With written informed consent, data were acquired in 33 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 T2-SPACE FLAIR[25]. All patients were scanned on a whole-body 3-Tesla MAGNETOM Prisma MRI scanner (Siemens Healthcare) equipped with a 20-channel head coil.

Standard FLAIR data were acquired using a 3D T2-SPACE FLAIR sequence with: 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, acquisition time=7.25 minutes. Wave-CAIPI FLAIR data were acquired using a prototype sequence with matched parameter values except for: echo time=392 ms, bandwidth=750 Hz/pixel, acceleration factor=3×2, acquisition time=2.75 minutes.

Image processing

Standard FLAIR images were non-linearly co-registered to Wave-CAIPI images using the "*reg_f3d*" function from the NiftyReg software[26,27], which was initialized with an affine transformation derived from NiftyReg's "*reg_aladin*" function. The non-linear registration was used to account for the subtle non-linear shifts of tissues in the images. Brain masks were created from Wave-CAIPI images using the unified segmentation algorithm in the Statistical Parametric Mapping software[28].

For each subject, a binary mask was created to exclude parts of the frontal lobe, temporal lobe, cerebellum, and brainstem where the residuals between Wave-CAIPI and standard FLAIR images were dominated by large image artifacts and geometric distortions rather than noise. Specifically, 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. Only the MSE loss within this mask was used to optimize the generator (i.e., referred to herein as "loss mask").

Generative adversarial network

A GAN was used to perform image quality transfer from Wave-CAIPI to standard FLAIR images. The GAN consisted of a generator and a discriminator. The generator synthesized images similar to standard FLAIR images from Wave-CAIPI images, while the discriminator tried to distinguish synthesized images from the acquired standard FLAIR images. The generator and the discriminator were trained in synchrony to compete against each other.

Generator

A modified 3D U-Net[29] was used as the generator (Fig. 1a) (2.3 million parameters), which predicted the residual between the Wave-CAIPI and standard image volume. Specifically, all max pooling, up-sampling, and batch normalization layers were removed, and the number of kernels at each layer was kept constant (n=64). This modified U-Net created 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. It represented an intermediate network between a plain network (e.g., VDSR[9] and DnCNN[10]) without any short paths and a densely connected network (e.g., DenseNet[30]) that comprehensively connects each layer to every other layer.

Discriminator

The discriminator of SRGAN[12] was adopted (13.1 million parameters) (Fig. 1b), with spectral normalization incorporated in each layer to stabilize training[31]. The discriminator was designed to classify 2D image slices rather than 3D blocks to increase the number of training samples to ensure that the GAN could be optimized on data from a limited number of subjects. Specifically, the discriminator used all axial, coronal, and sagittal image slices of the resultant 3D image blocks from the generator as separate samples during the training.

Loss function

The loss function $\mathcal{L}$ consisted of a content loss $\mathcal{L}_{content}$ from the generator and an adversarial loss $\mathcal{L}_{adversarial}$ from the discriminator as:

$$\mathcal{L} = \mathcal{L}_{content} + \lambda \mathcal{L}_{adversarial}, \quad (1)$$

where λ determines the contribution of $\mathcal{L}_{adversarial}$ to $\mathcal{L}$. $\mathcal{L} = \mathcal{L}_{content}$ when λ=0 and $\mathcal{L} = \mathcal{L}_{adversarial}$ when λ=∞. Intermediate λ values achieve mixed content loss and adversarial loss.

The content loss was defined as the voxel-wise MSE, calculated as:

$$\mathcal{L}_{content} = \frac{1}{N_b}\frac{1}{N_v}\sum_{n=1}^{N_b}\sum_{i,j,k\in\mathcal{M}}\left[I_{i,j,k}^{std} - G_{\theta_G}^{3D}(I^{wave})_{i,j,k}\right]^2, \tag{2}$$

where I^{std} and I^{wave} denote standard and Wave-CAIPI image volumes; $G_{\theta_G}^{3D}$ is the 3D generator parametrized by θ_G; i, j, k are coordinates within the loss mask $\mathcal{M}$ with minimum image artifacts and distortions; N_v is the number of voxels within the loss mask; N_b is the number of blocks for training.

The adversarial loss was defined as:

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

where $D_{\theta_D}^{2D}$ is the 2D discriminator parametrized by θ_D; $S_m(:)$ denotes the operation of selecting the mth slice out of all N_{acs} axial, coronal and sagittal slices from the synthesized image block $G_{\theta_G}^{3D}(I^{wave})$; $D_{\theta_D}^{2D}\{S_m[G_{\theta_G}^{3D}(I^{wave})]\}$ is the probability that an image slice is classified as a real standard FLAIR image. For better gradient behavior, $-\log D_{\theta_D}^{2D}\{S_m[G_{\theta_G}^{3D}(I^{wave})]\}$ was minimized rather than $\log\left\{1 - D_{\theta_D}^{2D}\{S_m[G_{\theta_G}^{3D}(I^{wave})]\}\right\}$[12].

GAN training

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 an additional 5 patients using an NVIDIA GeForce RTX 2080 Ti GPU.

Because of the limited GPU memory, the generator was trained on input and output image blocks consisting of 64×64×64 voxels (18-27 blocks from each patient), 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, Wave-CAIPI and standard FLAIR images were skull-stripped and standardized by subtracting the mean intensity and then dividing by the standard deviation of the image intensities of the brain voxels from Wave-CAIPI data. All data were flipped along the anatomical left-right direction to augment the training data.

Network parameters were optimized by minimizing the loss function defined in Equations 1-3 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. HDnGAN with eight different λ values (listed in Fig. 2) were trained and validated, each for 21 epochs and ~24 hours.

Image quality evaluation

Data from 8 separate patients were used for evaluation. For comparison, Wave-CAIPI images were also denoised by the state-of-the-art block-matching and 4D filtering (BM4D)[32,33] method (default parameters, https://www.cs.tut.fi/~foi/GCF-BM3D). The mean absolute error (MAE), PSNR, structural similarity index (SSIM)[34], and VGG perceptual loss[35] were used to quantify the similarity between Wave-CAIPI, BM4D-denoised, and HDnGAN-denoised images and standard FLAIR images.

Images were also evaluated by two neuroradiologists (C.N., S.C.) across five image quality metrics, including gray-white matter contrast, sharpness, SNR, lesion conspicuity, and overall quality. For HDnGAN, results generated using $\lambda=10^{-3}$ were used for comparison. Images were scored using a five-point scale: 1 nondiagnostic, 2 limited, 3 diagnostic, 4 good, 5 excellent. The group mean and 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. Bonferroni correction was performed to account for multiple comparisons, which resulted in an adjusted P threshold of 0.005 (10 comparisons in total).

Results

Different losses computed on the validation set smoothly decreased or increased and approached convergency at 21 epochs (Supplementary Information Fig. 1). Visually, the contribution of adversarial loss controlled the sharpness of resultant images (Fig. 2). For BM4D and HDnGAN with low λ values (i.e., 0, 10^{-5}, 10^{-4}) when MSE loss dominated, resultant images were smooth without high-frequency textures, as expected. When λ gradually increased, more textural details became apparent. Images from HDnGAN around $\lambda=10^{-3}$ were visually very similar to the standard FLAIR images.

The optimal λ varied for different image similarity metrics (Fig. 3). For MAE, PSNR, and SSIM, HDnGAN ($\lambda=0$) achieved the best performance (0.017±0.001, 32.512±0.323 dB, 0.935±0.008), substantially better than Wave-CAIPI input (0.025±0.001, 29.798±0.439 dB, 0.876±0.017) and slightly better than BM4D results (0.019±0.001, 31.656±0.363 dB, 0.922±0.011). When λ gradually increased, MAE, PSNR, and SSIM increased (or decreased) first, reached the maximum (or minimum) at $\lambda=10^{-1}$ and then became better. HDnGAN ($\lambda=10^{-3}$) achieved the best performance in terms of VGG loss (0.011±0.002), improving significantly upon Wave-CAIPI inputs (0.020±0.003) and BM4D results (0.017±0.002). When λ gradually increased, VGG perceptual loss first decreased, reached the minimum at $\lambda=10^{-3}$, then increased until $\lambda=10^{-1}$, and finally decreased until $\lambda=\infty$.

Figure 4 demonstrated that HDnGAN ($\lambda=10^{-3}$) generated high-quality image volumes with increased SNR, smooth transitions between imaging slices along all axial, sagittal, and coronal directions, and rich and realistic textural details. The difference and SSIM maps (Supplementary Information Fig. 2) showed that the residuals between the HDnGAN-denoised images and ground-truth high-SNR images did not contain anatomical structures.

The results of the reader study (Fig. 5, Supplementary Information Table 1) demonstrated that HDnGAN ($\lambda=10^{-3}$) significantly improved the input Wave-CAIPI images' gray-white contrast (4.188±0.544 vs. 3.438±0.512, $P=0.002$) and SNR (3.750±0.447 vs. 3.063±0.250, $P<0.001$). In terms of image sharpness, it outperformed BM4D (4.189±0.403 vs. 3.000±0.365, $P<0.001$) and HDnGAN ($\lambda=0$) (4.189±0.403 vs. 3.125±0.619, $P<0.001$). The overall scores from HDnGAN ($\lambda=10^{-3}$) were significantly higher than those from Wave-CAIPI (4.125±0.342 vs. 3.563±0.512, $P<0.001$), BM4D

(4.125±0.342 vs. 3.500±0.516, $P<0.001$), and HDnGAN ($\lambda=0$) (4.125±0.342 vs. 3.563±0.727, $p=0.003$), with no significant difference compared to the standard images (4.125±0.342 vs. 4.313±0.479, $p=0.083$).

Discussion

This study leverages Wave-CAIPI and GAN to achieve fast volumetric MRI with high-fidelity image quality similar to the standard images acquired in longer 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. Reader assessment demonstrates clinicians' preference for images from HDnGAN ($\lambda=10^{-3}$) over those from Wave-CAIPI, BM4D, and HDnGAN ($\lambda=0$).

The hybrid architecture of HDnGAN with a 3D generator and a 2D discriminator facilitates HDnGAN to be well-trained on limited empirical data (e.g., 25 patients here). The 3D discriminator requires substantially more data for training because its loss is computed for a single probability value that classifies the input image volume, while the loss is computed for each voxel for the generator. A deep 3D discriminator with millions of parameters (13.1 million for our 2D discriminator) cannot be optimized using only ~650 blocks from 25 MS patients. This intensive data requirement poses a significant barrier to training a 3D GAN on empirical data, which is challenging to collect. Therefore, in previous studies, 3D GANs were only trained on simulated noisy data obtained by adding synthetic noise to high-SNR data from large-scale public databases. The generalization of these GANs to empirical data with different contrasts, noise levels and distributions is unclear. The use of 2D discriminator in HDnGAN substantially increases the training samples by 64×3 times by classifying all axial, coronal, and sagittal slices, which provides a practical solution for concurrent benefits from a 3D generator and a 2D discriminator and also mimics the visual inspection by radiologists who read 2D images from different views.

We also systematically characterize the effects of adversarial loss on resultant image sharpness. Overall, larger adversarial loss weight leads to images with more textural details (Fig. 2) but higher MAE, PSNR, and SSIM (Fig. 3). In order to demonstrate this effect, we did not use the VGG perceptual loss in the optimization as in many previous studies, which confounds the origin of the resultant textural details. This effect enables an elegant way to control the output image sharpness, which could address the needs of radiologists who may have different preferences for the richness of textures. In our study, we select $\lambda=10^{-3}$ which generates output images that are most similar to the standard FLAIR images as quantified by the VGG perceptual loss and by visual inspection (Fig. 2) (even though VGG was trained on natural

images from ImageNet database). In practice, λ could be selected or adjusted by radiologists accordingly to their preferences.

Previous studies often perform a linear blending of the GAN generator network parameters[13] or output images of models optimized using only the MSE loss and the adversarial loss[13,21] to achieve different sharpness levels for image super-resolution. However, these two methods are sub-optimal (Supplementary Information Fig. 3-5). Visually, the network parameter blending fails to denoise Wave-CAIPI images effectively, while the textures from the image blending method are less realistic than those from HDnGAN ($\lambda=10^{-3}$). The VGG perceptual losses of both methods are higher than HDnGAN ($\lambda=10^{-3}$) results, demonstrating the superiority of our method on controlling resultant image sharpness.

HDnGAN can be simply extended for different noise levels, MRI sequences and contrasts, imaging tasks (e.g., super-resolution), and image reconstruction using a generator that reconstructs images directly from k-space data. The acceleration factor of the Wave-CAIPI data used for our experiment was chosen not very high (6×) due to the use of 20-channel head coil. For 32-channel coils, HDnGAN could be employed to denoise Wave-CAIPI images with 10× and even higher acceleration to further reduce the scan time potential to within a minute. Further, HDnGAN could be used for Wave-CAIPI magnetization prepared rapid gradient echo (MPRAGE)[36,37] and susceptibility-weighted imaging[38] or images reconstructed using other methods such as compressed sensing and LORAKS[3,4].

Conclusion

This study proposes a hybrid GAN for denoising entitled HDnGAN for improving the SNR of highly accelerated images while preserving realistic textural details. 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 time, with the lowest VGG perceptual loss and higher preference than Wave-CAIPI images, BM4D-denoised results and HDnGAN ($\lambda=0$) by neuroradiologists.

Conflict of Interest

W.L. is an employee of Siemens Medical Solutions. B.B. has provided consulting services to Subtle Medical.

References

1. De Coene B, Hajnal JV, Gatehouse P, et al. MR of the brain using fluid-attenuated inversion recovery (FLAIR) pulse sequences. *American Journal of Neuroradiology.* 1992;13(6):1555-1564.
2. Breuer FA, Blaimer M, Mueller MF, et al. Controlled aliasing in volumetric parallel imaging (2D CAIPIRINHA). *Magnetic Resonance in Medicine.* 2006;55(3):549-556.
3. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine.* 2007;58(6):1182-1195.
4. Haldar JP. Low-rank modeling of local k-space neighborhoods (LORAKS) for constrained MRI. *IEEE transactions on medical imaging.* 2013;33(3):668-681.
5. Bilgic B, Gagoski BA, Cauley SF, et al. Wave-CAIPI for highly accelerated 3D imaging. *Magnetic Resonance in Medicine.* 2015;73(6):2152-2162.
6. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine.* 1999;42(5):952-962.
7. Griswold MA, Jakob PM, Heidemann RM, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine.* 2002;47(6):1202-1210.
8. Dong C, Loy CC, He K, Tang X. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2016;38(2):295-307.
9. Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks. *IEEE Conference on Computer Vision and Pattern Recognition.* 2016:1646-1654.
10. Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing.* 2017;26(7):3142-3155.
11. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *Advances in Neural Information Processing Systems.* 2014;27:2672-2680.
12. Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network. *IEEE Conference on Computer Vision and Pattern Recognition.* 2017:4681-4690.
13. Wang X, Yu K, Wu S, et al. Esrgan: Enhanced super-resolution generative adversarial networks. *European Conference on Computer Vision (ECCV) Workshops.* 2018.
14. 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. *Biomedical Signal Processing and Control.* 2020;55:101632.
15. Wang H, Rivenson Y, Jin Y, et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nature Methods.* 2019;16:103-110.
16. Wolterink JM, Leiner T, Viergever MA, Išgum I. Generative adversarial networks for noise reduction in low-dose CT. *IEEE transactions on medical imaging.* 2017;36(12):2536-2545.
17. 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 transactions on medical imaging.* 2018;37(6):1348-1357.
18. Mardani M, Gong E, Cheng JY, et al. Deep generative adversarial neural networks for compressive sensing MRI. *IEEE transactions on medical imaging.* 2018;38(1):167-179.

19. 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. *International Conference on Medical Image Computing and Computer-Assisted Intervention.* 2018:91-99.
20. Ran M, Hu J, Chen Y, et al. Denoising of 3D magnetic resonance images using a residual encoder–decoder Wasserstein generative adversarial network. *Medical image analysis.* 2019;55:165-180.
21. 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.* 2020:3627-3636.
22. Chaudhari AS, Fang Z, Kogan F, et al. Super-resolution musculoskeletal MRI using deep learning. *Magnetic Resonance in Medicine.* 2018;80(5):2139-2154.
23. Tian Q, Bilgic B, Fan Q, et al. Improving in vivo human cerebral cortical surface reconstruction using data-driven super-resolution. *Cerebral Cortex.* 2020;31(1):463-482.
24. Tian Q, Bilgic B, Fan Q, et al. DeepDTI: High-fidelity six-direction diffusion tensor imaging using deep learning. *NeuroImage.* 2020:117017.
25. Ngamsombat C, Filho ALMG, Longo MGF, et al. Evaluation of Ultrafast Wave-CAIPI 3D FLAIR in the Visualization and Volumetric Estimation of Cerebral White Matter Lesions. *medRxiv.* 2021.
26. Modat M, Ridgway GR, Taylor ZA, et al. Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine.* 2010;98(3):278-284.
27. Modat M, Cash DM, Daga P, Winston GP, Duncan JS, Ourselin S. Global image registration using a symmetric block-matching approach. *Journal of Medical Imaging.* 2014;1(2):024003.
28. Uwano I, Kudo K, Yamashita F, et al. Intensity inhomogeneity correction for magnetic resonance imaging of human brain at 7T. *Medical Physics.* 2014;41(2).
29. Falk T, Mai D, Bensch R, et al. U-Net: deep learning for cell counting, detection, and morphometry. *Nature Methods.* 2019;16(1):67-70.
30. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. *IEEE conference on computer vision and pattern recognition.* 2017:4700-4708.
31. Miyato T, Kataoka T, Koyama M, Yoshida Y. Spectral normalization for generative adversarial networks. *arXiv preprint.* 2018;arXiv:1802.05957.
32. Maggioni M, Katkovnik V, Egiazarian K, Foi A. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE transactions on image processing.* 2012;22(1):119-133.
33. Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on image processing.* 2007;16(8):2080-2095.
34. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing.* 2004;13(4):600-612.
35. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint.* 2014;arXiv:1409.1556.
36. Polak D, Setsompop K, Cauley SF, et al. Wave-CAIPI for highly accelerated MP-RAGE imaging. *Magnetic resonance in medicine.* 2018;79(1):401-406.
37. Longo M, Conklin J, Cauley S, et al. Evaluation of ultrafast wave-CAIPI MPRAGE for visual grading and automated measurement of brain tissue volume. *American Journal of Neuroradiology.* 2020;41(8):1388-1396.
38. 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 American journal of neuroradiology.* 2019;40(12):2073.

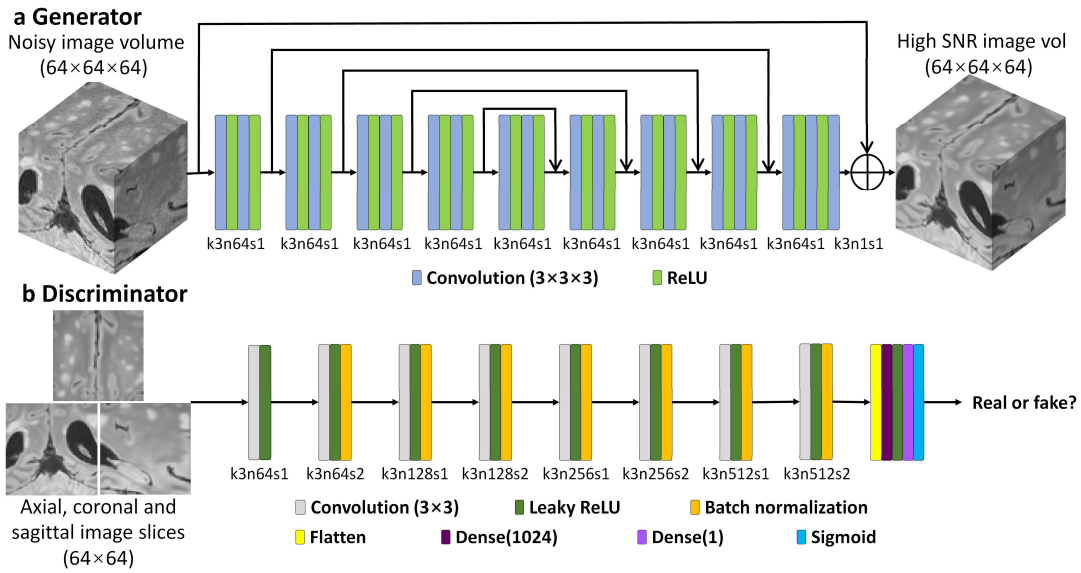


Figure 1. HDnGAN architecture. HDnGAN consists of a 3D generator (a) and a 2D discriminator (b). The 3D generator is modified from U-Net by removing all max-pooling, up-sampling, and batch normalization layers and keeping the number of kernels constant (n=64) across all layers. The input of the generator is a noisy image volume (64×64×64 voxels). The output of the generator is an image volume with high signal-to-noise ratio (SNR). The 2D discriminator adopts the discriminator of 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 the probability of the input image slice being classified as a real high-SNR image slice. The abbreviation k3n64s1 stands for a kernel size equal to 3×3 for the generator or 3×3×3 for the discriminator, a kernel number equal to 64 and stride equal to 1.

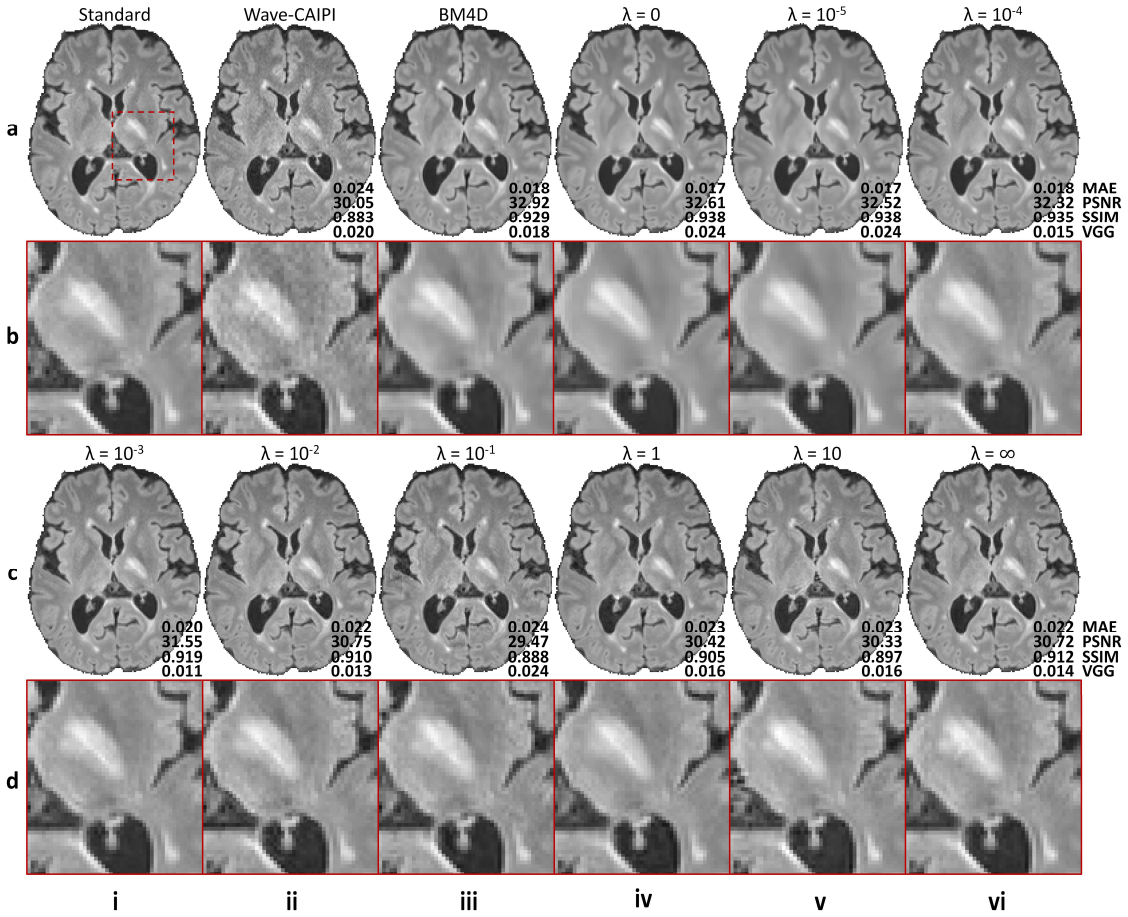


Figure 2. Effects of the adversarial loss on image quality. Representative axial image slices (rows a and c) and enlarged views of the left basal ganglia and left thalamus (rows b and d) from different methods and weights (λ) of the adversarial loss in a multiple sclerosis patient. For λ=0 (rows a, b, column iv), the training only minimizes the content loss (i.e., voxel-wise mean squared error). In this case, the GAN is effectively the generator. For λ=∞ (rows c, d, column vi), the training only minimizes the adversarial loss. Image similarity metrics including the mean absolute error (MAE), peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) and VGG perceptual loss (VGG) are listed to quantify the similarity between images from different methods and the standard FLAIR image.

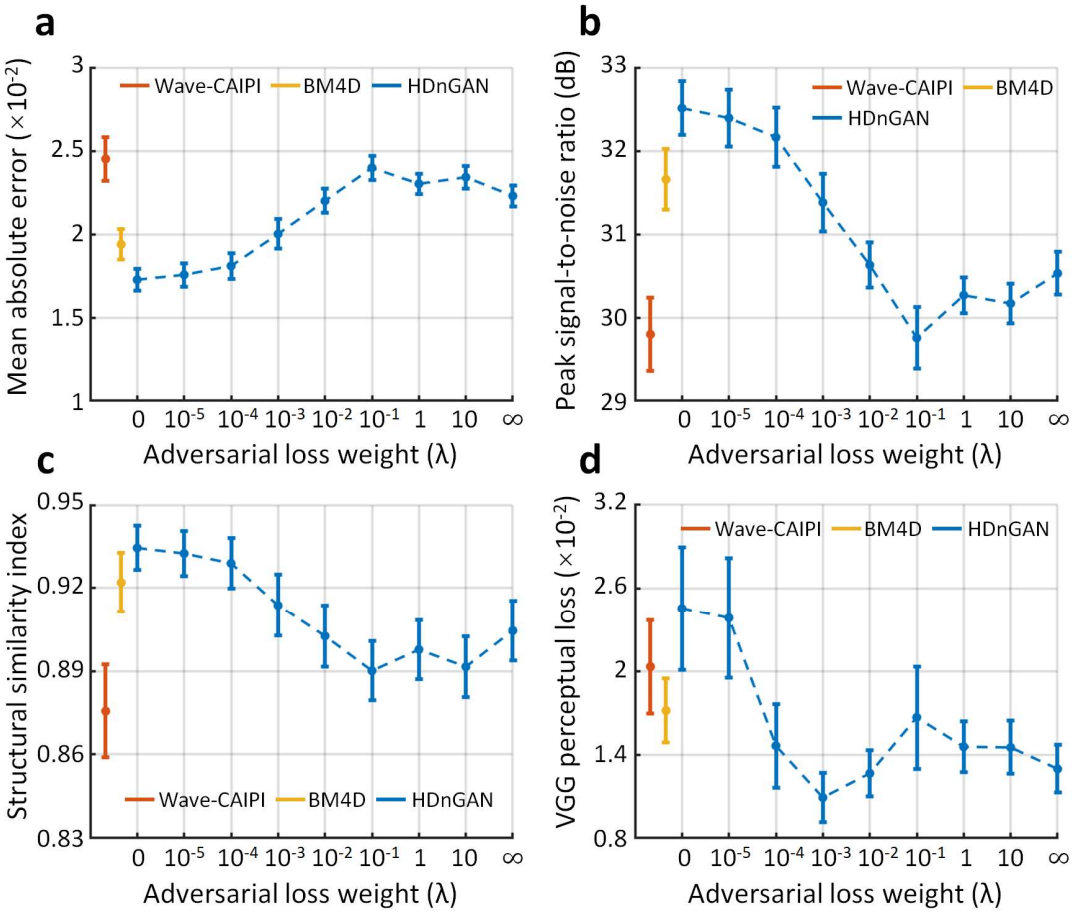


Figure 3. Quantification of the effects of the adversarial loss. The similarity between images derived from different methods and the standard FLAIR images is quantified using the mean absolute error (MAE) (a), peak signal-to-noise ratio (PSNR) (b), structural similarity index (SSIM) (c), and VGG perceptual loss (d). The red, yellow, and blue dots and error bars represent the group mean and standard deviation of different metrics for Wave-CAIPI images, BM4D-denoised results, and results of HDnGAN trained with different weights for the adversarial loss. The metrics were calculated from eight patients for evaluation which were not used for training and validation.

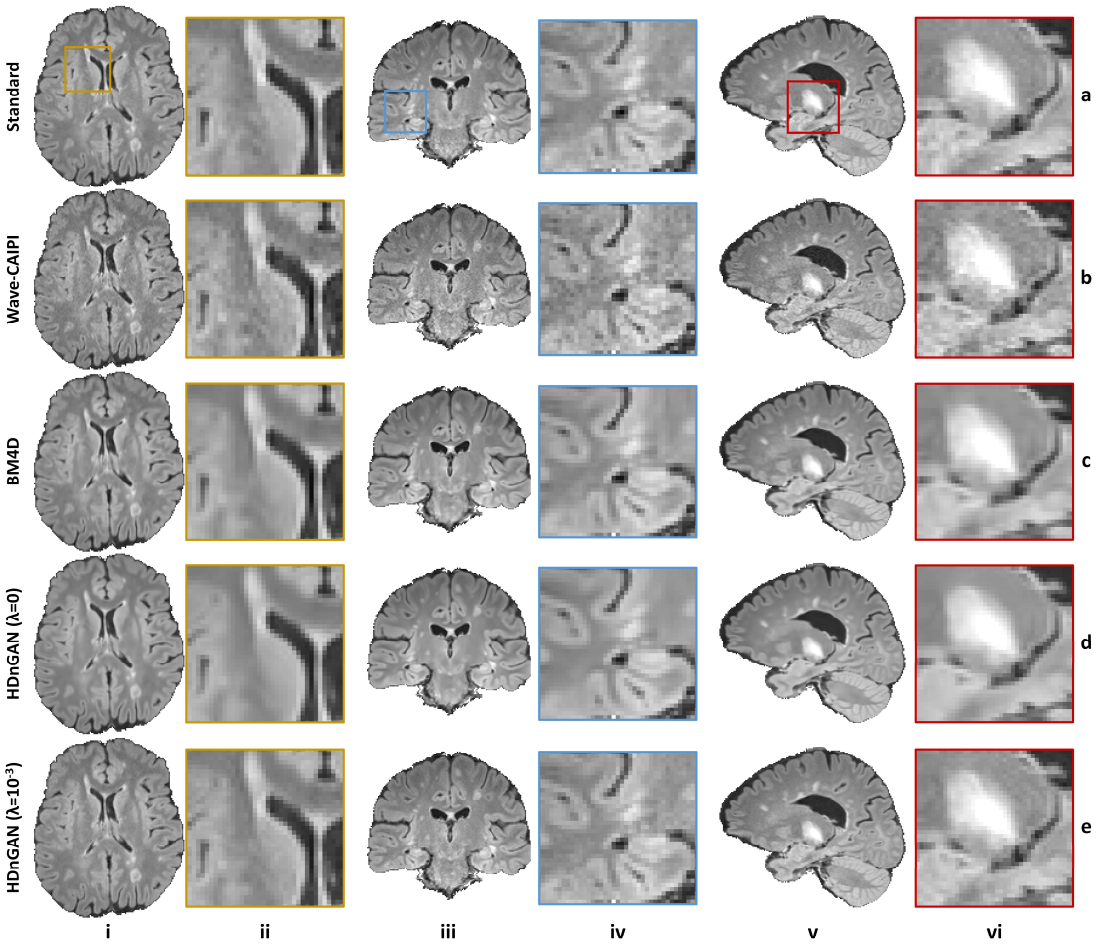


Figure 4. Resultant images along different directions. Representative axial (column i), coronal (column iii), and sagittal (column v) image slices and enlarged regions (columns ii, iv and vi) from standard T2-SPACE FLAIR data (row a), Wave-CAIPI data (row b), BM4D-denoised results (row c), HDnGAN (λ=0) results (row d) and HDnGAN ($\lambda=10^{-3}$) results (row e) from 3 different evaluation subjects.

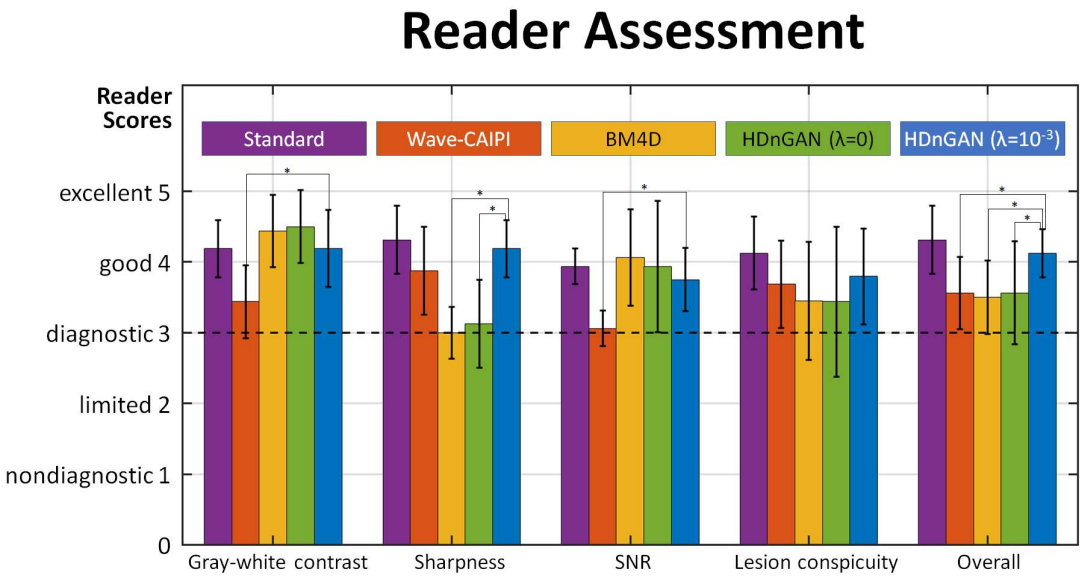
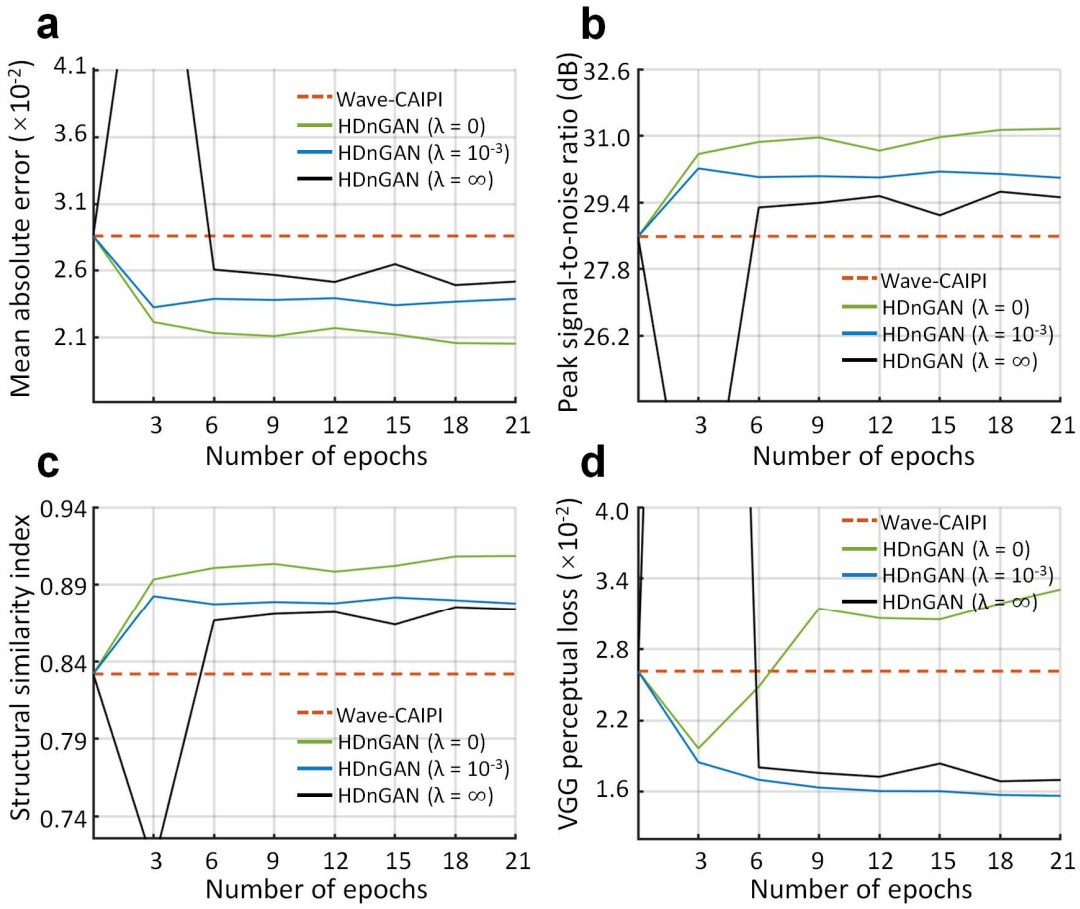
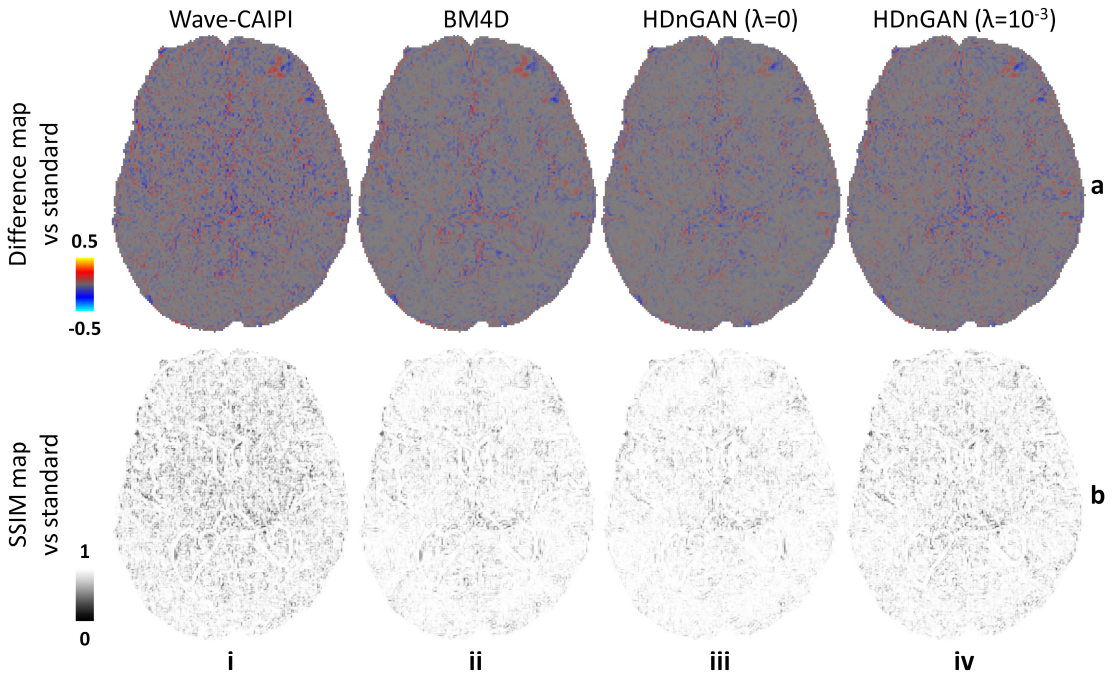


Figure 5. Results of the multi-reader assessment. The group mean and standard deviation of image quality (gray-white contrast, sharpness, SNR, lesion conspicuity, and overall quality) scores (1 nondiagnostic, 2 limited, 3 diagnostic, 4 good, 5 excellent) from two radiologists for the standard, Wave-CAIPI, BM4D, HDnGAN (λ=0), and HDnGAN ($\lambda=10^{-3}$) images of eight evaluation subjects for evaluation. t-tests assess whether scores of HDnGAN ($\lambda=10^{-3}$) images are significantly different from those for other images ($^{*}P < 0.005$ due to Bonferroni correction).

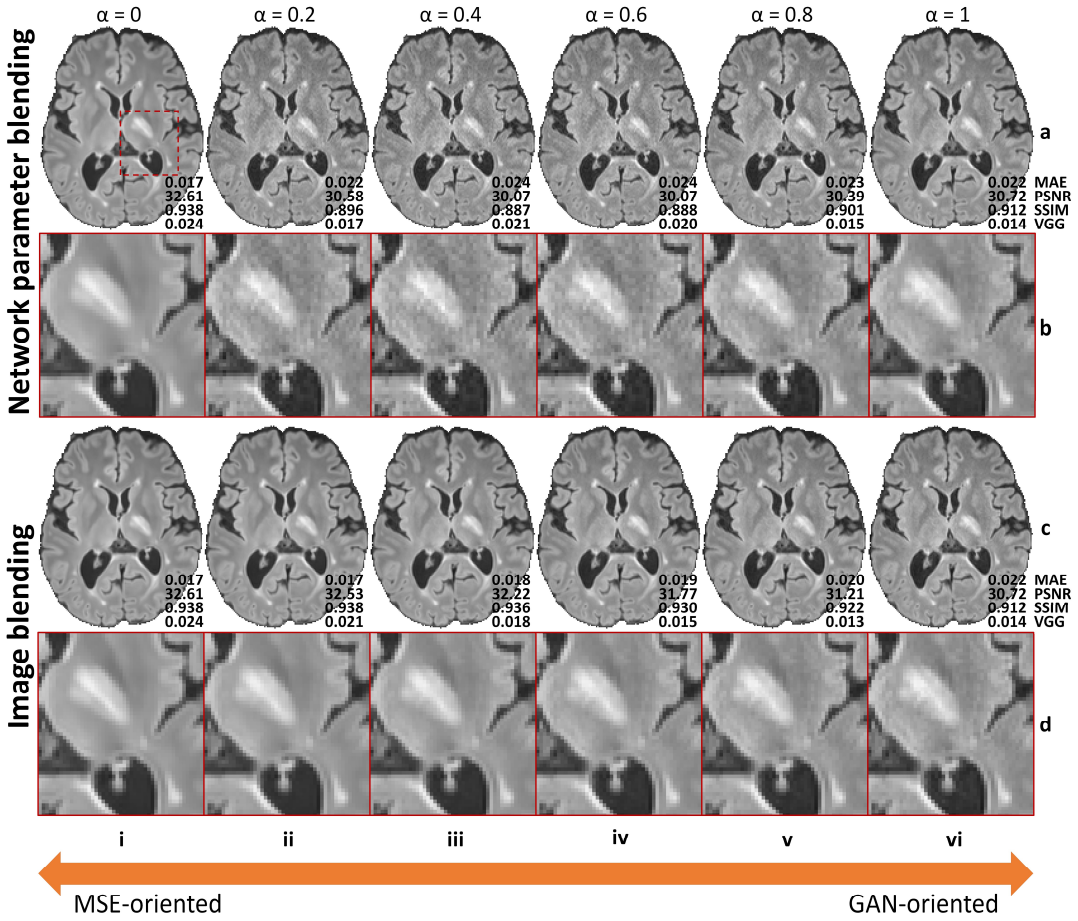
Supplementary Information



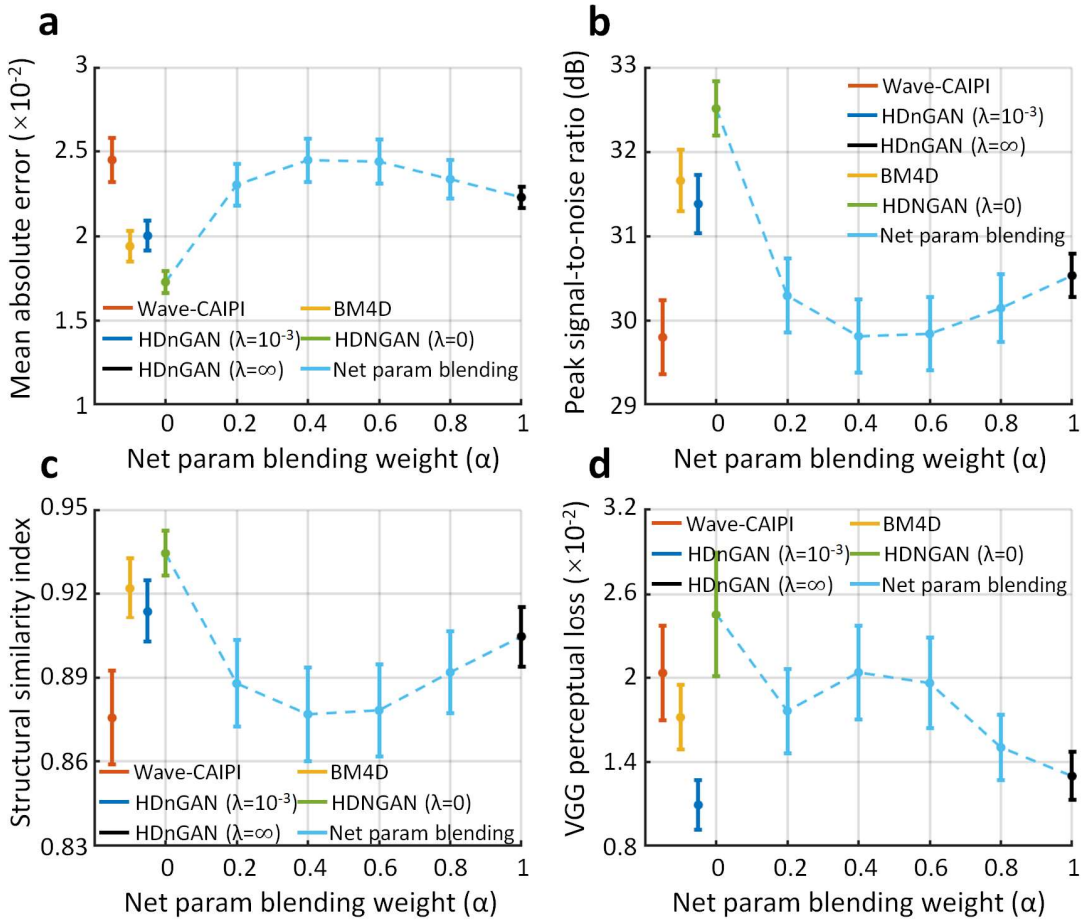
Supplementary Information Figure 1. Training convergence. The similarity between standard FLAIR images and Wave-CAIPI images (red dashed lines) as well as resultant images from HDnGAN trained with different weights for the adversarial loss at different epochs during the training (green, blue and black lines) is quantified using the mean absolute error (MAE) (a), peak signal-to-noise ratio (PSNR) (b), structural similarity index (SSIM) (c), and VGG perceptual loss (d). The metrics were calculated from five patients for validation.



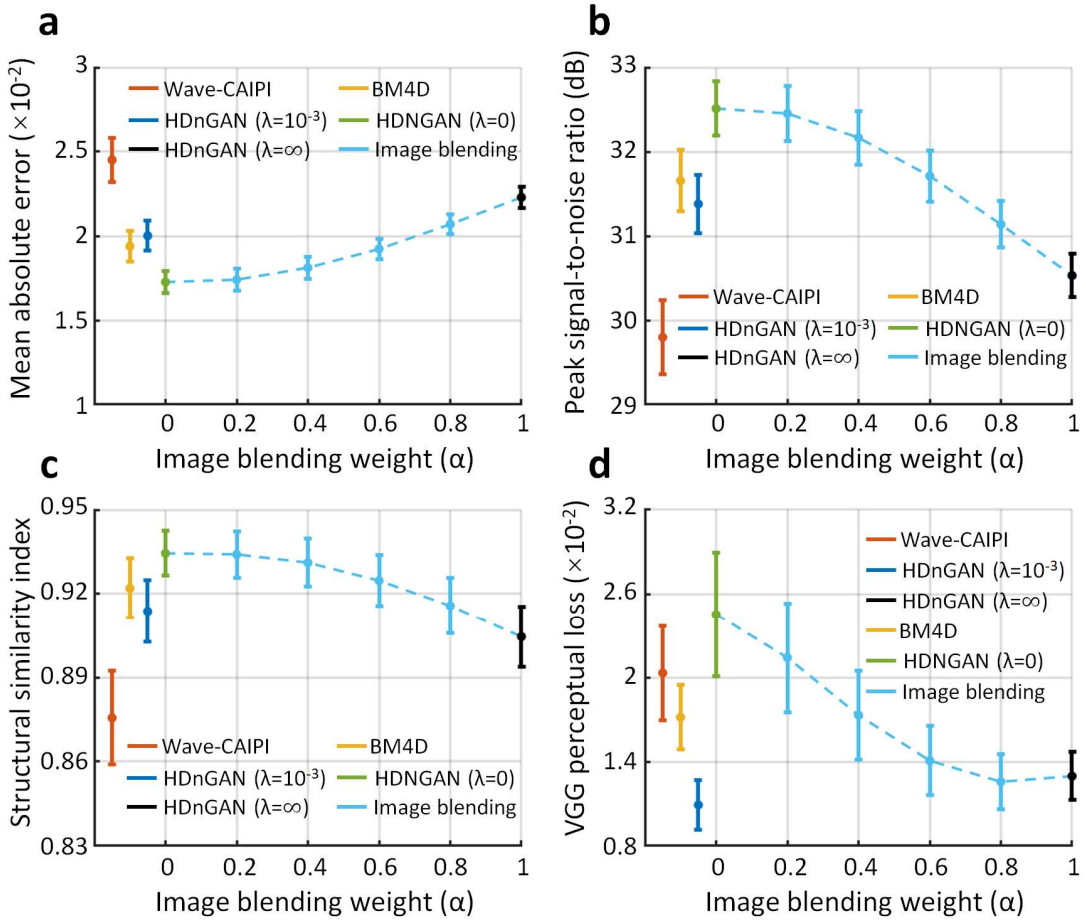
Supplementary Information Figure 2. Difference and SSIM maps. Maps of the difference (row a) and the structural similarity index (SSIM) (row b) between images from different methods and standard FLAIR images shown for a representative axial image slice from an evaluation subject.



Supplementary Information Figure 3. Effects of network parameter blending and image blending 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 GAN generator network parameter blending (rows a and b) and image blending (rows c and d) with different weights (α) of GAN generator parameters or images from GAN-oriented training in a multiple sclerosis patient. For α=0 (column i), the results are from mean squared error (MSE)-oriented training (i.e., only the voxel-wise MSE is minimized). For α=1 (column vi), the results are from GAN-oriented training (i.e., only the adversarial loss is minimized). Intermediate results are from linear weighted summation of either model parameters at each layer (network parameter blending, rows a and b) or voxel-wise image intensity (image blending, rows c and d). Image similarity metrics including the mean absolute error (MAE), peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) and VGG perceptual loss (VGG) are listed to quantify the similarity between images from different methods and the standard FLAIR image.



Supplementary Information Figure 4. Quantification of the effects of network parameter blending. The similarity between images derived from different methods and the standard FLAIR images is quantified using the mean absolute error (MAE) (a), peak signal-to-noise ratio (PSNR) (b), structural similarity index (SSIM) (c), and VGG perceptual loss (d). The red, yellow, blue, green, black, and light blue dots and error bars represent the group mean and standard deviation of different metrics for Wave-CAIPI images, BM4D results, HDnGAN ($\lambda=10^{-3}$) results, HDnGAN ($\lambda=0$) results, HDnGAN ($\lambda=\infty$) results and results of network parameter blending with different weights (α) of GAN generator parameters or images from GAN-oriented training. For $\alpha=0$, the results are from mean squared error-oriented training. The metrics were calculated from eight patients for evaluation.



Supplementary Information Figure 5. Quantification of the effects of image blending. The similarity between images derived from different methods and the standard FLAIR images is quantified using the mean absolute error (MAE) (a), peak signal-to-noise ratio (PSNR) (b), structural similarity index (SSIM) (c), and VGG perceptual loss (d). The red, yellow, blue, green, black, and light blue dots and error bars represent the group mean and standard deviation of different metrics for Wave-CAIPI images, BM4D results, HDnGAN ($\lambda=10^{-3}$) results, HDnGAN ($\lambda=0$) results, HDnGAN ($\lambda=\infty$) results and results of image blending with different weights (α) of GAN generator parameters or images from GAN-oriented training. For $\alpha=0$, the results are from mean squared error-oriented training. The metrics were calculated from eight patients for evaluation.

	Gray-white contrast	Sharpness	SNR	Lesion conspicuity	Overall
Standard vs Wave-CAIPI	7.01E-06	4.83E-02	3.62E-08	3.54E-03	1.09E-04
Standard vs BM4D	1.04E-01	2.93E-07	5.44E-01	6.92E-03	1.71E-04
Standard vs HDnGAN (λ=0)	1.97E-02	2.88E-07	1.00E+00	3.59E-02	5.26E-04
Standard vs HDnGAN ($\lambda=10^{-3}$)	1.00E+00	1.64E-01	1.88E-01	1.92E-02	8.25E-02
Wave-CAIPI vs BM4D	1.36E-05	4.59E-05	1.36E-05	4.24E-01	7.74E-01
Wave-CAIPI vs HDnGAN (λ=0)	2.19E-06	1.50E-03	2.34E-03	5.50E-01	1.00E+00
Wave-CAIPI vs HDnGAN ($\lambda=10^{-3}$)	1.50E-03	5.55E-02	3.70E-04	4.33E-01	5.26E-04
BM4D vs HDnGAN (λ=0)	3.33E-01	4.32E-01	5.44E-01	1.00E+00	7.18E-01
BM4D vs HDnGAN ($\lambda=10^{-3}$)	1.04E-01	2.88E-07	1.73E-01	5.52E-02	1.58E-04
HDnGAN (λ=0) vs HDnGAN ($\lambda=10^{-3}$)	5.55E-02	8.48E-08	5.30E-01	1.73E-01	2.76E-03

Supplementary Information Table 1. Comparison of scores from the multi-reader study. Raw *P* values from t-tests that assess whether there are significant differences between image quality scores from different methods on the eight evaluation subjects from two neuroradiologists. Raw *P* values that indicate significant differences ($P<0.005$ due to Bonferroni correction) are marked in red. All *P* values are two-sided.