Title

Undersampling artifact reduction for free-breathing 3D stack-of-radial MRI based on a deep adversarial learning network.

Authors

Gao, Chang
Ghodrati, Vahid
Shih, Shu-Fu
et al.

Publication Date

2023

DOI

10.1016/j.mri.2022.10.010

Undersampling Artifact Reduction for Free-Breathing 3D Stack-Of-Radial MRI Based on a Deep Adversarial Learning Network

Chang Gao[1,2*], Vahid Ghodrati[1,2*], Shu-Fu Shih[1,3], Holden H. Wu[1,2,3], Yongkai Liu[1,2], Marcel Dominik Nickel[4], Thomas Vahle[4], Brian Dale[5], Victor Sai[1], Ely Felker[1], Chuthaporn Surawech[1,6], Qi Miao[1,7], J. Paul Finn[1,2], Xiaodong Zhong[8], Peng Hu[1,2]

[1]Department of Radiological Sciences, University of California Los Angeles, Los Angeles, CA, United States, [2]Inter-Departmental Graduate Program of Physics and Biology in Medicine, University of California Los Angeles, Los Angeles, CA, United States, [3]Department of Bioengineering, University of California Los Angeles, Los Angeles, CA, United States, [4]MR Application Predevelopment, Siemens Healthcare GmbH, Erlangen, Germany, [5]MR R&D Collaborations, Siemens Medical Solutions USA, Inc., Cary, NC, United States, [6]Department of Radiology, Division of Diagnostic Radiology, Faculty of Medicine, Chulalongkorn University and King Chulalongkorn Memorial Hospital, Bangkok, Thailand, [7]Department of Radiology, The First Affiliated Hospital of China Medical University, Shenyang, Liaoning Province, China, [8]MR R&D Collaborations, Siemens Medical Solutions USA, Inc., Los Angeles, CA, United States

**Chang Gao and Vahid Ghodrati contributed equally to this work*

Corresponding author:
Peng Hu, PhD
David Geffen School of Medicine
University of California, Los Angeles
300 UCLA Medical Plaza Suite B119
Email: peng.hu@gmail.com

Keywords

deep learning, artifact suppression, radial acquisition, deep adversarial network, abdominal MRI

Word count: 4958 words

Abstract

Purpose: Stack-of-radial MRI allows free-breathing abdominal scans, however, it requires relatively long acquisition time. Undersampling reduces scan time but can cause streaking artifacts and degrade image quality. This study developed deep learning networks with adversarial loss and evaluated the performance of reducing streaking artifacts and preserving perceptual image sharpness.

Methods: A 3D generative adversarial network (GAN) was developed for reducing streaking artifacts in stack-of-radial abdominal scans. Training and validation datasets were self-gated to 5 respiratory states to reduce motion artifacts and to effectively augment the data. The network used a combination of three loss functions to constrain the anatomy and preserve image quality: adversarial loss, mean-squared-error loss and structural similarity index loss. The performance of the network was investigated for 3-5 times undersampled data from 2 institutions. The performance of the GAN for 5 times accelerated images was compared with a 3D U-Net and evaluated using quantitative NMSE, SSIM and region of interest (ROI) measurements as well as qualitative scores of radiologists.

Results: The 3D GAN showed similar NMSE (0.0657 vs. 0.0559, $p=0.5217$) and significantly higher SSIM (0.841 vs. 0.798, $p<.0001$) compared to U-Net. ROI analysis showed GAN removed streaks in both the background air and the tissue and was not significantly different from the reference mean and variations. Radiologists' scores showed GAN had a significant improvement of 1.6 point ($p=0.004$) on a 4-point scale in streaking score while no significant difference in sharpness score compared to the input.

Conclusion: 3D GAN removes streaking artifacts and preserves perceptual image details.

1. Introduction

Free-breathing abdominal MRI techniques can achieve volumetric coverage, high spatial resolution and high signal-to-noise ratio (SNR) for subjects with breath-hold difficulties, such as pediatric or elderly patients, and patients with disabilities, neurological disorders or inability to comply with operator instructions.[1–3] A promising approach to enabling abdominal free-breathing scans is 3D stack-of-radial MRI.[4–8] However, to reduce respiratory motion artifacts, stack-of-radial abdominal MR images require self-gating and sufficient number of radial spokes to achieve good image quality, resulting in a relatively long acquisition time compared to breath-hold techniques. The typical acquisition time of clinically used free-breathing stack-of-radial technique is approximately 7 times that of the clinically used breath-hold Cartesian technique.[9,10]

One of the practical acceleration approaches is to undersample the radial k-space and reconstruct images from the acquired radial data by using *a priori* information about the data, e.g., incoherency in the sampling pattern and redundant information in the temporal or channel direction of the data. Reconstruction methods such as parallel imaging[11,12] and compressed sensing[6,7,13] have been extensively studied. However, parallel imaging and compressed sensing methods cannot completely remove streaking artifacts with high acceleration rates.[14,15]

Deep neural networks such as convolutional neural networks (CNNs) and generative adversarial networks (GANs) have been recently used to reduce image artifacts and noise.[16–21] In particular, convolutional U-Nets have gained much attention in non-Cartesian image artifact reduction[17–19,22–24] and image reconstruction[25–28] problems. Hauptmann et al.[17] demonstrated the feasibility of using a residual U-Net to suppress streaking artifacts of undersampled real-time cardiovascular MRI. El-Rewaidy et al.[26] achieved fast and accurate reconstruction of dynamic cardiac MRI by using k-space and image domain CNNs and spatial-temporal information among neighboring time frames. However, these studies all used pixel-wise loss functions (i.e., L1/L2-norm) to train the networks, resulting in image blurring and loss of image details.[21,29,30] In addition, these networks were only tested using the data acquired in a single institution, while imaging parameters typically vary in different institution. Their performance on

data acquired in different institutions and with different acceleration factors remains to be investigated.

GANs have been shown to improve perceptual sharpness and image quality through an adversarial training process.[21,30–34] Yang et al.[21] demonstrated that the conditional GAN preserved image details for MRI de-aliasing tasks better than CNNs trained solely on pixel-wise losses. Mardani et al.[34] demonstrated that the combination of adversarial loss and pixel-wise loss resulted in high-resolution and visually appealing images. Although there has been extensive research on denoising and de-aliasing for accelerated Cartesian images,[21,34–39] the performance of the adversarial loss on destreaking tasks for non-Cartesian imaging, specifically radial k-space sampling trajectories, has not been extensively investigated. The streaking artifacts are high-frequency incoherent artifacts which are inherently different from noise and aliasing artifacts with Cartesian sampling. The controversy is that removing the streaking artifacts may remove the high-frequency image content at the same time. Thus, the feasibility of using the adversarial loss to remove streaking artifacts and preserve image details is of interest. Liu et al.[40] proposed to train the 2D adversarial network using varying undersampling patterns and showed increased robustness of the network in removing undersampling artifacts. However, the performance of a 3D network to utilize the inter-slice information has not been investigated and systematically evaluated.

This study aimed to investigate the feasibility and performance of GAN for reducing streaking artifacts in free-breathing undersampled stack-of-radial abdominal images. Specifically, a 3D GAN with the adversarial and image content losses, i.e., structural similarity index (SSIM) and L2 norm, was developed and trained. A novel data augmentation by respiratory gating into multiple respiratory states was proposed and implemented. The 3D GAN was compared to a traditional artifact-reduction pixel-wise U-Net approach with regard to streaking artifact reduction for undersampled radial k-space data with various acceleration factors. Lastly, the 3D GAN was assessed with preliminary testing data from another institution.

2. Material and Methods

2.1 Data Acquisition

The study was compliant with the Health Insurance Portability and Accountability Act and approved by the local Institutional Review Board or in accordance with the local regulatory provisions. Written informed consent was obtained for each subject. A prototype free-breathing 3D golden-angle stack-of-radial gradient echo sequence was used to acquire multi-echo and single-echo images.[8,41,9,42] At the first institution, 17 healthy subjects (Dataset A) and 15 patients (Dataset B) were scanned. Additionally, 4 healthy subjects were scanned at the second institution (Dataset C) for a preliminary testing of the generalization of the network on data from different institutions. In addition, multi-echo images and images from different areas were collected to increase the diversity of the training datasets with regard to image contrast and geometry.

2.1.1 Dataset A

Reference multi-echo and single-echo images were acquired on 17 healthy subjects (4 females, 27.3±6.3 years, body mass index (BMI): 23.0±2.8 kg/m^2; age and BMI were not recorded for two subjects) using two 3T MRI scanners (all on MAGNETOM Prisma^{fit} except one subject on MAGNETOM Skyra, Siemens Healthcare, Erlangen, Germany). To increase the size and diversity of the training and validation data, we scanned the pelvis, thigh and leg regions in addition to the abdomen region. There were totally 231 sets of 3D images (33 x 6 multi-echo and 33 single-echo images). The parameters for the multi-echo images were: TEs: 1.23, 2.46, 3.69, 4.92, 6.15, 7.38ms, TR=9ms, acquisition time=9.5 minutes, matrix size=160x160, 1500 radial spokes per partition (except 1306 radial spokes per partition for one subject), FOV=450x450mm^2, slice thickness=3.5mm, 64 slices with 2 times interpolation. The parameters for the single-echo images were the same as multi-echo images except that matrix size=256x256, 3500 spokes per partition, TE=1.23ms, TR=3ms, and acquisition time=7.2 minutes.

2.1.2 Dataset B

Reference multi-echo images were retrospectively obtained from 15 patients diagnosed with non-alcoholic fatty liver disease (6 females, 54.7±14.9 years, BMI: 29.3±2.3 kg/m^2) on a 3T MRI scanner (MAGNETOM Prisma^{fit}, Siemens Healthcare, Erlangen, Germany). The parameters for the multi-echo images were: TEs: 1.23, 2.46, 3.69, 4.92, 6.15, 7.38ms, TR=8.85ms, acquisition time=2-3 minutes, matrix size=256x256 (except

288x288 for 4 subjects), 404 radial spokes per partition, FOV=410x410 mm^2, slice thickness=5mm, 36 slices (except 40 slices for 1 subject and 48 slices for 1 subject).

2.1.3 Dataset C

Reference 2-echo MR images for 4 healthy subjects (a total of 9 acquisitions at different locations; biographical information was not recorded) on a 3T scanner (Biograph mMR, Siemens Healthcare, Erlangen, Germany) at the second institution were retrospectively collected. The acquisition parameters were: TEs: 1.29, 2.52ms, TR=4.12ms, acquisition time 5.5 minutes, matrix size=256x256, 1700 radial spokes per partition, FOV=400x400mm^2, slice thickness=3 mm, 72 slices with 2 times interpolation.

2.2 Data Preparation

A prototype retrospective undersampling framework was implemented with the image reconstruction framework of the MRI scanner (Siemens Healthcare, Erlangen, Germany), MATLAB (MathWorks, Natick, MA, USA), and command scripts. As illustrated in Figure 1, undersampling was performed by extracting a specified number of spokes from the reference data. The reference and undersampled data were then sorted into M respiratory states using self-gating and corresponding images were reconstructed using gridding followed by inverse Fourier Transform.[9,42–44] Gridding of the radial data was completed by using convolution kernels to weigh the contribution of each radial data point in a rectilinear neighborhood.[45] An adaptive coil combination method was used to reconstruct the phased-array MR images to improve the image quality under low signal-to-noise and reserve the phase data of the image.[46] The deep learning networks did not handle undersampling in slice direction.

The workflow of generating the training, validation and testing datasets is shown in Supporting Information Figure 3. Dataset A was divided into training (111 sets of 3D images from 8 subjects), validation (29 sets of 3D images from 2 subjects), and testing (91 sets of 3D images from 7 subjects) subsets. Datasets B and C were completely used as the testing data (108 sets of 3D images). There was no overlap between training and testing datasets. To increase the size of the training and validation datasets, data augmentation was applied by extracting and reconstructing 5 respiratory states, each consisting of approximately 20% of the total acquisition.

As shown in Figure 1, the streaking artifacts in different respiratory states differed, which equivalently augmented the data size by 5 times. The 5 respiratory states were independently fed into the network and yielded 555 pairs of 3D complex images for training and 145 pairs of 3D complex images for validation. According to Nyquist sampling theorem, fully sampled radial images need $sampling\ points \times \frac{\pi}{2}$ spokes,[6] corresponding to 251 and 402 spokes for matrix sizes of 160x160 and 256x256, respectively. The images with matrix sizes of 160x160 and 256x256 had 50 and 80 spokes on average after self-gating, respectively, which was equivalent to an acceleration factor of 5 with respect to the fully sampled radial k-space data.

Respiratory gating of a 40% acceptance rate achieved a good balance between motion artifact removal and scan efficiency[9,42] and was used for the testing dataset. To test the performance of the networks with different acceleration factors, we applied 3x, 4x, and 5x accelerations on the testing dataset, which corresponded to 84, 62, and 50 spokes for a matrix size of 160x160; 134, 100, and 80 spokes for a matrix size of 256x256; and 151, 114, and 91 spokes for a matrix size of 288x288, respectively. For single-echo image acquisitions, 3x, 4x, and 5x accelerations correspond to 41s, 30s, and 24s acquisition times for a matrix size of 256x256, and 26s, 19s, and 15s for a matrix size of 160x160. For multi-echo image acquisitions, 3x, 4x, and 5x accelerations correspond to 123s, 90s, and 72s acquisition times for a matrix size of 256x256, and 78s, 57s, and 45s for a matrix size of 160x160. In total, there were 199 sets of 3D testing images for each acceleration factor.

2.3 Network Architecture

The proposed GAN consisted of a generator network G and a discriminator network D. The generator G tried to imitate the distribution of a reference dataset x with the input of artifact-contaminated images x_u. The discriminator D aimed to distinguish between the *real* artifact-free images x and the synthesized *fake* images $G(x_u)$. The adversarial setting was a competing process between the generator $G_{\theta_G}(x_u): x_u \mapsto x$ and the discriminator $D_{\theta_D}(x): x \mapsto [0, 1]$, where θ_G and θ_D are the parameters of networks G and D, respectively.

As shown in Figure 2, the generator G was adapted from a 3D U-Net. The U-Net consisted of contracting and expanding paths and same-scale skip connections from the contracting path to the expanding path. The size of the 3D convolutional kernels was 3x3x3. Leaky Rectified Linear Unit (LeakyReLU) was used as a nonlinear activation function. The output of the generator and the reference images were inputs of the discriminator. The discriminator D consisted of 5 convolutional layers of size 3x3x3 interleaved with 3 average pooling layers, followed by 2 fully-connected layers with a 0.5 dropout. The output of the discriminator was *fake* or *real* (0 or 1).

The input data were complex images with real and imaginary parts as two channels (N_C) and the output data were magnitude images with one channel. Training and validation images were randomly cropped to 64x64x64x2 (64x64x64x1 for the output) ($N_x \times N_y \times N_z \times N_C$) patches, while testing images were directly fed into the network without cropping.

2.4 Network Loss Functions

The training process of a general GAN can be mathematically formulated as a minmax optimization problem to find the optimal network parameters θ_G and θ_D:

$$\min_{\theta_G} \max_{\theta_D} \mathcal{L}_{GAN}(\theta_G, \theta_D) = \mathbb{E}_{x \sim p_{train}(x)}\left[\log D_{\theta_D}(x)\right] + \mathbb{E}_{x_u \sim p_G(x_u)}\left[\log\left(1 - D_{\theta_D}\left(G_{\theta_G}(x_u)\right)\right)\right] \quad (1)$$

where $\mathbb{E}[\cdot]$ is the expectation operator. The generator loss $\mathcal{L}_G$ tries to minimize the above function, while the discriminator loss $\mathcal{L}_D$ tries to maximize it.

The adversarial loss shown in Eq. (1) can imitate the distributions of the reference image but may cause anatomical mismatch by overemphasizing the high frequency texture. Therefore, a pixel-wise mean-squared-error (MSE) loss, i.e., L2 loss, and a SSIM loss were added to constrain the output of the generator. MSE loss aims at minimizing the squared Euclidian distance between the output and the reference:

$$\min_{\theta_G} \mathcal{L}_{l_2}(\theta_G) = \left\| G_{\theta_G}(x_u) - x \right\|_2^2 \quad (2)$$

where $\|\cdot\|_2$ denotes the L2 norm. SSIM loss promotes the similarities between the local patches of the output of the generator and the reference images:

$$\min_{\theta_G} \mathcal{L}_{SSIM}(\theta_G) = -SSIM(x, G_{\theta_G}(x_u)) \quad (3)$$

Supporting Information Document S1 details the $SSIM(\cdot)$ function.

By combining the adversarial loss of the generator, L2 loss and SSIM loss, the total loss can be defined as:

$$\mathcal{L}_{TOTAL} = \alpha\mathcal{L}_G + \beta\mathcal{L}_{l_2} + \gamma\mathcal{L}_{SSIM} \quad (4)$$

Here α, β, γ are the weights of each loss respectively, and $\alpha + \beta + \gamma = 1$. Different combinations of the weights were investigated. The weights for each training were: $[\alpha, \beta, \gamma] = \{[0.6, 0.2, 0.2], [0.4, 0.4, 0.2], [0.2, 0.6, 0.2]\}$, which were referred to as GAN1, GAN2 and GAN3, respectively. Based on our experience, a weight of 0.2 for the SSIM loss was sufficient to constrain the local structures, therefore the SSIM weight of 0.2 was used for the experiments.

2.5 Network Training

The network was trained with 100 epochs using an adaptive moment estimation optimization (Adam) algorithm[47] with a momentum parameter β = 0.9. A mini-batch training was performed with 16 batches per iteration. Initial learning rates were 0.0001 for the generator and 0.00001 for the discriminator, respectively. An exponential decay annealing procedure with a decay rate of 0.96 for every 2000 iterations was implemented. Weights for the networks were initiated with random normal distributions with a variance σ = 0.01 and mean μ = 0.

Implementation and training of the GAN were completed in Python version 3.5 (Python Software Foundation, https://www.python.org/) using the Pytorch library version 1.4.0.[48] The training and testing were performed on a commercially available graphics processing unit (NVIDIA Titan RTX, 24GB memory). The total training time was about 72 hours. To determine if there were sufficient number of subjects in the training dataset, we mainly looked at two indicators: (1) if the network could converge to an equilibrium state by monitoring the training loss, and (2) if the network is overfitted to the training data by testing on unseen data (Datasets B and C).

2.6 Performance Evaluation

To evaluate the performance of the proposed network compared to the state-of-the-art deep learning method, a 3D U-Net with the same self-gated reference and

undersampled data was also trained as a baseline comparison. The 3D U-Net had the same structure as the generator with minimal adjustments to stabilize and optimize the network, including adding instance normalization layers before the convolutional layers and having 2 channel complex outputs. The MSE loss was used to train the 3D U-Net using an Adam optimizer with an initial learning rate of 0.01 which was reduced with a factor of 0.7 if no improvement for 200 iterations and a maximum of 100 epochs.

To test the networks for different acceleration factors, 3x, 4x and 5x accelerated images were used for testing while the network was trained only with 5x accelerated images. In addition, the networks were tested retrospectively with 2-echo images (Dataset C) to evaluate its performance on datasets acquired on a different scanner at a different institution.

The performance of the proposed network compared to the input (undersampled gridding reconstruction) and U-Net with respect to the reference was evaluated quantitatively by the normalized MSE (NMSE) and SSIM metrics, where lower NMSE and higher SSIM values indicated higher similarity compared to the reference.

Single-echo abdominal images from each testing subject were used for quantitative ROI analysis and qualitative image quality evaluations. A total of 28 abdominal images were selected for evaluation. Quantitative ROI analysis was conducted to compare the mean, standard deviation (SD), and coefficient of variation (CV=SD/mean) of GAN, U-Net and the input with respect to the reference. Two ROIs in the background air and two ROIs inside the liver were selected for three representative slices (near-dome, mid-level, and inferior liver).

The image quality was evaluated in a two-step process by three abdominal radiologists with an average of 8 years of abdominal image reading experience. First, to compare the performance of three GANs, ranking of 30 sets of single-echo abdominal images was performed. Resultant images of three GANs were blindly presented to the radiologists in a random order, and each radiologist independently selected the best images among the three GANs to evaluate the streaking artifacts and the perceptual sharpness. The GAN with the highest ranking was identified for further scoring. Second, 28 sets of images of four methods (the GAN with the highest ranking in the first step, U-Net, the input and the reference) were blindly presented to the radiologists. Each

radiologist independently scored each 3D image with a scale of 1 to 4 based on these criteria: (1) streaking artifacts (1=non-diagnostic, 2=severe, 3=moderate, 4=mild/minimal) and (2) visual sharpness (1=poor, 2=adequate, 3=good, 4=excellent). The radiologists rated the images for the entire abdomen and individual abdominal organs including the liver, gallbladder, and kidneys.

2.7 Statistical Analysis

All statistical analysis was done in R.[49]

For quantitative evaluations, both NMSE and SSIM data were fitted to a full factorial linear model with the image method and the acceleration as fixed effects. For the ROI analysis, mean, SD, and CV values were fitted to a linear mixed-effects model with the method and the tissue type being the fixed effects and subjects being a random effect.

For image ranking of the 3 GANs, the results were fitted to a linear mixed-effects model with the method being the fixed effect and both radiologists and subjects being random effects. For image scoring of four methods, Fleiss' kappa score was calculated[50] for both streaking and sharpness scores to assess the inter-rater reliability scores. Both streaking and sharpness scores were fitted to a linear mixed-effects model with the image method and health status (healthy vs patient) being the fixed effects and both radiologists and subjects being random effects. The random effects structure was kept maximal following the approach of Barr et al.[51]

For all linear model analysis, data were *post hoc* analyzed to determine estimated marginal means (emmeans) and their differences.[52] All possible pairwise comparisons between methods were calculated and tested for statistical significance. The p-values were corrected using Tukey's method.[53] P-values < 0.05 were considered significant.

3. Results

Figure 3 shows a healthy subject example of the 5x accelerated single-echo images of three GANs, U-Net, and the input compared to the reference. All three GANs and U-Net showed fewer streaking artifacts compared to the input image. Statistical analysis of the

radiologists' ranking showed no significant difference between the three GANs ($p>0.82$), except one radiologist preferred GAN1 over GAN3 ($p=0.0432$). The three GANs had similar rankings, and GAN1 had slightly fewer streaking artifacts compared to the other two GANs. Therefore, GAN1 was chosen for the remaining evaluations, and referred to as "GAN" for simplicity hereafter in this paper.

Supporting Information Figure S1 shows the training loss and intermediate results of GAN1. For the generator, the adversarial loss is plotted, which indicates the performance of the generator in fooling the discriminator. The discriminator contains two components associated with classification performance for both real and fake images. As shown in Figure (A), all three components converge to an equilibrium state (0.7). Figure (B) shows examples of the intermediate results, which shows the image quality improved steadily as the training progressed.

Figure 4 shows GAN, U-Net, and input single-echo images of a healthy subject (Dataset A) compared to the reference for 3-5x accelerations. GAN and U-Net trained only with 5x accelerated images successfully reduced streaking artifacts for 3x and 4x accelerated images. Consistent with 5x acceleration results, GAN and U-Net had fewer streaking artifacts compared to the input, while GAN better preserved the image details compared to U-Net.

Figure 5 shows representative 5x accelerated multi-echo images of two patients reconstructed using GAN and U-Net, along with the input and reference images. GAN and U-Net have fewer streaking artifacts compared to the input, and GAN has sharper images than U-Net. The second patient was identified with a lobulated hyperintensity lesion on the liver, as pointed out by the red arrow. This hyperintensity lesion could be detected on both the input and the GAN images, however, the lesion was better delineated on the GAN images compared to on the input images.

Figure 6 shows the performance of GAN, U-Net and the input compared to the reference on 5x accelerated 2-echo images acquired at a different institution (Dataset C). Consistent with the previous results, GAN and U-Net had fewer streaking artifacts compared to input and GAN preserved image details compared to U-Net.

Quantitative assessment of all the testing data by NMSE and SSIM are shown in Table 1. The GAN and U-Net images did not have a significantly different ($p>0.521$)

NMSE regardless of acceleration. Both U-Net and GAN had a significantly lower ($p<0.01$) NMSE than the input image for all acceleration factors. At all acceleration factors GAN had significantly higher SSIM ($p<.0001$) than both the U-Net and the input images (5x acceleration: 0.841 vs. 0.798 and 0.729, 4x acceleration: 0.857 vs. 0.807 and 0.779, 3x acceleration: 0.864 vs. 0.806 and 0.830).

Quantitative ROI analysis showed different performance of GAN and U-Net in the background air and the liver in the images. In both the air and the liver, the input images have an elevated mean (air: 12.8, $p=0.0684$; liver: 79.77, $p=0.0062$) compared to the reference images (air: 8.19, liver: 73.56). In the destreaking process, GAN removed this elevated signal and approached the mean signal of the reference images (air: 7.82, $p=0.9977$; liver: 68.65, $p=0.0599$), whereas U-Net removed the elevated signal only in the air (2.88, $p=0.0287$) and maintained a higher mean in the liver (80.92, $p=0.0009$). For the SD in the air, both GAN (2.17, $p<.0431$) and U-Net (0.80, $p<.0001$) were significantly lower than the input (6.56) and the reference (3.335). Both U-Net (0.07, $p=0.001$) and GAN (0.08, $p<.0001$) methods had a lower CV in the liver compared to the input images (0.15), meaning that both methods reduced the streaking. In the liver, both GAN ($p=1$) and U-Net ($p=0.9496$) had similar CVs as the reference (0.05), indicating that the tissue signal variations were not influenced.

Qualitative assessment by radiologists' scores of the 5x accelerated abdominal images is shown in Figure 7. Radiologist c appeared to differ substantially from the others for the streaking scores of both reference and U-Net and for the sharpness scores of the U-Net (blue markers in Figure 7). The Fleiss' kappa was 0.337 for the streaking scores, and 0.303 for the sharpness scores, indicating the disagreement between the radiologists. For streaking scores, the GAN images were 1.6 points better ($p=0.004$) than the input images but 0.8 points worse ($p=0.008$) than the reference images, and not significantly different from the U-Net images ($p=0.9218$). The U-Net images were 1.5 points better ($p=0.005$) than the input images, but not significantly different ($p>0.083$) from the reference images. There was no significant ($p>0.055$) difference in sharpness between GAN images and either the input or the U-Net images. The U-Net images were 1.4 points less ($p=0.015$) sharp than the input images. In summary, despite the radiologist reading variation in this study, both U-Net ($p=0.005$)

and GAN (p=0.004) significantly improved streaking relative to the input images, and GAN was not different from the input images in terms of sharpness (p=0.055) while U-Net was significantly blurrier than the input images (p=0.015).

4. Discussion

In this study, a 3D deep adversarial learning–based network was developed to remove streaking artifacts in free-breathing undersampled stack-of-radial MR images. The developed 3D GAN showed significantly lower NMSE and significantly higher SSIM than the input, indicating that it reduced the streaking artifacts and preserved the perceptual sharpness. The GAN also achieved significantly higher SSIM compared to deep neural networks based on the pixel-wise loss. Radiologists' evaluation showed GAN was not significantly different from the input images in terms of sharpness while U-Net was significantly blurrier than the input images.

Inspired by the recent development of deep adversarial networks, the adversarial loss was utilized in this study to learn the different distributions of the streaking artifacts and the tissue. Although studies have shown GANs could remove aliasing artifacts for Cartesian trajectories,[21,34,35,54] removing streaking artifacts was still challenging because of their high-frequency nature. The capability of the proposed GAN to reduce streaking artifacts and preserve perceptual sharpness is likely attributed to the ability of the adversarial network to distinguish the different signal distributions of the streaking artifacts and the normal tissue.[21,31,54] To preserve the consistency in the image domain, we constrained the network by adding L2 loss and SSIM loss. In choosing the loss weights, we used a narrow search of the different combinations of the weights. The main obstacle in finding the proper weights for the different loss components is the lack of a well-defined quantitative image quality metric[34,55], which could be a good direction for future studies.

In this study, 3D convolutional kernels were applied to utilize the volumetric information of 3D images. It is known that 3D CNNs have the ability to leverage interslice context and can lead to improved performance.[56–58] In Supporting Information Figure S2, we show comparison of the 2D and 3D GAN results. The 2D GAN images suffer from severe residual streaking artifacts. Our finding was consistent with the

literature which shows that the 3D network outperformed the 2D network for volumetric data,[59] especially when there was a relatively large number of training data and when the inter-slice resolution was close to the inner-slice resolution.

Generally, deep learning relies on a large amount of training data, which can be challenging to acquire or access. In this work, a novel data augmentation approach was implemented to use multi-bin self-gating to reconstruct images of multiple respiratory states for training. Although typical data augmentation methods such as scaling, flipping, rotation and translation can expand the data size, they do not apply to specific scenarios such as different breathing states. A previous study[40] demonstrated data augmentation by varying sampling pattern could improve the network robustness. We further improved the data augmentation by generating multi-bin data using respiratory gating. Self-gating not only minimized the motion artifacts in each respiratory state but also generated five volumes with different radial spokes and different motion states. Those images had inherently different streaking artifacts, which made the multi-bin self-gating an effective data augmentation method. In this study, the data were self-gated to 5 bins, but more bins could potentially be used to generate the training data if sufficient radial spokes were acquired to maintain a high SNR and minimal streaks in the originally acquired data. Additionally, multi-echo images with different image contrasts and images with different anatomical locations were acquired to increase the size and diversity of the training data. With all these data augmentation strategies, even the training data size was not significantly large in this study, the proposed network showed promising performance to reduce streaking artifacts and preserve image quality.

The networks in this study were trained solely on 5x accelerated images. Yet it was demonstrated that the proposed networks could reduce streaking artifacts on 3-5x accelerated testing data. The golden-angle radial trajectory enabled flexible undersampling and self-gating strategies and resulted in varying streaking artifact patterns in the reconstructed images. It was beneficial to train the network to learn a wide variety of streaking artifact patterns. This nature contributed to the flexibility of the proposed network to handle input data with various acceleration factors. Based on our experience, an acceleration factor of 5 was feasible and recommended for accelerated

radial acquisitions, and the proposed network could effectively reduce the streaking artifacts due to undersampling.

A common concern of deep learning models is whether the network trained with data from one institution can generate good results for the input data acquired from a different institution. In general, trained neural networks can be viewed as combinations of sequential weights, biases, and activation functions that are potentially sensitive to minor changes, e.g., intensity variations in the input data. One solution for this issue could be to expand the diversity of the training datasets, which can be achieved by acquiring more data from a large population, preferably from different institutions, but this could be very challenging practically. Another possible remedy might be the data augmentation strategies. Although the data size was very limited, this study showed promising preliminary results of the proposed network when applied on data from a different institution. Future studies are warranted to evaluate the proposed network using test data with a larger size from multiple institutions.

In image reconstruction tasks, especially Cartesian-type problems, it is common to include MR physics-based prior information to enforce data consistency by implementing a forward operation. Although this could improve the performance of the neural network, defining a forward operation model for the radial acquisition is challenging and requires a separate study. The proposed network in this study is an image-to-image operation without directly dealing with the k-space data. This offers special advantages in practice. For example, this proposed approach could be used in tandem to remove any residual streaking artifacts after conventional streaking-artifact reduction methods and/or image reconstruction algorithms. Alternatively, it could be further trained and applied to reduce the streaking artifacts from different sources.

In this study, the radiologists were not in particularly good agreement for the scores of different methods. Qualitative evaluation can directly assess the image quality but has the disadvantage of being subjective and not reproducible. We designed our study to mitigate this issue as much as possible, through our detailed instructions and a radiologist training session designed to elicit comparable responses between radiologists. However, radiologist c appeared to have distinct preferences and concepts of image quality and differ substantially from the others for the streaking scores of the

reference and the U-Net and for the sharpness scores of the U-Net, particularly in the patient population. Despite the variation, both U-Net and GAN significantly improved streaking relative to the input images, and GAN was not different from the input images in terms of sharpness while U-Net was significantly blurrier than the input images. The GAN method achieved more consistent agreement among radiologists and was more preferable to obtain a consensual improvement in destreaking.

During the data acquisition, the various breathing patterns of the subjects were not investigated. Subjects were instructed to breathe normally and avoid coughing during the scan. The various breathing patterns could impact the self-gating accuracy and the image quality. While the impact of breathing patterns and/or sudden movement on the self-gated images is of interest, our study focused on removing streaking artifacts caused by radial undersampling rather than self-gating imperfections. It is out of scope in this study and requires a separate investigational study in the future. In addition, the network investigated in this study generate only magnitude images. Future studies are warranted to output complex (real/imaginary) images which can be used for fat/water separation and PDFF estimation purposes.

5. Conclusion

A 3D deep adversarial network was developed to reduce streaking artifacts and preserve perceptual image details for free-breathing radial MRI and was evaluated in healthy subjects and patients. This study showed the feasibility and performance of the proposed network for destreaking compared to the results of a traditional U-Net approach, using undersampled radial data with different acceleration factors, as well as testing data from an institution different from where the training data were acquired. This proposed method allows accelerated free-breathing stack-of-radial MRI with reduced streaking artifacts and preserved perceptual image sharpness.

Declarations of interest

This study was supported by Siemens Medical Solutions USA, Inc. The authors MDN, TV, BD and XZ are Siemens employees.

Abbreviations

GAN: generative adversarial network

CNN: convolutional neural network

ROI: region of interest

SNR: signal-to-noise ratio

FOV: field of view

MSE: mean squared error

NMSE: normalized Mean squared error

SSIM: structural similarity index

Adam: adaptive moment estimation optimization

References

1. Block KT, Chandarana H, Milla S, et al. Towards Routine Clinical Use of Radial Stack-of-Stars 3D Gradient-Echo Sequences for Reducing Motion Sensitivity. *J Korean Soc Magn Reson Med*. 2014;18(2):87. doi:10.13104/jksmrm.2014.18.2.87

2. Fujinaga Y, Kitou Y, Ohya A, et al. Advantages of radial volumetric breath-hold examination (VIBE) with k-space weighted image contrast reconstruction (KWIC) over Cartesian VIBE in liver imaging of volunteers simulating inadequate or no breath-holding ability. *Eur Radiol*. 2016;26(8):2790-2797. doi:10.1007/s00330-015-4103-7

3. Chandarana H, Block TK, Rosenkrantz AB, et al. Free-Breathing Radial 3D Fat-Suppressed T1-Weighted Gradient Echo Sequence: A Viable Alternative for Contrast-Enhanced Liver Imaging in Patients Unable to Suspend Respiration. *Investigative Radiology*. 2011;46(10):648-653. doi:10.1097/RLI.0b013e31821eea45

4. Lin W, Guo J, Rosen MA, Song HK. Respiratory motion-compensated radial dynamic contrast-enhanced (DCE)-MRI of chest and abdominal lesions. *Magn Reson Med*. 2008;60(5):1135-1146. doi:10.1002/mrm.21740

5. Chandarana H, Feng L, Block TK, et al. Free-Breathing Contrast-Enhanced Multiphase MRI of the Liver Using a Combination of Compressed Sensing, Parallel Imaging, and Golden-Angle Radial Sampling: *Investigative Radiology*. 2013;48(1):10-16. doi:10.1097/RLI.0b013e318271869c

6. Feng L, Grimm R, Block KT, et al. Golden-angle radial sparse parallel MRI: Combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and

flexible dynamic volumetric MRI: iGRASP: Iterative Golden-Angle RAdial Sparse Parallel MRI. *Magn Reson Med*. 2014;72(3):707-717. doi:10.1002/mrm.24980

7. Feng L, Axel L, Chandarana H, Block KT, Sodickson DK, Otazo R. XD-GRASP: Golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing: XD-GRASP: Extra-Dimensional Golden-angle Radial Sparse Parallel MRI. *Magn Reson Med*. 2016;75(2):775-788. doi:10.1002/mrm.25665

8. Armstrong T, Dregely I, Stemmer A, et al. Free-breathing liver fat quantification using a multiecho 3D stack-of-radial technique: Free-Breathing Radial Liver Fat Quantification. *Magn Reson Med*. 2018;79(1):370-382. doi:10.1002/mrm.26693

9. Zhong X, Armstrong T, Nickel MD, et al. Effect of respiratory motion on free-breathing 3D stack-of-radial liver relaxometry and improved quantification accuracy using self-gating. *Magn Reson Med*. 2020;83(6):1964-1978. doi:10.1002/mrm.28052

10. Zhong X, Armstrong T, Gao C, et al. Accelerated k-space shift calibration for free-breathing stack-of-radial MRI quantification of liver fat and. *Magn Reson Med*. Published online August 19, 2021:mrm.28981. doi:10.1002/mrm.28981

11. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med*. 1999;42(5):952-962.

12. Griswold MA, Jakob PM, Heidemann RM, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med*. 2002;47(6):1202-1210. doi:10.1002/mrm.10171

13. 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. doi:10.1002/mrm.21391

14. Du J, Thornton FJ, Fain SB, et al. Artifact reduction in undersampled projection reconstruction MRI of the peripheral vessels using selective excitation. *Magn Reson Med*. 2004;51(5):1071-1076. doi:10.1002/mrm.20036

15. Xue Y, Yu J, Kang HS, Englander S, Rosen MA, Song HK. Automatic coil selection for streak artifact reduction in radial MRI. *Magn Reson Med*. 2012;67(2):470-476. doi:10.1002/mrm.23023

16. Tamada D. Review: Noise and artifact reduction for MRI using deep learning. *arXiv:200212889 [physics]*. Published online February 28, 2020. Accessed June 23, 2020. http://arxiv.org/abs/2002.12889

17. Hauptmann A, Arridge S, Lucka F, Muthurangu V, Steeden JA. Real-time cardiovascular MR with spatio-temporal artifact suppression using deep learning–proof of concept in congenital heart disease. *Magn Reson Med*. 2019;81(2):1143-1156. doi:10.1002/mrm.27480

18. Nezafat M, El-Rewaidy H, Kucukseymen S, Hauser TH, Fahmy AS. Deep convolution neural networks based artifact suppression in under-sampled radial acquisitions of myocardial T_1 mapping images. *Phys Med Biol*. 2020;65(22):225024. doi:10.1088/1361-6560/abc04f

19. Fan L, Shen D, Haji-Valizadeh H, et al. Rapid dealiasing of undersampled, non-Cartesian cardiac perfusion images using U-net. *NMR in Biomedicine*. 2020;33(5). doi:10.1002/nbm.4239

20. Gao C, Ghodrati V, Nguyen D, et al. Streaking artifact reduction of free-breathing undersampled stack-of-radial MRI using a 3D generative adversarial network. In: *Proeedings of the 29th Annual Meeting of ISMRM*. 2021:p.3523.

21. Yang G, Yu S, Dong H, et al. DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction. *IEEE Trans Med Imaging*. 2018;37(6):1310-1321. doi:10.1109/TMI.2017.2785879

22. Shen D, Ghosh S, Haji-Valizadeh H, et al. Rapid reconstruction of highly undersampled, non-Cartesian real-time cine *k* -space data using a perceptual complex neural network (PCNN). *NMR in Biomedicine*. 2021;34(1). doi:10.1002/nbm.4405

23. Kofler A, Dewey M, Schaeffter T, Wald C, Kolbitsch C. Spatio-Temporal Deep Learning-Based Undersampling Artefact Reduction for 2D Radial Cine MRI With Limited Training Data. *IEEE Trans Med Imaging*. 2020;39(3):703-717. doi:10.1109/TMI.2019.2930318

24. Chen D, Schaeffter T, Kolbitsch C, Kofler A. Ground-truth-free deep learning for artefacts reduction in 2D radial cardiac cine MRI using a synthetically generated dataset. *Phys Med Biol*. 2021;66(9):095005. doi:10.1088/1361-6560/abf278

25. Malavé MO, Baron CA, Koundinyan SP, et al. Reconstruction of undersampled 3D non-Cartesian image-based navigators for coronary MRA using an unrolled deep learning model. *Magn Reson Med*. 2020;84(2):800-812. doi:10.1002/mrm.28177

26. El-Rewaidy H, Fahmy AS, Pashakhanloo F, et al. Multi-domain convolutional neural network (MD-CNN) for radial reconstruction of dynamic cardiac MRI. *Magn Reson Med*. 2021;85(3):1195-1208. doi:10.1002/mrm.28485

27. Kofler A, Haltmeier M, Schaeffter T, Kolbitsch C. An End-To-End-Trainable Iterative Network Architecture for Accelerated Radial Multi-Coil 2D Cine MR Image Reconstruction. *Medical Physics*. Published online March 2, 2021:mp.14809. doi:10.1002/mp.14809

28. Terpstra ML, Maspero M, d'Agata F, et al. Deep learning-based image reconstruction and motion estimation from undersampled radial k-space for real-time MRI-guided radiotherapy. *Phys Med Biol*. 2020;65(15):155015. doi:10.1088/1361-6560/ab9358

29. Zhao H, Gallo O, Frosio I, Kautz J. Loss Functions for Image Restoration With Neural Networks. *IEEE Trans Comput Imaging*. 2017;3(1):47-57. doi:10.1109/TCI.2016.2644865

30. Zhu JY, Krähenbühl P, Shechtman E, Efros AA. Generative Visual Manipulation on the Natural Image Manifold. *arXiv:160903552 [cs]*. Published online December 16, 2018. Accessed March 30, 2021. http://arxiv.org/abs/1609.03552

31. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks. *arXiv:14062661 [cs, stat]*. Published online June 10, 2014. Accessed March 14, 2021. http://arxiv.org/abs/1406.2661

32. Ghodrati V, Bydder M, Ali F, et al. Retrospective respiratory motion correction in cardiac cine MRI reconstruction using adversarial autoencoder and unsupervised learning. *NMR in Biomedicine*. 2021;34(2). doi:10.1002/nbm.4433

33. Jiang W, Liu Z, Lee KH, et al. Respiratory Motion Correction in Abdominal MRI using a Densely Connected U-Net with GAN-guided Training. *arXiv:190609745 [eessIV]*. Published online June 24, 2019. Accessed July 20, 2021. https://arxiv.org/abs/1906.09745

34. 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. doi:10.1109/TMI.2018.2858752

35. 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. doi:10.1016/j.media.2019.05.001

36. Shitrit O, Riklin Raviv T. Accelerated Magnetic Resonance Imaging by Adversarial Neural Network. In: Cardoso MJ, Arbel T, Carneiro G, et al., eds. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Vol 10553. Lecture Notes in Computer Science. Springer International Publishing; 2017:30-38. doi:10.1007/978-3-319-67558-9_4

37. Quan TM, Nguyen-Duc T, Jeong WK. Compressed Sensing MRI Reconstruction using a Generative Adversarial Network with a Cyclic Loss. *IEEE Trans Med Imaging*. 2018;37(6):1488-1497. doi:10.1109/TMI.2018.2820120

38. Shaul R, David I, Shitrit O, Riklin Raviv T. Subsampled brain MRI reconstruction by generative adversarial neural networks. *Medical Image Analysis*. 2020;65:101747. doi:10.1016/j.media.2020.101747

39. Deora P, Vasudeva B, Bhattacharya S, Pradhan PM. Structure Preserving Compressive Sensing MRI Reconstruction using Generative Adversarial Networks. *arXiv:191006067 [cs, eess]*. Published online April 26, 2020. Accessed July 20, 2021. http://arxiv.org/abs/1910.06067

40. Liu F, Samsonov A, Chen L, Kijowski R, Feng L. SANTIS: Sampling-Augmented Neural neTwork with Incoherent Structure for MR image reconstruction. *Magn Reson Med*. 2019;82(5):1890-1904. doi:10.1002/mrm.27827

41. Armstrong T, Ly KV, Murthy S, et al. Free-breathing quantification of hepatic fat in healthy children and children with nonalcoholic fatty liver disease using a multi-echo 3-D stack-of-radial MRI technique. *Pediatr Radiol*. 2018;48(7):941-953. doi:10.1007/s00247-018-4127-7

42. Zhong X, Hu HH, Armstrong T, et al. Free-Breathing Volumetric Liver and Proton Density Fat Fraction Quantification in Pediatric Patients Using Stack-of-Radial MRI With Self-Gating Motion Compensation. *J Magn Reson Imaging*. 2021;53(1):118-129. doi:10.1002/jmri.27205

43. Grimm R, Block KT, Hutter J, et al. Self-gating Reconstructions of Motion and Perfusion for Free-breathing T1-weighted DCE-MRI of the Thorax Using 3D Stack-of-stars GRE Imaging. In: *Proceedings of the 20th Annual Meeting of ISMRM*. 2012:p.3814.

44. Grimm R, Bauer S, Kiefer B, Hornegger J, Block T. Optimal Channel Selection for Respiratory Self-Gating Signals. In: *Proeedings of the 21st Annual Meeting of ISMRM*. 2013:p.3749.

45. Fessler JA, Sutton BP. Nonuniform fast fourier transforms using min-max interpolation. *IEEE Trans Signal Process*. 2003;51(2):560-574. doi:10.1109/TSP.2002.807005

46. Vladimir J. Phase-dependent magnetic resonance imaging with multiple coils. German patent No. DE102009053522B4. German Patent and Trademark Office, Munich, Germany.

47. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *arXiv:14126980 [cs]*. Published online January 29, 2017. Accessed March 15, 2021. http://arxiv.org/abs/1412.6980

48. Paszke A, Gross S, Massa F, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv:191201703 [cs, stat]*. Published online December 3, 2019. Accessed March 15, 2021. http://arxiv.org/abs/1912.01703

49. R Core Team. R: A language and environment for statistical computing. Published online 2020. https://www.R-project.org/

50. Gamer M, Lemon J, Fellows I, Singh P. irr: Various coefficients of interrater reliability and agreement. R package version 0.84.1. Published online 2019. https://CRAN.R-project.org/package=irr

51. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*. 2013;68(3):255-278. doi:https://doi.org/10.1016/j.jml.2012.11.001

52. Lenth R. Emmeans: Estimated marginal means, aka least-squares means. R package version 1.5.4. Published online 2021. https://CRAN.R-project.org/package=emmeans

53. Haynes W. Tukey's Test. In: Dubitzky W, Wolkenhauer O, Cho KH, Yokota H, eds. *Encyclopedia of Systems Biology*. Springer New York; 2013:2303-2304. doi:10.1007/978-1-4419-9863-7_1212

54. Yu S, Dong H, Yang G, et al. Deep De-Aliasing for Fast Compressive Sensing MRI. *arXiv:170507137 [cs]*. Published online May 19, 2017. Accessed July 20, 2021. http://arxiv.org/abs/1705.07137

55. Ghodrati V, Shao J, Bydder M, et al. MR image reconstruction using deep learning: evaluation of network structure and loss functions. *Quant Imaging Med Surg*. 2019;9(9):1516-1527. doi:10.21037/qims.2019.08.10

56. Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *arXiv:160604797 [cs]*. Published online June 15, 2016. Accessed July 17, 2021. http://arxiv.org/abs/1606.04797

57. Pham CH, Ducournau A, Fablet R, Rousseau F. Brain MRI super-resolution using deep 3D convolutional networks. In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE; 2017:197-200. doi:10.1109/ISBI.2017.7950500

58. Du J, Wang L, Gholipour A, He Z, Jia Y. Accelerated Super-resolution MR Image Reconstruction via a 3D Densely Connected Deep Convolutional Neural Network. In: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE; 2018:349-355. doi:10.1109/BIBM.2018.8621073

59. Ghodrati V, Bydder M, Bedayat A, et al. Temporally aware volumetric generative adversarial network-based MR image reconstruction with simultaneous respiratory motion compensation: Initial feasibility in 3D dynamic cine cardiac MRI. *Magn Reson Med*. Published online July 13, 2021:mrm.28912. doi:10.1002/mrm.28912

Figures and tables:

Figure 1. Data preparation for building the proposed network including retrospective undersampling and self-gated reconstruction. Undersampling was performed by extracting specified number of radial spokes from the reference k-space. The reference and undersampled k-space data were sorted into M respiratory states using self-gating and gridding followed by inverse Fourier transform. Different respiratory states had different angle distribution of the spokes, resulting in different streaking artifacts. The k-space images were density compensated for illustration purposes only. The window/level settings of different images were adjusted separately for optimal visualization.

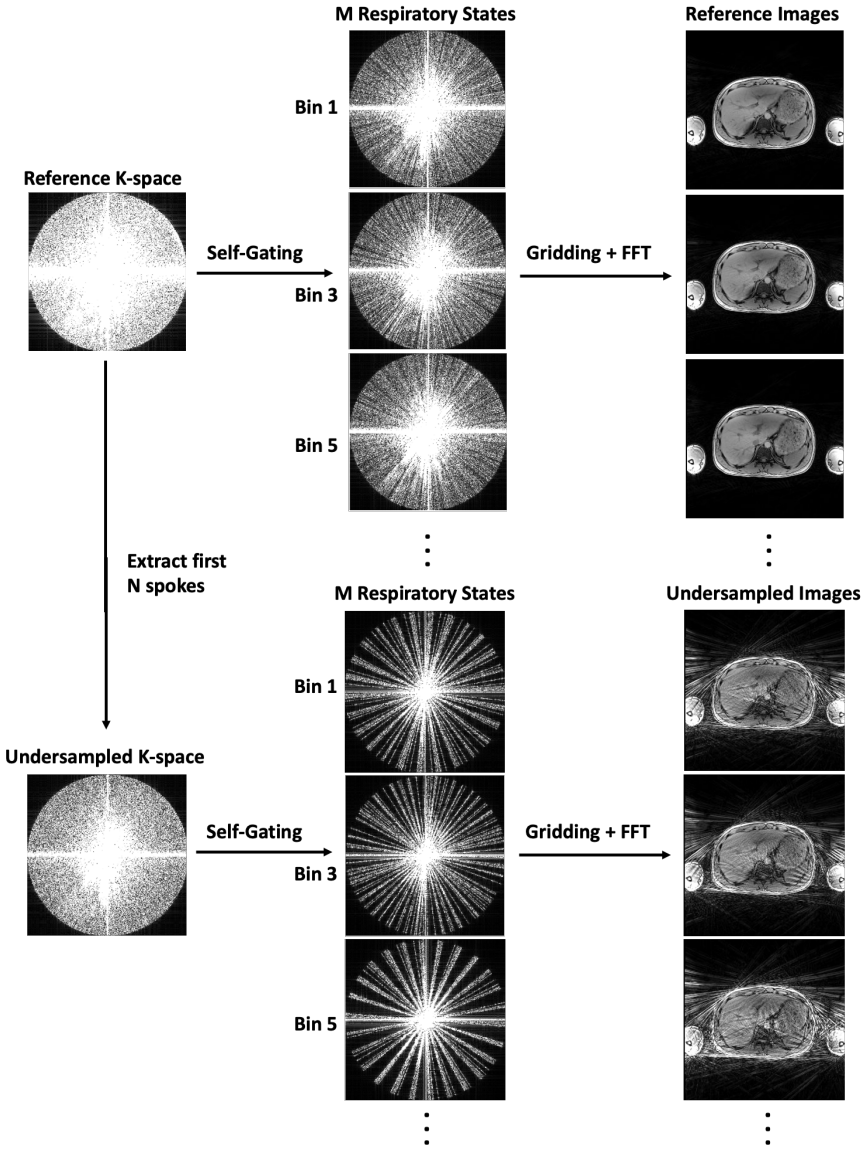

Figure 2. The 3D GAN architecture for streaking artifact reduction. The network consists of a generator G and a discriminator D. The generator is adapted from a 3D U-Net. The numbers on top of the blue cubes denote the number of channels of each layer. The complex input images are cropped to 64x64x64x2 ($N_x \times N_y \times N_z \times N_c$) patches before they are input into the network. The output and the reference images are 64x64x64 magnitude images and are inputs of the discriminator. The discriminator consists of five 3x3x3 convolution layers (with LeakyReLU) interleaved with three average pooling layers to downsample the images followed by two fully-connected (FC) layers to classify the images as *real* or *fake*.

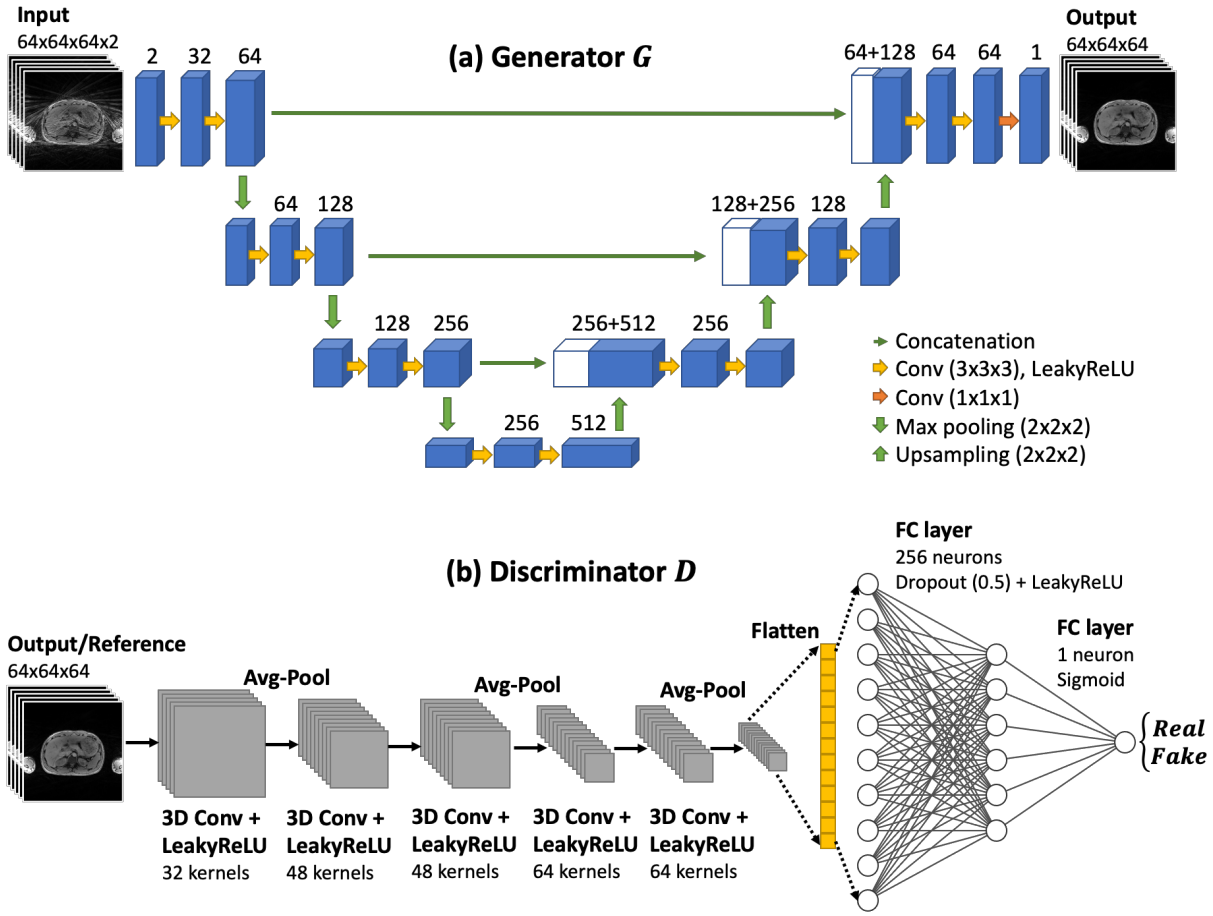

Figure 3. Representative 5x accelerated single-echo images of a healthy subject based on the three GANs tested and the U-Net, along with the input and reference images. The GANs and U-Net have superior image quality and fewer streaking artifacts compared to the input. GANs preserve the perceptual sharpness better than U-Net.

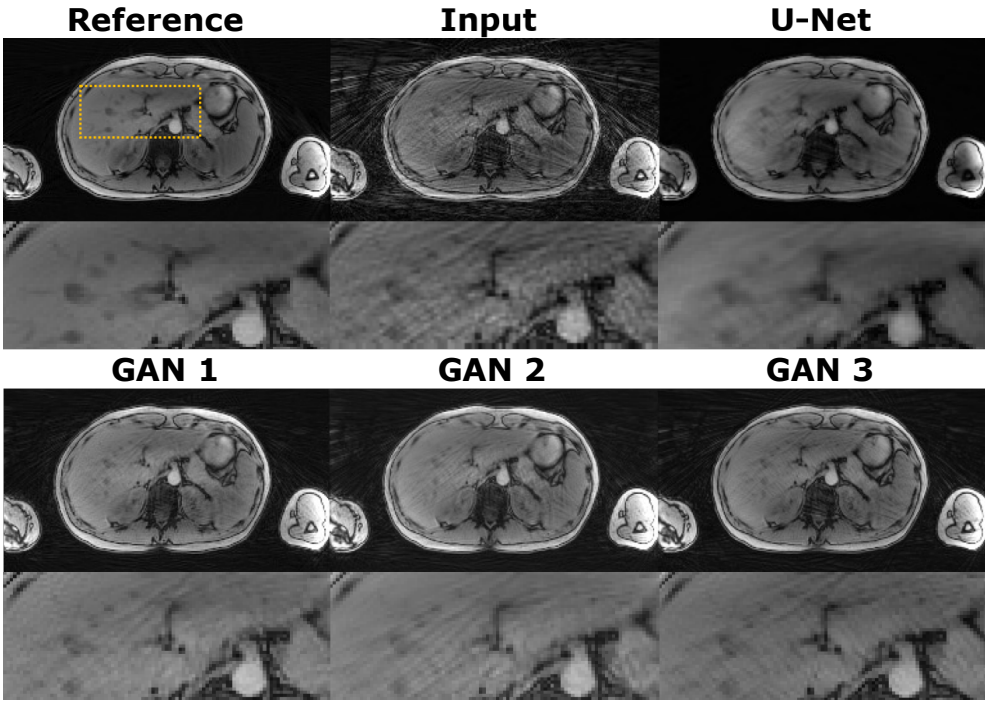

Figure 4. Representative single-echo images of a healthy subject based on GAN 1, U-Net and the undersampled input images of 3-5x accelerations compared to the reference. The top panel shows the reference image, followed by images with different acceleration factors. Comparing images of each acceleration factor, GAN and U-Net have fewer streaking artifacts compared to the input while GAN appears sharper than U-Net.

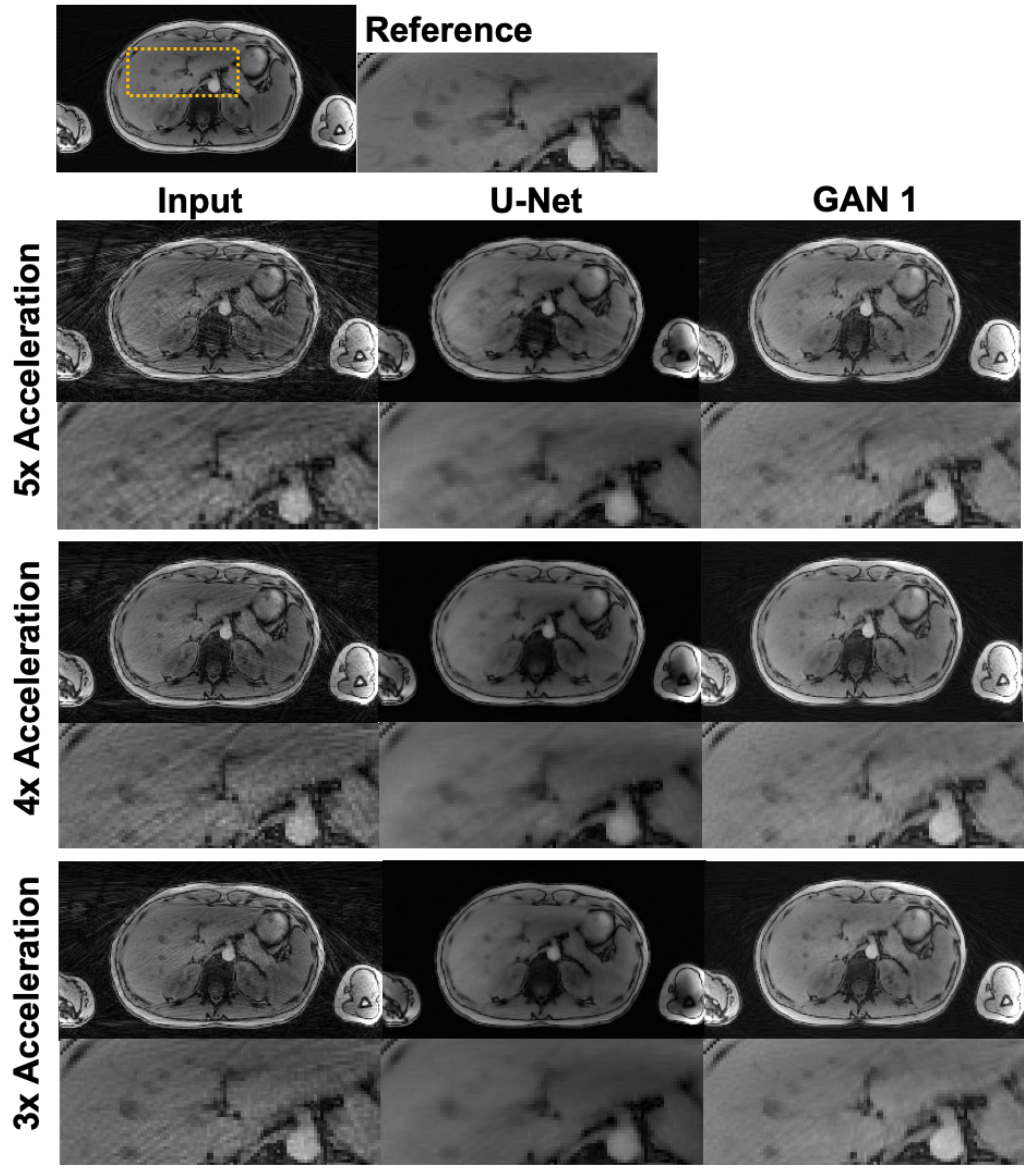

Figure 5. Representative 5x accelerated multi-echo images of two patients reconstructed using GAN 1 and U-Net, along with the input and reference images. The two panels show the images of two patients, respectively. GAN and U-Net have fewer streaking artifacts compared to the input, and GAN has sharper images than U-Net. The second patient was identified with a lobulated hyperintensity lesion on the liver, as pointed out by the red arrow. This hyperintensity lesion could be detected on both the input and the GAN images, however, the lesion was better delineated on the GAN images compared to on the input images.

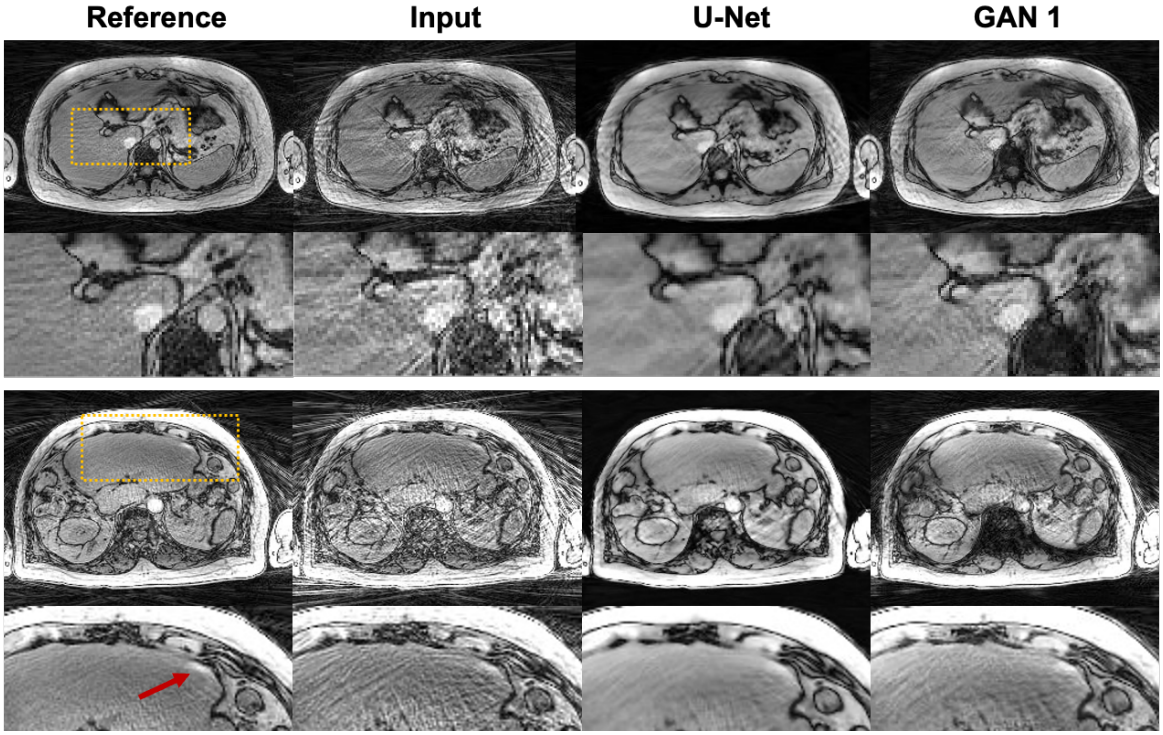

Figure 6. Representative 5x accelerated images reconstructed using GAN 1 and U-Net, along with the input and reference images for a healthy subject scanned on an MR-PET scanner the second institution. The results show that GAN and U-Net are effective at removing streaking artifacts for multi-center images and appear to be generalizable to a different platform at a different institution. GAN consistently shows better perceptual sharpness compared to U-Net.

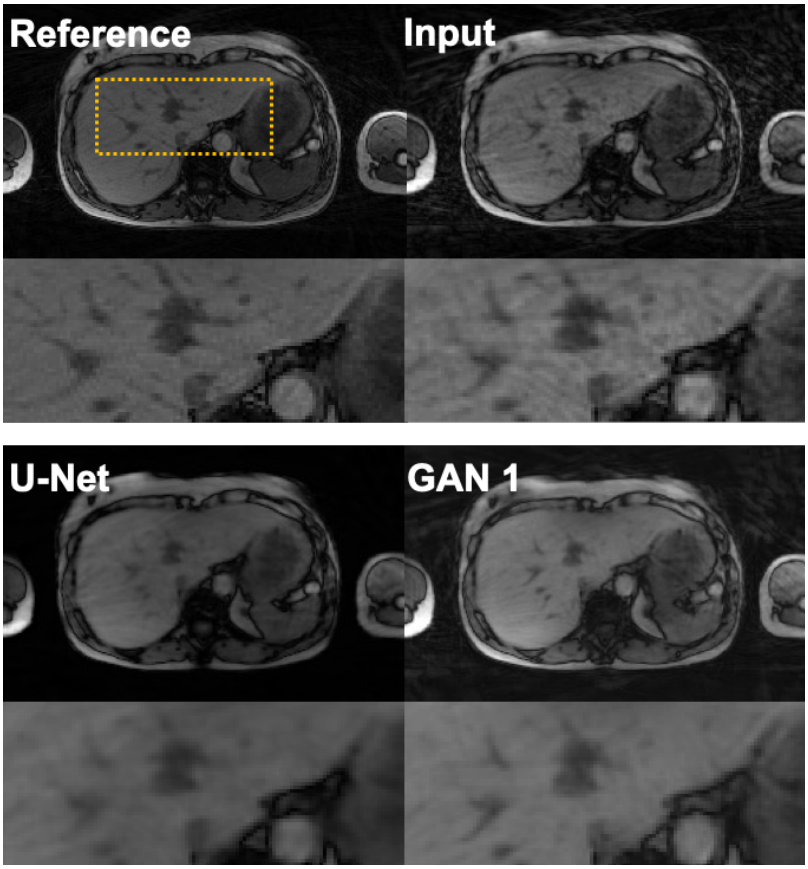

Figure 7. Radiologists' scores for streaking artifacts and perceptual sharpness. For streaking scores, the GAN images were 1.6 points better (p=0.004) than the input images but 0.8 points worse (p=0.008) than the reference images, and not significantly different from the U-Net images (p=0.9218). The U-Net images were 1.5 points better (p=0.005) than the input images, but not significantly different (p>0.083) from the reference images. There was no significant (p>0.055) difference in sharpness between GAN images and either the input or the U-Net images. The U-Net images were 1.4 points less (p=0.015) sharp than the input images.

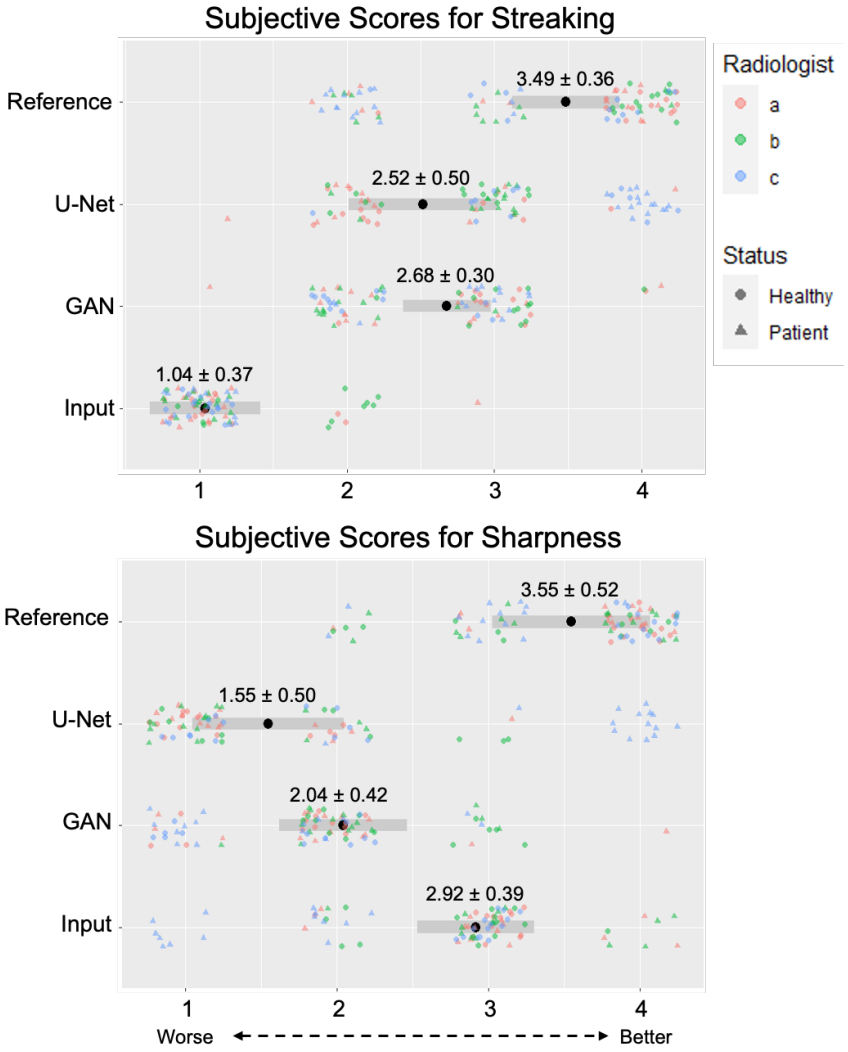

Table 1. Quantitative assessment by NMSE and SSIM of GAN, U-Net and the input. P-values are reported by pairwise comparisons between the GAN and the other methods. Estimates are reported as estimated marginal mean ± margin of error. The values in the bold text marked with asterisks (*) denote statistically significant differences compared to GAN (p-value < 0.05).

	NMSE		SSIM	
	Estimate	p-value	Estimate	p-value
3x Acceleration				
GAN	0.0571 ± 0.0124	-	0.864 ± 0.011	-
U-Net	0.0504 ± 0.0124	0.7321	**0.806 ± 0.011**	**<.0001***
Input	**0.0832 ± 0.0124**	**0.0100***	**0.830 ± 0.011**	**<.0001***
4x Acceleration				
GAN	0.0553 ± 0.0124	-	0.857 ± 0.011	-
U-Net	0.0477 ± 0.0124	0.6745	**0.807 ± 0.011**	**<.0001***
Input	**0.1111 ± 0.0124**	**<.0001***	**0.779 ± 0.011**	**<.0001***
5x Acceleration				
GAN	0.0657 ± 0.0124	-	0.841 ± 0.011	-
U-Net	0.0559 ± 0.0124	0.5217	**0.798 ± 0.011**	**<.0001***
Input	**0.1636 ± 0.0124**	**<.0001***	**0.729 ± 0.011**	**<.0001***

List of Supporting Information:

Supporting Information Document S1. Definition of SSIM function.

Supporting Information Figure S1. Training loss and examples of intermediate results of the GAN at iterations 1210 (a), 13723 (b) and 31089 (c).

Supporting Information Figure S2. Comparison of the 2D GAN and 3D GAN results.

Supporting Information Figure S3. Workflow of the generation of training, validation and testing datasets.