Three-dimensional self-attention conditional GAN with spectral normalization for multimodal neuroimaging synthesis

Haoyu Lan[1], **Alzheimer Disease Neuroimaging Initiative**, **Arthur W. Toga**[1,2], **Farshid Sepehrband**[1,2]

[1]Laboratory of NeuroImaging, USC Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Los Angeles, California, USA

[2]Alzheimer's Disease Research Center, Keck School of Medicine, University of Southern California, Los Angeles, California, USA

Abstract

Purpose: To develop a new 3D generative adversarial network that is designed and optimized for the application of multimodal 3D neuroimaging synthesis.

Methods: We present a 3D conditional generative adversarial network (GAN) that uses spectral normalization and feature matching to stabilize the training process and ensure optimization convergence (called SC-GAN). A self-attention module was also added to model the relationships between widely separated image voxels. The performance of the network was evaluated on the data set from ADNI-3, in which the proposed network was used to predict PET images, fractional anisotropy, and mean diffusivity maps from multimodal MRI. Then, SC-GAN was applied on a multidimensional diffusion MRI experiment for superresolution application. Experiment results were evaluated by normalized RMS error, peak SNR, and structural similarity.

Results: In general, SC-GAN outperformed other state-of-the-art GAN networks including 3D conditional GAN in all three tasks across all evaluation metrics. Prediction error of the SC-GAN was 18%, 24% and 29% lower compared to 2D conditional GAN for fractional anisotropy, PET and mean diffusivity tasks, respectively. The ablation experiment showed that the major contributors to the improved performance of SC-GAN are the adversarial learning and the self-attention module, followed by the spectral normalization module. In the superresolution multidimensional diffusion experiment, SC-GAN provided superior predication in comparison to 3D Unet and 3D conditional GAN.

Conclusion: In this work, an efficient end-to-end framework for multimodal 3D medical image synthesis (SC-GAN) is presented. The source code is also made available at https://github.com/Haoyulance/SC-GAN.

Correspondence Haoyu Lan, Laboratory of NeuroImaging, USC Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. haoyulan@usc.edu, *Haoyu Lan* @HaoyuLan.

Keywords

3D GAN; MRI; PET; self-attention; spectral normalization; synthesis

1 | INTRODUCTION

Medical image synthesis is a technique for generating new parametric images from other medical image modalities that contain a degree of similarity or mutual information. In recent years, deep learning methods have been vastly used in medical image synthesis or medical image transformation tasks, which are similar from a methodological point of view. These tasks include MR image reconstruction from k-s pace,[1] image superresolution to improve image resolution from low resolution,[2] image denoising by generating low-noise images from high-noise images,[3] and image synthesis by generating one image modality from one or multiple different image modalities.[4] Sparse reconstruction from k-s pace could potentially save scanning time; image denoising and superresolution can benefit diagnosis by improving image quality; PET modality synthesis from MRI modalities can reduce the patient's radiant dose; and synthetic, image-b ased data augmentation can improve lesion segmentation accuracy.[5–10] In this work, the goal was to propose a generalized deep-l earning algorithm for neuroimage transformation across image domains (eg, between MRI and PET). Generative adversarial network (GAN),[11] in particular, has been shown to be one of the effective and reliable deep-l earning algorithms for image synthesis.[12] Variants of GAN, such as conditional GAN[13] and cycle GAN,[14] have also been proposed to generalize GAN to different tasks and circumstances, including medical image synthesis.

Medical image synthesis with deep convolutional neural networks is often implemented using encoder-decoder networks, GAN, or its variants. For example, Nie et al[15] proposed a deep convolutional adversarial network to synthesize CT images from MR images. Chen et al[16] implemented an encoder-decoder convolutional neural network to synthesize PET from ultralow-dose PET and MRI. Ouyang et al[17] used conditional GAN with task-specific perceptual loss to synthesize PET from ultralow-dose PET. However, using a 2D approach on 3D data is suboptimal and inefficient because it does not incorporate the 3D spatial information in the image, and/or because it requires multiple independent implementations of the same network along different image axes.

To the best of our knowledge, Wang et al[4] first expanded the medical image synthesis GAN from 2D to 3D by using 3D convolution and transposed convolution to achieve high-quality PET image estimation from low-dose PET images. The 3D network was proposed to address the limitations of the 2D and 2.5D networks for the purpose of image synthesis. The 3D implementation of conditional GAN with no specific modification/addition to the network elements or its optimizers, however, creates an inconsistency problem due to the large differences in feature distributions,[4] negatively affecting network reliability and sometimes network fails to converge. We anticipate that a self-attention module[18] could ameliorate these limitations and further improve the performance of 3D GAN.

Here, we proposed a 3D self-attention conditional GAN (SC-GAN) constructed as follows: First, we extended 2D conditional GAN into 3D conditional GAN. Next, we added a 3D

self-attention module to generate 3D images with preserved brain structure and reduced blurriness in the synthesized images. We also introduced spectral normalization,[19] feature matching loss[9] and brain area RMS error (RMSE) loss to stabilize the network training process and prevent overfitting. To further test the effectiveness of our proposed method, SC-GAN was then tested on a challenging application of multidimensional diffusion MRI superresolution, and displayed superior performance to conventional GANs. The SC-GAN network is an end-t o-end medical image synthesis network that can be applied to high-resolution, multimodal input images (eg, $256 \times 256 \times 256$). The SC-GAN source code is made available at https://github.com/Haoyulance/SC-GAN.

The main novelties of this technique are as follows:

i. It combines 3D self-attention module into 3D conditional GAN to generate high-accuracy synthesis results with stable training process. A smooth training is then achieved by using of a series of stabilization techniques and a modified loss function; and

ii. The SC-GAN code was tested on multiple data sets across different synthesis tasks and enables multimodal input, which can be generalized to a wide range of image synthesis applications.

2 | THEORY

In this section, we introduce the 3D SC-GAN and present the relevant theory.

2.1 | Three-dimensional conditional GAN

For the main body of the SC-GAN, we used conditional GAN, which is shown to be the optimum choice of GAN for medical image synthesis and reconstruction with paired images.[3,4,17,20] A conditional GAN uses the following loss function:

$$L_{cGAN}(G, D) = \mathbb{E}_{(x, y)}[\log D(x, y)] + \mathbb{E}_{(x, z)}[\log(1 - D(x, G(x, z)))], \tag{1}$$

where x is the input modality image (also the conditional image for the conditional GAN); y is target image; and z is the noise vector. We also use G for generator and D for discriminator in the following text. As stated in Isola et al[21] and Ouyang et al,[17] noise vector z does not explicitly affect the results, and the generator would easily learn to ignore the noise vector z. We followed the same implementation principle as Isola et al[21] did, where noise vector z is no longer provided to the generator. The loss function is formulated as

$$L_{cGAN}(G, D) = \mathbb{E}_{(x, y)}[\log D(x, y)] + \mathbb{E}_{(x)}[\log(1 - D(x, G(x)))]. \tag{2}$$

We adopted the pix2pix network[21] as the network structure of 3D conditional GAN. The objective function is as follows:

$$\underset{G}{\text{argmin}} \left(\underset{D}{\text{argmax}} L_{cGAN}(G, D) + \mu L_1(G) \right), \tag{3}$$

where $L_1(G) = \mathbb{E}_{(x, y)}(\| y - G(x) \|_1)$ is the L_1 loss between the ground truth and generated image, and μ is the regularization term for the L_1 loss.

Generator and discriminator forward and backward propagate alternately until the training process reaches the Nash equilibrium and the network converges.[22]

2.2 | Three-dimensional self-attention

Self-attention allows GAN to efficiently model relationships between widely separated spatial regions,[18] to ensure that generated images contain realistic details. The image feature map $x \in R^{C \times h * w * d}$ from one intermediate hidden layer of 3D cGAN was transformed into two feature spaces, $f(x) = W_f x$ and $g(x) = W_g x$, to calculate attention. Next, the third feature space $h(x) = W_h x$ was used to calculate the attention feature map. Because the purpose of using self-attention is to measure the similarity of each voxel to the target voxel, we used the similarity scores (attentions) as weights to calculate the weighted sum representation of each target voxel. The 3D self-attention module structure is presented in Figure 1. The similarity score (attention) was calculated as follows:

$$\beta_{j, i} = \frac{\exp(S_{j, i})}{\sum_{i = 1}^{N} \exp(S_{j, i})}, \quad \text{where} S_{j, i} = f(x_j)^T g(x_i), \tag{4}$$

where $\beta_{j,i}$ is voxel j's attention to voxel i. We then calculated the attention feature for each voxel j as follows:

$$O_j = \upsilon \left(\sum_{i = 1}^{N} \beta_{j, i} h(x_i) \right), \quad \text{where } \upsilon(x) = W_\upsilon x. \tag{5}$$

The final output of the attention layer is

$$y_j = \alpha O_j + x_j. \tag{6}$$

In these formulations,

$$W_f \in R^{\bar{C} \times C}, \ W_g \in R^{\bar{C} \times C}, \ W_h \in R^{\bar{C} \times C},$$
$$W_\upsilon \in R^{C \times \bar{C}}, \ O \in R^{C \times h * w * d}, \tag{7}$$

where W_f, W_g, W_h, W_v are learned weight matrices by $1 \times 1 \times 1$ 3D convolutions; $\bar{C}$ is the number of original channels; C equals $C/8$ for memory efficiency; $h * w * d$ is the number of voxels in one feature map; and α is a learnable scalar initialized to 0.

In our network, self-attention is implemented in both the generator and the discriminator, as shown in Figure 2. When comparing our results with U-net, we added self-attention at both the encoder and the decoder of the generator to improve the synthesis performance.

2.3 | Feature matching loss

To stabilize the training, we incorporated a feature matching loss.[9] Feature matching loss is described as follows:

$$L_{FM}(G, D) = \mathbb{E}_{(x, y)} \sum_{i=1}^{T} \frac{1}{N_i} \left\| D^i(x, y) - D^i(x, G(x)) \right\|_1, \tag{8}$$

where D^i is the ith layer's feature map; T is the total number of layers of discriminator; and N_i is the number of elements in ith layer's feature map.

Feature matching loss was added only to the generator loss, because only the L_{FM} is required to be minimized at generator's optimization. The objective function with feature matching loss is

$$\underset{G}{\operatorname{argmin}} \left(\underset{D}{\operatorname{argmax}} L_{cGAN}(G, D) + \mu L_1(G) + \lambda L_{FM}(G, D) \right), \tag{9}$$

where regularization term (λ) controls the importance of the feature matching loss.

2.4 | Brain area RMSE loss

Error calculation was performed on brain voxels and the background was excluded. We calculated the RMSE between masked G and masked y and subsequently added the RMSE to the generator loss. We obtained the brain area ($mask_y$) from the ground-truth y; this was then used to calculate brain-area RMSE (B-rmse) loss as follows:

$$L_{B-rmse}(G) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(mask_y(y)^i - mask_y(G(x))^i \right)^2}, \tag{10}$$

where $mask_y(y)^i$ is the ith voxel of $mask_y(y)$, and N is the number of total voxels. The objective function of B-rmse loss is

$$\underset{G}{\operatorname{argmin}} \left(\underset{D}{\operatorname{argmax}} L_{cGAN}(G, D) + \mu L_1(G) + \lambda L_{FM}(G, D) + \gamma L_{B-rmse}(G) \right), \tag{11}$$

where γ controls the regularization term for the brain-a rea RMSE loss. In the ablation study, we found that B-rmse loss contributed to the improvement of the network performance and improved the accuracy of the synthesis. Note that B-rmse loss is not the only loss for the generator; there are combinations of L_1 loss, B-rmse loss, and feature-matching loss as well. The L_1 loss focuses on the difference between whole output and the target, whereas B-rmse loss focuses only on the brain-area difference between output and target.

2.5 | Spectral normalization

Spectral normalization is first implemented in GAN as in Miyato et al,[19] which is implemented in each layer $g: h_{in} \rightarrow h_{out}$ of the neural networks to normalize the weight matrix between two connected layers. Under the definition of Lipschitz continuity,[23] the Lipschitz norm $\|g\|_{Lip} = sup_h \sigma(\nabla g(h))$, where $\sigma(\cdot)$ is the spectral norm (the largest singular value).

Suppose a neural network $f(x, W, a) = W^{L+1} a_L(W^L(a_{L-1}(W^{L-1}(...a_1(W^1 x)...))))$, where $\{W^1, W^2, ..., W^{L+1}\}$ is the weights set and $\{a_1, a_2, ..., a_L\}$ is the element-wise nonlinear activation functions. For the linear layer $g(h) = Wh$, the norm is given by

$$\|g\|_{Lip} = sup_h \sigma(\nabla g(h)) = sup_h \sigma(W) = \sigma(W). \tag{12}$$

If the Lipschitz norm of the activation function $\|a_L\|_{Lip}$ is equal to 1, based on the inequality $\|g1 \circ g2\|_{Lip} \leq \|g1\|_{Lip} \cdot \|g2\|_{Lip}$, the following bound can be ‖derived:

$$\begin{aligned}
\left\| f \right\|_{Lip} &\leq \|g_{L+1}\|_{Lip} \cdot \|a_L\|_{Lip} \cdot \|g_L\|_{Lip} \cdots \|a_1\|_{Lip} \cdot \|g_1\|_{Lip} = \prod_{l=1}^{L+1} \|g_l\|_{Lip} \\
&= \prod_{l=1}^{L+1} \sigma(W_l).
\end{aligned} \tag{13}$$

The spectral normalization normalizes the spectral norm of the weight matrix W_l to obtain $W_{SN} = W_l / \sigma(W_l)$. Thus, if W_l is normalized as W_{SN}, then $\|f\|_{Lip} \leq \prod_{l=1}^{L+1} \sigma(W_{SN}) = 1$, which means that $\|f\|_{Lip}$ is bounded by 1. Miyato et al[19] showed the importance of Lipschitz continuity in assuring the boundness of statistics. We used spectral normalization in both the generator and the discriminator of SC-GAN.

2.6 | Regularization

To prevent overfitting, we added L2 norm regularizations to the generator and the discriminator, resulting in a final objective function of SC-GAN:

$$\begin{aligned}
\underset{G}{\mathrm{argmin}}\Big(&\underset{D}{\mathrm{argmax}}(L_{cGAN}(G, D) - \upsilon_D L_2(D)) + \mu L_1(G) + \lambda L_{FM}(G, D) \\
&+ \gamma L_{B-rmse}(G) + \upsilon_G L_2(G)\Big),
\end{aligned} \tag{14}$$

where ν_D and ν_G control the importance of L_2 regularization. Because during the training process we minimize the negative discriminator loss for the discriminator training, this objective function uses $- \nu_D L_2 (D)$ to regularize the discriminator. Note that $L_2(D)$ and $L_2(G)$ are the constraints on trainable values of discriminator and generator; however, $L_1(G)$ is the L_1 distance between generated output and ground truth.

3 | METHODS

3.1 | Study data

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative 3 (ADNI-3) database (http://adni.loni.usc.edu).[24] We downloaded de-identified MRI and PET data from ADNI-3 participants. All of the available data from ADNI-3 at the time this study was conducted were used (ADNI-3 is an ongoing project). For the PET synthesis task, 265 subjects were selected and randomly split into 207 training subjects and 58 testing subjects. For FA and MD synthesis tasks, 497 subjects were selected and randomly split into 398 training subjects and 99 testing subjects. For MRI data, T_1-weighted (T1w) and fluid-attenuated inversion-recovery (FLAIR) structural MRI and diffusion-weighted MRI were used. For PET data, we used amyloid PET. For PET synthesis, an input data set with complete T1w, FLAIR, and a target amyloid PET data set—of acceptable quality based on ADNI guidelines—were included in the analysis. For diffusion-weighted MRI synthesis, an input data set with complete T1w, FLAIR, and target diffusion-weighted MRI data set were used (all images were visually inspected).

3.2 | Magnetic resonance imaging data collection and preprocessing

Magnetic resonance imaging of the ADNI-3 was performed exclusively on 3T scanners (Siemens [Munich, Germany], Philips Healthcare [Amsterdam, Netherlands, and General Electric [Boston, MA]) using a standardized protocol. Three-dimensional T1w with 1-mm^3 resolution was acquired using an MPRAGE sequence (on Siemens and Philips scanners) and fast spoiled gradient echo (on GE scanners). For FLAIR images, a 3D sequence with similar resolution to the T1w images was acquired, providing an opportunity for accurate intrasubject intermodal co-registration. The MPRAGE T1w MRI scans were acquired using the following parameters: TR = 2300 ms, TE = 2.98 ms, FOV = 240×256 mm^2, matrix = 240×256 (variable slice number), TI = 900 ms, flip angle = 9, and effective voxel resolution = $1 \times 1 \times 1$ mm^3. The fast spoiled gradient-echo sequence was acquired using sagittal slices, with TR = 7.3 ms, TE = 3.01 ms, FOV = 256×256 mm^2, matrix = 256×256 (variable slice number), TI = 400 ms, flip angle = 11, and effective voxel resolution = $1 \times 1 \times 1$ mm^3. The 3D FLAIR images were acquired using sagittal slices, TR = 4800 ms, TE = 441 ms, FOV = 256×256 mm^2, matrix = 256×256 (variable slice number), TI = 1650 ms, flip angle = 120, and effective voxel resolution = $1 \times 1 \times 1.2$ mm^3.

The T1w preprocessing and parcellation were performed using the freely available *FreeSurfer* (ver. 5.3.0) software package,[25] and data processing was conducted using the Laboratory of Neuro Imaging pipeline system (http://pipeline.loni.usc.edu),[26–29] similar to Sta Cruz et al[30] and Sepehrband et al.[31] Field-corrected, intensity-normalized images were filtered using nonlocal mean filtering to reduce noise, and the outputs were used for the analysis. The FLAIR images for each individual were corrected for nonuniform field inhomogeneity using the N4ITK module[32] of Advanced Normalization Tools (ANTs).[33] The FLAIR images were then co-registered to T1w images using the *antsIntermodalityIntrasubject* ANTs module.

Diffusion MRI is a quantitative modality and contains microstructural information about brain tissues.[34–36] Thus, it is challenging to predict quantitative voxel-level information from the structural data, which contain only relative signal intensity values (as opposed to per-voxel quantitative values). Diffusion MRI data were acquired using the following parameters: 2D echo-planar axial imaging, with a slice thickness of 2mm, in-plane resolution of 2×2 mm^2 (matrix size of 1044×1044), flip angle of 90º, 48 diffusion-weighted images with 48 uniformly distributed diffusion encodings with b-value = 1000 s/mm^2 and 7 non-diffusion-weighted images. Diffusion MRI preprocessing and DTI fitting were performed as described in Sepehrband et al.[37,38] In brief, images were corrected for eddy current distortion and for involuntary movement using FSL TOPUP and EDDY tools.[39,40] The DTI was then fitted to diffusion data using the Quantitative Imaging Toolkit.[41] The FA and MD maps were used for the synthesis task.

3.3 | Positron emission tomography data collection and preprocessing

Amyloid PET analysis was performed according to the UC Berkeley PET methodology for quantitative measurement.[42–45] Participants were imaged with florbetapir (^{18}F-AV-45; Avid Radiopharmaceuticals, Philadelphia, PA) or ^{18}F-Florbetaben (NeuraCeq; Piramal Pharma Solutions, Mumbai, India). Six 5-minute frames of PET images were acquired 30–60 minutes following injection. Each extracted frame was co-registered to the first extracted frame and then combined into a single image, which lessened subject motion artifacts. The combined image had the same image resolution as did the original PET image (2-mm isotropic voxels). All PET images were co-registered on T1w MRI. Quantitative measurement was performed based on the standard uptake value ratio (SUVR). The brain mask, which was obtained from T1w analysis, was applied on co-registered T1w, FLAIR, and PET images. Examples of a set of input and target images are presented in Figure 3.

3.4 | Implementation, baseline models

To rigorously assess the performance of SC-GAN, we have compared it with current, well-developed medical image synthesis networks, including 2D cGAN, 3D cGAN, and attention cGAN (Att cGAN). The 2D cGAN was adopted from Ouyang et al,[17] who proposed the technique for the PET synthesis task. The 3D cGAN was initially proposed by Wang et al[4] for PET image synthesis from low-dose PET images. Attention cGAN was designed based on the attention module proposed by Oktay et al,[46] who incorporated the 3D attention module in the U-net architecture for applying pancreas segmentation (assisted by the image synthesis task). The same 3D attention module was also adopted by Liu et al[47] in the CycleGAN medical image synthesis network. To produce a fair comparison, we incorporated the aforementioned 3D attention module in conditional GAN (referred to here as Att cGAN) and compared it with SC-GAN. The main difference between Att cGAN and SC-GAN is that Att cGAN uses gated attention[46] at each skip connection of the generator, whereas SC-GAN uses self-attention in the down-sampling and up-sampling paths of the generator as well as in the discriminator. Another difference between the two techniques is that gated attention has two different inputs (ie, input features and gating signal [Supporting information Figure S1]), whereas self-attention contains only input features. All three baseline models and SC-GAN were implemented using *TensorFlow* (1.12.2) and deployed training on an NVIDIA GPU cluster (Santa Clara, CA) equipped with eight V100

GPUs (Cisco UCS C480 ML, San Jose, CA). All four sets of results are used to analyze and compare different networks' performances.

3.5 | Training and testing

The PET and DTI were up-s ampled to have the same resolution as the T1w and FLAIR (ie, $256 \times 256 \times 256$). Synthesis results generated by convolutional neural networks could be improved by adding an intensity normalization preprocessing step before training, but the synthesis results are robust for the different choices of the normalization methods.[48] We implemented Z-score normalization for all four tasks, then applied min-max rescaling to scale the voxels' intensity from between 0 to 1 before the training. The T1w and FLAIR were used as inputs for MD, FA, and PET synthesis tasks.

The 2D cGAN was implemented similar to Ouyang et al.[17] The 3D cGAN was implemented similar to Wang et al,[4] and Att cGAN was implemented similar to Oktay et al[46] and Liu.[47] We performed 5-fold cross-validation during the hyperparameter tuning phase for all four networks to obtain the optimal hyperparameters.

For SC-GAN, the optimal result was obtained with the following hyperparameters: $\mu = 200$, $\gamma = 200$, $\lambda = 20$, $\nu_G = 0.001$, $\nu_D = 0.001$, and batch size = 1. The learning rate began at 0.001, and cosine decay was used to continuously shrink the learning rate during the training process. For 2D cGAN, the hyperparameters were $\mu = 100$, $\nu_G = 0.01$, $\nu_D = 0.01$, batch size = 4, and learning rate = 0.0002; for 3D cGAN, the hyperparameters were $\mu = 200$, $\nu_G = 0.001$, $\nu_D = 0.001$, batch size = 1, and learning rate = 0.002; and for Att cGAN, the hyperparameters were $\mu = 200$, $\nu_G = 0.001$, $\nu_D = 0.001$, batch size = 1, and learning rate = 0.001.

3.6 | Evaluation criteria

Three image-quality metrics were used to evaluate the performance of the synthesis task: normalized RMSE (NRMSE), peak SNR, and structural similarity. The NRMSE reflects the normalized error without being affected by the range of the voxel values. Thus, NRMSE could be used to compare the performances of the network on different tasks directly. To enable a direct comparison between 2D cGAN and 3D networks, we evaluated the 3D output of the 2D network directly.

3.7 | Ablation study

To analyze the contribution of each component of SC-GAN, we performed an ablation study and evaluated results on the test data set of PET synthesis task. Five ablation tests were conducted for the proposed network: SC-GAN (1) without self-attention module, (2) without adversarial learning, (3) without brain area RMSE loss, (4) without spectral normalization, and (5) without feature matching loss.

3.8 | Evaluating synthesized PET

A secondary analysis was performed to compare SC-GAN results against ground-truth PET. Amyloid-b ($A\beta$) uptake was estimated from PET and synthesized PET. The $A\beta$ uptake values were then compared across clinically relevant regions. Although the focus of the

study was primarily on proposing and optimizing a neuroimage synthesis technique, this evaluation was performed to examine whether PET synthetization from MRI data can substitute for actual PET imaging. The SUVR of the $A\beta$ was calculated across subcortical and cortical regions of 10 randomly selected individuals from the ADNI-3 cohort. The SUVR values for 110 regions per participant were compared between PET and synthesized PET. The SUVRs across these regions of interest were derived using the Desikan-Killiany atlas, which was parcellated on T1w images using the *FreeSurfer* pipeline, as explained in the section 3.2. The PET images used for training were normalized using the min-max normalization approach; thus, test PET images were also normalized using the same approach before comparison.

3.9 | Superresolution application

The utility of SC-GAN in a practical application was tested for superresolution of multidimensional diffusion MRI (MUDI) as part of the Computational Diffusion MRI Workshop 2020.[49,50] multidimensional diffusion MRI enables additional sensitivity and specificity toward tissue microstructure, but is time-consuming to obtain. As such, a computational technique that allows reliable superresolution of accelerated low-resolution MUDI would be valuable. The challenge consists of two tasks: isotropic down-s ampled image superresolution and anisotropic down-sampled image superresolution. As for the acquisition protocol of the data set from the challenge, each data set contains 1344 volumes distributed over four b-shells $b \in \{500, 1000, 2000, 3000\}$ s/mm^2 with 106 uniformly spread directions; three TEs $TE \in \{80, 105, 130\}$ ms; 28 TIs $TI \in [20, 7322]$ ms; $TR = 7.5$ seconds; resolution = 2.5 mm isotropic; FOV = $220 \times 230 \times 140$ mm; SENSE = 1.9; half-scan = 0.7; multiband factor 2; and a total acquisition time of 52 minutes. Our proposed solution was developed based on SC-GAN, with additional DTI acquisition protocol information as an adaptive manner to superresolve the low-resolution DTI image to a specific resolution of DTI image. Here we describe the method we developed for this challenge and the comparison experiments using different backbone modules among 3D Unet, 3D cGAN, and SC-GAN.

In our proposed method, adaptive instance normalization (AdaIN),[51] was used to incorporate acquisition protocol information. Protocol information was added as a one-dimensional vector with six elements, including gradient-encoding directions (three elements: x, y, and z), b-value, TE, and TI. We assumed that each protocol vector had a one-to-one mapping to each volume of the MUDI. Thus, adding protocol information to the discriminator could potentially strengthen the discriminator during the adversarial process and force the generator to minimize prediction error. The AdaIN method was added to each feature map of the discriminator as follows:

$$\text{AdaIN}(x, y) = \sigma(\text{affine}(y))\left(\frac{x - \mu(x)}{\sigma(x)}\right) + \mu(\text{affine}(y)),$$

where x is the feature map of the discriminator; $\mu(x)$ is the mean and $\sigma(x)$ is the SD of the feature maps, computed across the spatial dimension independently for each sample and feature channel; and y is the protocol vector from the vector space and trained affine

transformations map protocol vector to *affine*(*y*). The values of σ and μ were then extracted to normalize the feature maps.

Training was performed using a batch size of 4 and a learning rate of 0.001 with cosine decay. Four MUDI image data sets each with image size $41 \times 46 \times 28 \times 1344$ were used for training data, and 1 subject with the same image size was used for validation data. The training patch we used is the independent volume of the subsampled MUDI image with size $41 \times 46 \times 28$, which provides the requisite sample size to train a deep neural network. The target resolution was $82 \times 92 \times 56$ We followed a progressive training strategy in two steps: (1) Train the first SC-GAN to superresolve the low-resolution data to high-resolution data, and (2) train the second SC-GAN to refine the reconstructed high-resolution data from the previous step to further reduce the mean squared error.

To facilitate a fair comparison between SC-GAN and other GAN models, we replaced SC-GAN with 3D Unet and 3D cGAN using this presented method.

4 | RESULTS

The learning curves of the GANs that were used for the PET, FA, and MD synthesis tasks are presented in Figure 4. Learning curves demonstrate the performance of different networks across training epochs. The average performance when applying the trained network on the test data is presented in Table 1, and the qualitative assessments are presented in Figure 5.

4.1 | Quantitative assessment

The learning curves demonstrate that all networks were successfully optimized, reaching a plateau within the range of the study epochs (Figure 4). The 3D cGAN and SC-GAN networks showed smooth and stable patterns in their optimization curves, whereas 2D cGAN and Att cGAN demonstrated a degree of fluctuation during their learning. The learning-curve pattern across tasks was similar in structural similarity and NRMSE. However, the peak SNR was slightly different across tasks, with PET tasks resulting in the highest peak SNR (Figure 4).

Regardless of the evaluation metric or synthesis task used, SC-GAN outperformed other networks, resulting in the lowest NRMSE and the highest peak SNR and structural similarity of any technique (Table 1). The NRMSE results showed that SC-GAN's error was 18%, 24%, and 29% lower than that of 2D cGAN across FA, PET and MD tasks, respectively. Across all tasks, the 2D network produced the lowest performance. We also displayed difference maps for the FA and MD synthesis task in Figure 5. To understand the synthesis performance in various brain regions, we measured the NRMSE of the FA task results in white matter, gray matter, and CSF regions using anatomical masks generated by ANTs.[33] The whole-brain NRMSE (mean/SD) of SC-GAN results was 0.078/0.012, and white matter, gray matter, and CSF NRMSEs were 0.088/0.012, 0.062/0.01 and 0.073/0.014, respectively.

All 3D networks outperformed the 2D network. We compared the 2D version of SC-GAN with other networks and displayed the results in Supporting Information Table S1 to

highlight the importance of incorporating 3D information into deep learning networks. The SC-GAN outperformed 3D cGAN and Att cGAN in all three tasks across all evaluation metrics. The increased performance of SC-GAN was more evident in the PET task, followed by smaller performance increases in the FA and MD tasks.

The ablation test showed that the major contributors to SC-GAN's improved performance were the adversarial learning and the self-attention module, followed by the B-rmse and spectral normalization modules (Figure 6 and Table 2). Spectral normalization and feature matching contributed to the stabilization of the SC-GAN training loss at multiples scales.

The time costs of SC-GAN and baseline models are as follows: 2D cGAN: 40 minutes per epoch and 800 minutes to reach plateau; 3D cGAN: 60 minutes per epoch and 1200 minutes to reach plateau; Att cGAN: 65 minutes per epoch and 1300 minutes to reach plateau; and SC-GAN: 65 minutes per epoch and 1300 minutes to reach plateau.

4.2 | Qualitative assessment

Figure 5 compares the studied networks qualitatively. To assess the quality of 3D synthesis images, results were presented in different planes: axial images for PET synthesis (Figure 5A), coronal images for FA synthesis (Figure 5B), and sagittal images for MD (Figure 5C). Because 2D cGAN was trained on the sagittal images, the sagittal view of the synthesized result returned the best result for the 2D network (eg, MD task) (Figure 5C), whereas the axial and coronal views showed visual discontinuity and distortion (eg, PET and FA tasks) (Figure 5A,B). Even in the sagittal view, 2D GAN generated sharp artificial boundaries (eg, ventricle boundaries in Figure 5C). The 3D networks did not suffer from either of these shortcomings, returning stable results across image dimensions.

The SC-GAN results were also visually closest to the ground-truth data in comparison with those of other networks. In particular, SC-GAN was able to capture certain image details that were hidden to other networks. For example, structural boundaries at the brainstem in the FA images were captured by SC-GAN (green dotted circle in Figure 5B), but these details were smoothed out by other networks. Cingulum bundle (blue arrows, Figure 5B) and superficial white matter (red arrow, Figure 5B) were not generated with 3D cGAN and 2D cGAN, respectively; however, these details were successfully generated by SC-GAN. We also noticed that Att cGAN failed to capture high-intensity FA across white matter (yellow arrows, Figure 5B), whereas SC-GAN demonstrated a similar intensity profile to the ground truth. It should be noted that SC-GAN also did not generate an exact match to the ground truth; artificial and incorrect features were still observed. Results from MD synthesis (Figure 5C) also showed that SC-GAN resulted in the generation of a map closer to the ground truth than those of other networks, and the map contained a higher degree of detail and fewer artifacts.

We noticed a significant correlation between PET and synthesis PET across subcortical and cortical regions (Figure 7; $P < .0001$ across all 10 tested participants). The results were consistent across all test data, with correlation coefficients ranging from $r = 0.67$ to $r = 0.95$ (all at $P < .0001$). Although synthesis PET SUVR values were significantly correlated with those of ground-truth PET, we observed that the error rate was higher when the SUVRs of

the PET images were high. These SUVR ranges correspond to regions with high clinical value, reflecting neurodegenerative pathology (high Aβ uptake).

4.3 | Application: MUDI superresolution

Figure 8 summarizes the performance of three deep neural network techniques for the application of MUDI superresolution: 3D Unet, 3D cGAN, and SC-GAN. We measured the mean squared error between output and target with the brain mask. The SC-GAN performance was superior to that of 3D Unet and 3D cGAN. The outputs of the experiments had average mean squared errors of 2.52 for 3D Unet, 2.41 for 3D cGAN, and 2.15 for SC-GAN. The inference time of these three networks on the test data was less than 3 minutes. The learning curve (Figure 8B) also demonstrated that SC-GAN can reach a lower mean RMSE across 1344 volumes compared with 3D Unet and 3D cGAN. The absolute difference maps of the generated images (one volume output of 1344 volumes; Figure 8A) showed that areas with high error rates are related to image regions affected by EPI distortion (eg, the top of the axial slice), which are therefore more challenging to predict using GAN. From a qualitative perspective, SC-GAN was also superior to 3D Unet and 3D cGAN, with lower quantities of noise and higher spatial homogeneity.

5 | DISCUSSION

Here we presented an efficient end-to-end framework for multimodal 3D medical image synthesis (SC-GAN) and validated its usefulness in PET, FA, and MD synthesis applications. To design and optimize the network, we added a 3D self-attention module to conditional GAN (cGAN), which models the similarity between adjacent and widely separated voxels of a 3D image. We also used spectral normalization and feature matching to stabilize the training process and ensure that SC-GAN could generate image details. The SC-GAN was designed to handle multimodal (multichannel) 3D images as inputs. We showed that SC-GAN significantly outperformed state-of-the-art techniques, enabling reliable and robust deep learning–based medical image synthesis for a wide range of applications.

Recent work has shown that 3D GAN can be used to improve the accuracy of medical imaging synthesis.[4,47] To evaluate the benefits of 3D implementation, we compared the performances of 2D cGAN and 3D networks. We observed intensity discontinuity and distortion in the synthesis results of 2D cGAN, highlighting the importance of using 3D neural network implementation for medical image applications. To rigorously assess SC-GAN, two existing 3D synthesis methods (3D cGAN and Att cGAN) were compared with SC-GAN. The SC-GAN technique achieved the highest performance and most stable learning curves.

Although adding the attention gate module improved 3D cGAN, the technique nevertheless returned less accurate results than SC-GAN, which uses the self-attention module. The Att cGAN method used the attention gate that filters the features propagated through the skip connections to enhance the feature maps in the up-sampling phase. Because the training process of Att cGAN is also guided by the attention gate module, the network performance

was superior to that of 3D cGAN. Qualitative results also showed that Att cGAN can generate better results compared with 3D cGAN.

The self-attention feature provided the SC-GAN network with context awareness, granting an additional degree of freedom to the synthesis process. Spectral normalization was used to stabilize the training process and prevent the training from collapsing. The ablation experiment conducted in this study, meanwhile, showed that the self-attention module contributed most to the improvement of 3D cGAN. Previous studies have shown that the self-attention module can be effective in other medical image analysis applications. Zhao et al[20] combined an object recognition network and self-attention-guided GAN into a single training process to handle the tumor detection task, whereas Li et al[3] incorporated self-attention and autoencoder perceptual loss into a convolutional neural network to denoise low-dose CT.

Three-dimensional medical image processing tasks often face dimensionality challenges, and GAN is no exception.[52] For example, 3D cGAN resulted in oversmoothed images in the FA synthesis task and created a large quantity of striping artifacts that blurred the image edges in PET and MD synthesis tasks. The SC-GAN method uses a series of regularization and stabilization techniques, namely, feature matching loss, spectral normalization loss, L1 loss, and brain-area RMSE loss. This permits stable training on high-d imensional input data (eg, the input image size of $N \times 256 \times 256 \times 256 \times 2$ that was used in this study).

It should be noted that while neuroimaging synthesis has dramatically improved over the past 5 years, our qualitative results suggest that synthesis PET cannot substitute for PET imaging yet, as pathological and clinically relevant molecular information revealed by PET may not be detected by synthesizing PET obtained from MRI data (which primarily contains structural information). This limitation does not dampen the significance of medical image synthesis, but rather calls for careful design/application when image synthesis is used. For example, studies have shown that a reliable transformation can be achieved when incorporating low-dose PET as synthesis input.[4,16,17]

The application of SC-GAN for MUDI superresolution showed that SC-GAN has the potential to be easily extended to other deep learning–based 3D medical image transformation and reconstruction tasks. The SC-GAN backbone outperformed other 3D GAN and Unet networks, resulting into a reliable MUDI superresolution that could shorten acquisition time or improve image quality through careful use of the redundant information in high-dimensional images. For the superresolution application, we lacked the additional acquisition protocol data required to perform further experiments and analyses. Correlation analysis between acquisition protocol data and DTI superresolution performance could be a future study.

Computational cost is the main limitation of SC-GAN; it requires lengthy training and heavy computational resources, such as GPU memory. Sparse attention matrix computation could be a potential solution. A future research direction could focus on generalizing SC-GAN, such as by adopting knowledge distillation mechanism into SC-GAN. As for the DTI image synthesis tasks, target images were normalized to between 0 and 1 for stable training in

the experiments, but given the quantitative nature of the DTI metrics, the normalization negatively affects the quantitative value of the metrics. Therefore, for synthesizing quantitative modalities, the normalization step should be avoided. Our in-house test suggested no normalization dependency in SC-GAN (results are not presented); therefore, the normalization was included in the DTI experiment for methodology consistency among experiments.

6 | CONCLUSIONS

The focus of this work was on enabling multimodal 3D neuroimage synthetization with GAN. The proposed method (SC-GAN) was evaluated on the challenging tasks of PET and DTI synthesis as well as MUDI superresolution, to aid in rigorous optimization of the network. The SC-GAN method was designed and assessed to enable robust and stable multimodal 3D neuroimaging synthesis. Future work could explore other SC-GAN applications; for example, SC-GAN may be used to combine MRI with low-dose PET to improve the efficacy of existing techniques.[16,17] We also expect that neuroimaging techniques with high numbers of repetitions, such as functional and diffusion MRI,[53] may benefit from SC-GAN; this is a future direction of our work.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

REFERENCES

1. Lee J, Kim H, Chung HJ, Ye JC. Deep learning fast MRI using channel attention in magnitude domain. In: Proceedings of the International Symposium on Biomedical Imaging, Iowa City, Iowa, 2020. pp 917–920.

2. Pham CH, Ducournau A, Fablet R, Rousseau F. Brain MRI superresolution using deep 3D convolutional networks. In: Proceedings of the International Symposium on Biomedical Imaging, Melbourne, Australia, 2017. pp 197–200.

3. Li M, Hsu W, Xie X, Cong J, Gao W. SACNN: self-attention convolutional neural network for low-dose CT denoising with self-supervised perceptual loss network. IEEE Trans Med Imaging. 2020;39:2289–2301. [PubMed: 31985412]

4. Wang Y, Biting Y, Wang L, et al. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. Physiol Behav. 2019;176:139–148.

5. Shin H-C, Tenenholtz NA, Rogers JK, et al. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: Proceedings of the International Workshop on Simulation and Synthesis Medical Imaging, Granada, Spain, 2018. pp 1–11.

6. Hiasa Y, Otake Y, Takao M, et al. Cross-modality image synthesis from unpaired data using CycleGAN. In: Proceedings of the International Workshop on Simulation and Synthesis Medical Imaging, Granada, Spain, 2018. pp 31–41.

7. Roy S, Carass A, Jog A, Prince JL, Lee J. MR to CT registration of brains using image synthesis. In: Proceedings of SPIE International Society for Optics and Photonics, San Diego, California, 2014. p 903419.

8. Nie D, Trullo R, Lian J, et al. Medical image synthesis with context-a ware generative adversarial networks. In: Proceedings from the International Conference on Medical Image Computing and Computer-Assisted Intervention, Québec City, Canada, 2017. pp 417–425.

9. Wang TC, Liu MY, Zhu JY, Tao A, Kautz J, Catanzaro B. High-resolution image synthesis and semantic manipulation with conditional GANs. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, 2018. pp 8798–8807.

10. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: a review. Med Image Anal. 2019;58:101552. [PubMed: 31521965]

11. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. Adv Neural Inf Process Syst. 2014;3:2672–2680.

12. Huang H, Yu PS, Wang C. An Introduction to Image Synthesis with Generative Adversarial Nets. 2018. arXiv:1803.04469v2 [cs.CV].

13. Mirza M, Osindero S. Conditional Generative Adversarial Nets. 2014. arXiv:1411.1784v1 [cs.LG].

14. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017. pp 2242–2251.

15. Nie D, Trullo R, Lian J, et al. Medical image synthesis with deep convolutional adversarial networks. IEEE Trans Biomed Eng. 2018;65:2720–2730. [PubMed: 29993445]

16. Chen KT, Gong E, Bezerra F, Macruz DC, Xu J. Ultra-low-dose 18 F-Florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs. Radiology. 2019;290:649–656. [PubMed: 30526350]

17. Ouyang J, Chen KT, Gong E, Pauly J, Zaharchuk G. Ultra-low-dose PET reconstruction using generative adversarial network with feature matching and task-specific perceptual loss. Med Phys. 2019;46:3555–3564. [PubMed: 31131901]

18. Zhang H, Goodfellow I, Metaxas D, Odena A. Self-attention generative adversarial networks. 2018. arXiv:180508318.

19. Miyato T, Kataoka T, Koyama M, Yoshida Y. Spectral normalization for generative adversarial networks. 2018. arXiv:1802.05957.

20. Zhao J, Li D, Kassam Z, et al. Tripartite-G AN: synthesizing liver contrast-enhanced MRI to improve tumor detection. Med Image Anal. 2020;63:101667. [PubMed: 32375101]

21. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, 2017. pp 5967–5976.

22. Nash JF. Equilibrium points in n-person games. Proc Natl Acad Sci USA. 1950;36:48–49. [PubMed: 16588946]

23. Hager WW. Lipschitz-continuity for constrained processes. SIAM J Control Optim. 1979;17:321–338.

24. Weiner MW, Veitch DP, Aisen PS, et al. The Alzheimer's disease neuroimaging initiative 3: continued innovation for clinical trial improvement. Alzheimer's Dement. 2017;13:561–571. [PubMed: 27931796]

25. Fischl B FreeSurfer. Neuroimage. 2012;62:774–781. [PubMed: 22248573]

26. Dinov I, Lozev K, Petrosyan P, et al. Neuroimaging study designs, computational analyses and data provenance using the LONI pipeline. PLoS One. 2010;5:e13070.

27. Dinov ID, Van Horn JD, Lozev KM, et al. Efficient, distributed and interactive neuroimaging data analysis using the LONI pipeline. Front Neuroinform. 2009;3:22. [PubMed: 19649168]

28. Moon SW, Dinov ID, Kim J, et al. Structural neuroimaging genetics interactions in Alzheimer's disease. J Alzheimer's Dis. 2015;48:1051–1063. [PubMed: 26444770]

29. Torri F, Dinov ID, Zamanyan A, et al. Next generation sequence analysis and computational genomics using graphical pipeline workflows. Genes (Basel). 2012;3:545–575. [PubMed: 23139896]

30. Sta Cruz S, Dinov ID, Herting MM, et al. Imputation strategy for reliable regional MRI morphological measurements. Neuroinformatics. 2020;18:59–70. [PubMed: 31054076]

31. Sepehrband F, Lynch KM, Cabeen RP, et al. Neuroanatomical morphometric characterization of sex differences in youth using statistical learning. Neuroimage. 2018;172:217–227. [PubMed: 29414494]

32. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. 2010;29:1310–1320. [PubMed: 20378467]

33. Avants BB, Tustison N, Song G. Advanced normalization tools (ANTS). In: OR Insight. London, United Kingdom: Palgrave Macmillan; 2009. pp 1–35.

34. Sepehrband F, Clark KA, Ullmann JFP, et al. Brain tissue compartment density estimated using diffusion-weighted MRI yields tissue parameters consistent with histology. Hum Brain Mapp. 2015;36:3687–3702. [PubMed: 26096639]

35. Sepehrband F, O'Brien K, Barth M. A time-efficient acquisition protocol for multipurpose diffusion-weighted microstructural imaging at 7 Tesla. Magn Reson Med. 2017;78:2170–2184. [PubMed: 28191681]

36. Le Bihan D, Mangin J-F, Poupon C, et al. Diffusion tensor imaging: concepts and applications. J Magn Reson Imaging. 2001;13:534–546. [PubMed: 11276097]

37. Sepehrband F, Cabeen RP, Choupan J, Barisano G, Law M, Toga AW. Perivascular space fluid contributes to diffusion tensor imaging changes in white matter. Neuroimage. 2019;197:243–254. [PubMed: 31051291]

38. Sepehrband F, Cabeen RP, Barisano G, et al. Nonparenchymal fluid is the source of increased mean diffusivity in preclinical Alzheimer's disease. Alzheimer's Dement Diagnosis, Assess Dis Monit. 2019;11:348–354.

39. Andersson JLR, Skare S, Ashburner J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage. 2003;20:870–888. [PubMed: 14568458]

40. Andersson JLR, Xu J, Yacoub E, Auerbach E, Moeller S, Ugurbil K. A comprehensive Gaussian process framework for correcting distortions and movements in diffusion images. In: Proceedings of the 20th Annual Meeting of ISMRM-ESMRMB, Melbourne, Australia, 2012. p 2426.

41. Cabeen RP, Laidlaw DH, Toga AW. Quantitative imaging toolkit: software for interactive 3D visualization, data exploration, and computational analysis of neuroimaging datasets. In: Proceedings of the Joint Annual Meeting of ISMRM-E SMRMB, Paris, France, 2018. pp 12–14.

42. Landau SM, Thomas BA, Thurfjell L, et al. Amyloid PET imaging in Alzheimer's disease: a comparison of three radiotracers. Eur J Nucl Med Mol Imaging. 2014;41:1398–1407. [PubMed: 24647577]

43. Schöll M, Lockhart SN, Schonhaut DR, et al. PET imaging of tau deposition in the aging human brain. Neuron. 2016;89:971–982. [PubMed: 26938442]

44. Baker SL, Lockhart SN, Price JC, et al. Reference tissue–based kinetic evaluation of 18F-AV-1451 for tau imaging. J Nucl Med Soc Nuclear Med. 2017;58:332–338.

45. Landau SM, Fero A, Baker SL, et al. Measurement of longitudinal β-amyloid change with 18F-florbetapir PET and standardized uptake value ratios. J Nucl Med Soc Nuclear Med. 2015;56: 567–574.

46. Oktay O, Schlemper J, Folgoc LL, et al. Attention U-Net: learning where to look for the pancreas. 2018. arXiv:1804.03999v3 [cs. CV].

47. Liu X, Wei X, Yu A, et al. Unpaired data based cross-domain synthesis and segmentation using attention neural network. In: Proceedings of the 11th Asian Conference on Machine Learning, Nagoya, Japan, 2019. pp 987–1000.

48. Reinhold JC, Dewey BE, Carass A, Prince JL. Evaluating the impact of intensity normalization on MR image synthesis. 2019. arXiv:1812.04652.

49. Pizzolato M, Palombo M, Bonet-C arne E, et al. Acquiring and predicting multidimensional diffusion (MUDI) data: an open challenge. In: Bonet-Carne E, Hutter J, Palombo M, Pizzolato M, Sepehrband F, Zhang F, eds. Computational Diffusion MRI. Cham, Switzerland: Springer International Publishing; 2020:195–208.

50. Hutter J, Slator PJ, Christiaens D, et al. Integrated and efficient diffusion-relaxometry using ZEBRA. Sci Rep. 2018;8:1–13. [PubMed: 29311619]

51. Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017. pp 1510–1519.

52. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. Z Med Phys. 2019;29:102–127. [PubMed: 30553609]

53. Ning L, Bonet-Carne E, Grussu F, et al. Muti-shell diffusion MRI harmonisation and enhancement challenge (MUSHAC): progress and results. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 2018. pp 217–224.
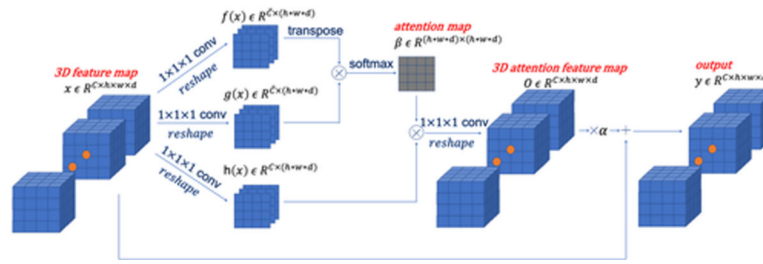
FIGURE 1.

Schematic view of the self-attention conditional generative adversarial network (SC-GAN). The first layer represents the input data. The attention map exploits the similarity of each pair of convolved images and combines it with the input data to generate the output of the self-attention module. Abbreviation: conv, convolution
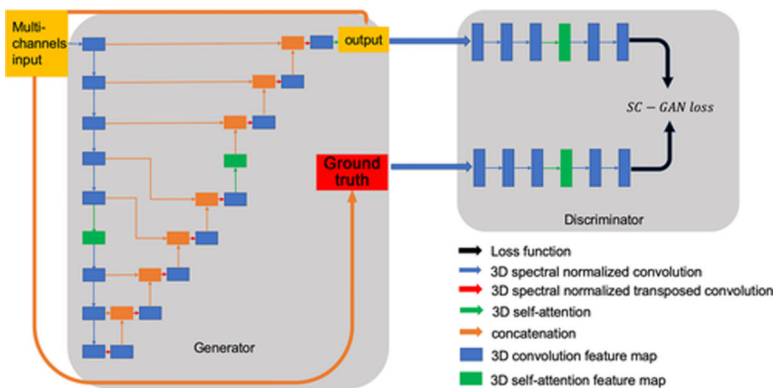
FIGURE 2.

The SC-GAN structure with 3D self-attention module. The network structure of SC-GAN consists of two parts: generator and discriminator. The generator is a Unet-like eight-layer encoder-decoder with a 3D self-attention module in the middle of the encoder and decoder. The discriminator is a five-layer patch GAN with 3D self-attention. The self-attention module empowers both generator and discriminator in the adversarial learning strategy
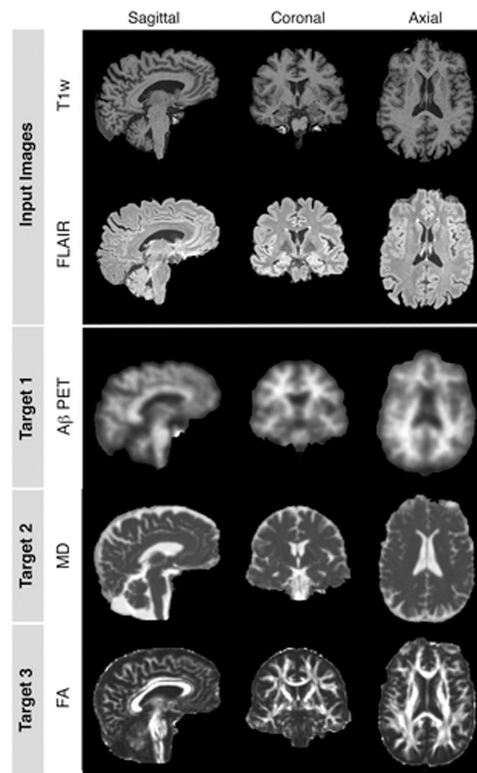
FIGURE 3.
Multimodal (multichannel) input. Examples of different neuroimaging data from single individual are presented. T_1-weighted (T1w) and fluid-attenuated inversion recovery (FLAIR) were used as input for different synthesis tasks. For each the study tasks, a different target was used, shown as outputs 1–3: mean diffusivity (MD), fractional anisotropy (FA), and amyloid-beta (Aβ) PET. Data were preprocessed and co-registered (see section 2 for details) and are shown from three anatomical views (from left to right: axial, coronal, and sagittal)
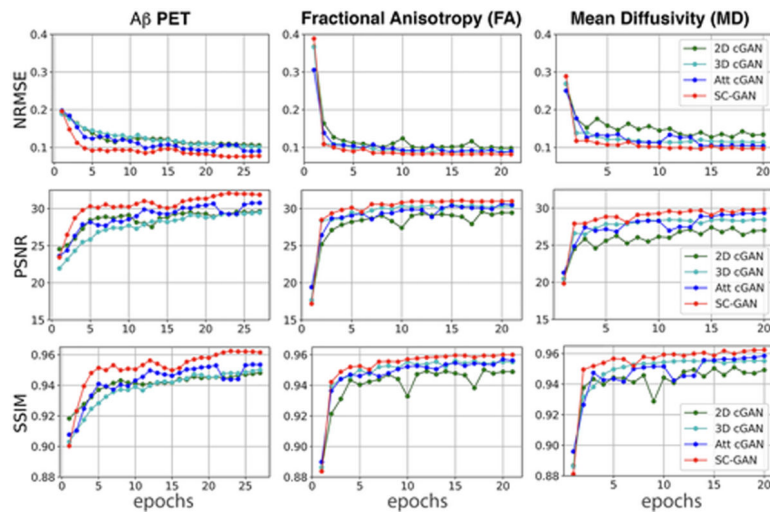
FIGURE 4.

Learning-curve SC-GANs compared with other synthesis GANs across different tasks. Plots demonstrate learning curves of four convolutional neural networks that were evaluated in this study: 2D GAN, 3D GAN, 3D conditional GAN with attention gate (Att cGAN), and SC-GAN. The T1w and FLAIR data were used for three tasks: (1) synthesizing AβPET (n = 242, first column); (2) synthesizing FA (n = 480, second column); and (3) synthesizing MD (n = 480, third column). Three different evaluation metrics were used: First row shows normalized RMS error (NRMSE); second row shows peak SNR (PSNR); and third row shows structural similarity (SSIM). Note that all networks reached their plateau around epoch 20
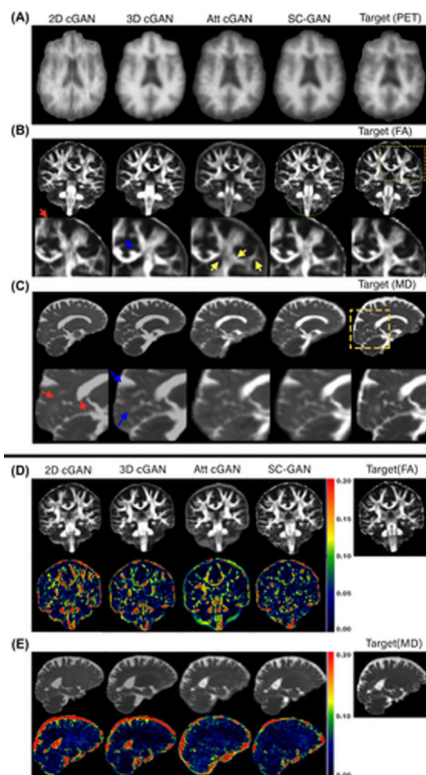
FIGURE 5.

Qualitative assessment of three tasks. Images are the result of applying different GANs on T1w and FLAIR input images to predict Aβ PET (A), FA (B), MD (C), and absolute value error maps between synthesis results and target for FA (D) and MD (E) tasks. Target PET/FA/MD are also illustrated for comparison. Target image is normalized to the [0 1] range for training, and an equal color range of [0 1] is used for visualization. Note that SC-GAN was able to synthesize the most similar results in comparison with other networks. In the FA task (B), a 2D network demonstrated continuous distortion (red arrow), and 3D cGAN resulted in an oversmoothed image (see blue arrow showing partial-volume effect between fiber bundles of cingulum and corpus callosum). Attention cGAN failed to capture high-intensity FA across the white matter (yellow arrows). Green dotted circle shows that, unlike other networks, SC-GAN was able to capture brainstem details. In the MD task (C), 2D generated artificial sharp boundaries (red arrow) and 3D cGAN resulted in a large amount of striping artifact (blue arrow). D, The absolute value error maps for FA synthesis task. E, The absolute error maps for MD synthesis task. High error rates at brain boundaries are related to the regions affected by the EPI distortion and caused by imperfect brain masks used in the experiment. The figure shows that the result generated by SC-GAN has the lowest intensity error compared with other networks' results
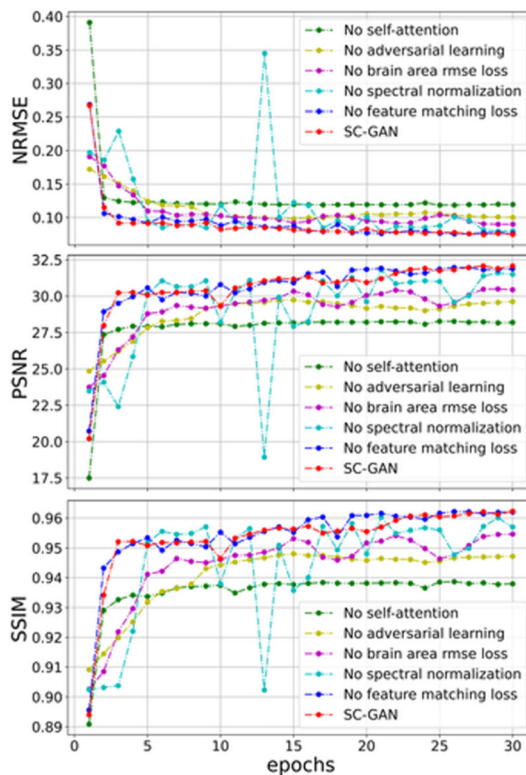
FIGURE 6.
Ablation study on test data across modules of SC-GAN. The SC-GAN with and without different network modules were assessed on the Aβ PET synthesis task, and learning curves across different evaluation criteria are presented here. Plots demonstrate NRMSE, PSNR, and SSIM. The self-attention module appeared to have the highest contribution to the achieved improvement, followed by spectral normalization and non–brain-loss function exclusion
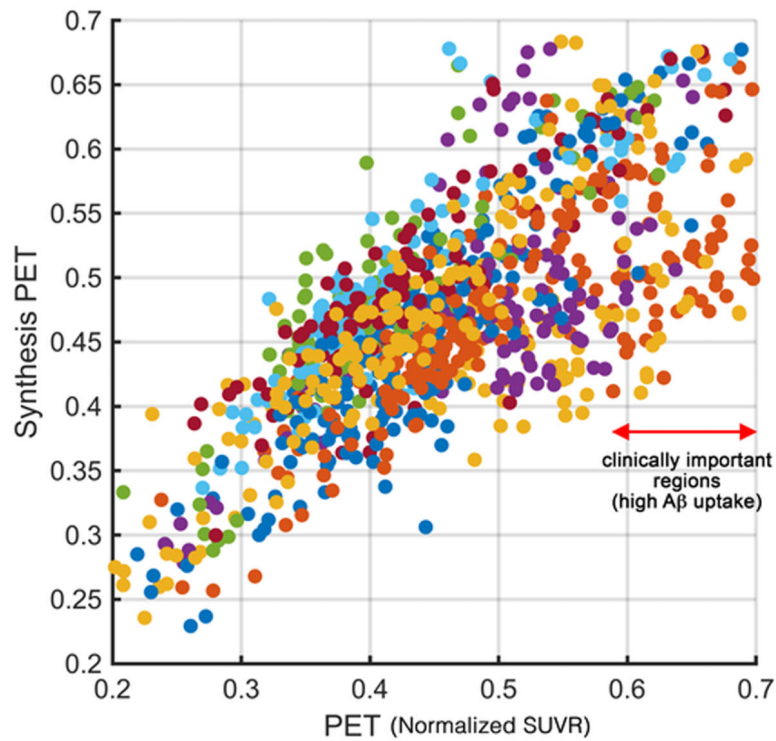
FIGURE 7.

Correlation between PET and synthesis PET. Plot shows the correlation between Aβ standard uptake value ratio (SUVR) across subcortical and cortical regions of 10 test participants (each color represents regions of each participants). The PET images that were used for training were normalized using the min-max normalization approach. Therefore, test PET images were also normalized using the same approach before comparison. Note that on the region with high load of Aβ (shown with red arrow), the synthesis error is higher, suggesting that synthesis PET could not substitute PET imaging
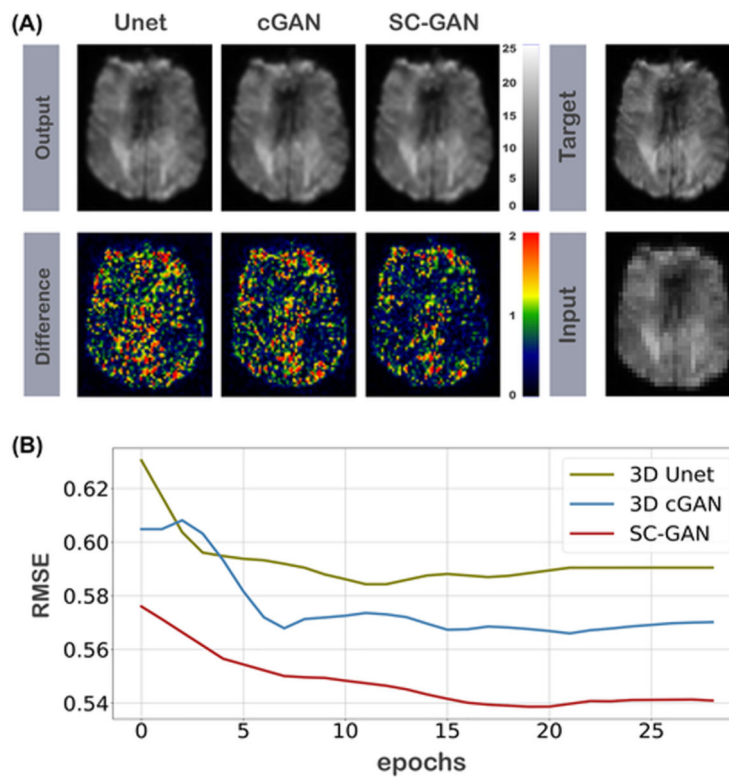
FIGURE 8.
Qualitative assessment of MUDI superresolution. A, Images show axial slices of outputs of 3D Unet, 3D cGAN and SC-GAN, and absolute difference map between output and target for each of them. Axial slice of input (size $41 \times 46 \times 28$) and target (size $82 \times 92 \times 56$) data are also presented here. B, Learning curves on the validation data of the second step of progressive training

TABLE 1

Comparison among different networks

Synthesis task (target image)	Method	NRMSE mean (± SD)	PSNR mean (± SD)	SSIM mean (± SD)
PET	2D cGAN	0.100 ± 0.028	29.80 ± 1.59	0.948 ± 0.010
	3D cGAN	0.099 ± 0.022	29.69 ± 1.96	0.950 ± 0.012
	Att cGAN	0.086 ± 0.024	31.03 ± 2.34	0.955 ± 0.014
	SC GAN	**0.076 ± 0.017**	**32.14 ± 1.10**	**0.962 ± 0.008**
FA	2D cGAN	0.100 ± 0.014	29.29 ± 1.23	0.948 ± 0.008
	3D cGAN	0.089 ± 0.015	30.39 ± 1.47	0.955 ± 0.008
	Att cGAN	0.086 ± 0.014	30.65 ± 1.41	0.956 ± 0.008
	SC GAN	**0.082 ± 0.013**	**31.00 ± 1.12**	**0.959 ± 0.007**
MD	2D cGAN	0.135 ± 0.019	26.98 ± 1.38	0.949 ± 0.010
	3D cGAN	0.121 ± 0.018	27.93 ± 1.42	0.953 ± 0.010
	Att cGAN	0.108 ± 0.014	28.74 ± 1.19	0.954 ± 0.009
	SC GAN	**0.096 ± 0.014**	**29.75 ± 1.25**	**0.963 ± 0.009**

Note: This table shows statistic values of NRMSE, PSNR, and SSIM among test images after the networks reached the plateau and the hyperparameters were optimized. Statistically significant results are highlighted in bold font.

TABLE 2

Ablation study of SC-GAN

Ablation study	NRMSE mean (± SD)	PSNR mean (± SD)	SSIM mean (± SD)
No self-attention	0.118 ± 0.016	28.34 ± 1.200	0.939 ± 0.011
No adversarial learning	0.102 ± 0.018	29.72 ± 1.583	0.947 ± 0.012
No brain-area RMS error loss	0.092 ± 0.017	30.27 ± 1.627	0.953 ± 0.010
No spectral normalization	0.080 ± 0.017	31.57 ± 1.203	0.957 ± 0.010
No feature matching	0.078 ± 0.019	32.03 ± 1.174	0.960 ± 0.013
SC-GAN	**0.076 ± 0.017**	**32.14 ± 1.100**	**0.962 ± 0.008**

Note: This table shows the ablation study of different components of SC-GAN on the $A\beta$ PET synthesis task.