



# Style-Based Manifold for Weakly-Supervised Disease Characteristic Discovery

Siyu Liu<sup>1(✉)</sup>, Linfeng Liu<sup>2</sup>, Craig Engstrom<sup>1</sup>, Xuan Vinh To<sup>2</sup>,  
Zongyuan Ge<sup>3,4,5</sup>, Stuart Crozier<sup>1</sup>, Fatima Nasrallah<sup>2</sup>,  
and Shekhar S. Chandra<sup>1</sup>

<sup>1</sup> School of Electrical Engineering and Computer Science, The University  
of Queensland, Brisbane, Australia  
siyu.liu1@uq.edu.au

<sup>2</sup> The Queensland Brain Institute, The University of Queensland, Brisbane, Australia

<sup>3</sup> Department of Data Science and AI, Monash University, Melbourne, Australia

<sup>4</sup> Monash-Airdoc Research Centre, Monash University, Melbourne, Australia

<sup>5</sup> AIM for Health Lab, Monash University, Melbourne, Australia

**Abstract.** In Alzheimer’s Disease (AD), interpreting tissue changes is key to discovering disease characteristics. However, AD-induced brain atrophy can be difficult to observe without Cognitively Normal (CN) reference images, and collecting co-registered AD and CN images at scale is not practical. We propose Disease Discovery GAN (DiDiGAN), a style-based network that can create representative reference images for disease characteristic discovery. DiDiGAN learns a manifold of disease-specific style codes. In the generator, these style codes are used to “stylize” an anatomical constraint into synthetic reference images (for various disease states). The constraint in this case underpins the high-level anatomical structure upon which disease features are synthesized. Additionally, DiDiGAN’s manifold is smooth such that seamless disease state transitions are possible via style interpolation. Finally, to ensure the generated reference images are anatomically correlated across disease states, we incorporate anti-aliasing inspired by StyleGAN3 to enforce anatomical correspondence. We test DiDiGAN on the ADNI dataset involving CN and AD magnetic resonance images (MRIs), and the generated reference AD and CN images reveal key AD characteristics (hippocampus shrinkage, ventricular enlargement, cortex atrophies). Moreover, by interpolating DiDiGAN’s manifold, smooth CN-AD transitions were acquired further enhancing disease visualization. In contrast, other methods in the literature lack such dedicated disease manifolds and fail to synthesize usable reference images for disease characteristic discovery.

**Keywords:** Disease learning · Weak supervision · Alzheimer’s Disease

**Supplementary Information** The online version contains supplementary material available at [https://doi.org/10.1007/978-3-031-43904-9\\_36](https://doi.org/10.1007/978-3-031-43904-9_36).

# 1 Introduction

Discovering disease characteristics from medical images can yield important insights into pathological changes. However, subtle disease characteristics are difficult to discover from imaging data. The reduction in brain volume due to Alzheimer’s Disease (AD) is one such example. In this case, comparing co-registered AD and Cognitively Normal (CN) reference magnetic resonance images (MRIs) is the most effective way to reveal subtle AD characteristics. Unfortunately, collecting such co-registered reference images at scale is not practical.

Generative Adversarial Networks (GANs) has the potential to synthesize the reference images required for disease discovery. Specifically, style-based generative frameworks [4] could be suitable as they employ mechanisms to “stylize” a share anatomical structure into different disease states. However, as we will show, many GAN frameworks (style and non-style based) are unable to produce satisfactory reference AD and CN images for AD characteristic discovery.

In this work, we propose Disease Discovery GAN (DiDiGAN), a specialized style-based network for disease characteristic discovery on medical image data. On the Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset [7], DiDiGAN can create not only reference AD and CN MRIs, but also smooth animated transitions between AD and CN. The highlights of DiDiGAN are:

- A learnt disease manifold that captures and encodes AD and CN disease state distributions. Style codes sampled from this manifold control the disease expression in the output reference images.
- The disease manifold is naturally smooth such that seamless transitions of AD to CN are possible via style interpolation.
- The generator uses a low-resolution input image as the source anatomical constraint (to provide coarse structural guidance), it is low-resolution to leave sufficient room for the generator to synthesize disease features.
- Anti-aliasing as a key mechanism to maintain anatomical correspondence across generated reference images. Without anti-aliasing, reference images exhibit inconsistent anatomical structures invalidating visual comparisons.
- DiDiGAN is a weakly-supervised framework requiring only class labels for images rather than pixel (voxel) labels.

DiDiGAN learns to generate representative reference AD and CN images by training on 2D coronal brain slices labelled with AD and CN. The generated reference images and animations clearly show systematic changes in the hippocampus, ventricle, and cortex areas while maintaining anatomical correspondence. The discovered characteristics are corroborated by independent tests which solidify the strengths of the findings. There has not been a previous work that can i) produce visualizations on par with DiDiGAN’s and ii) learn a dedicated manifold for disease characteristic discovery.

## 2 Related Work

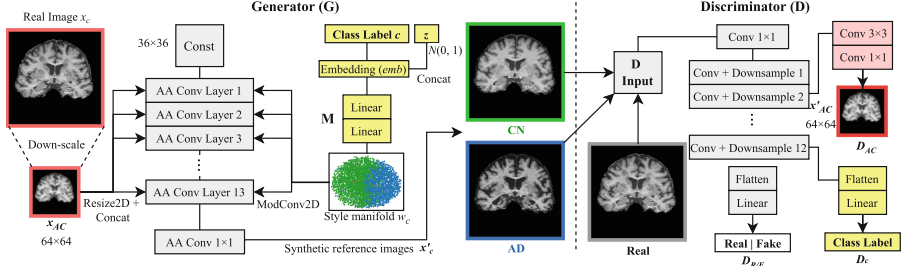
Various generative methods have attempted to synthesize pathological changes for disease modelling. HerstonNet [18] and Bernal et al. [2] synthesize brain MRIs with atrophy, but they rely on induced atrophy by altering the Partial Volume (PV)-maps and segmentation maps. Hence they do not learn disease characteristics on their own. Ravi et al. [17] and ADESyn [8] learn AD characteristics from data, and they can synthesize brain images at different stages of AD. However, ADESyn fails to capture ventricle enlargement which is a key AD while Ravi et al. failed to capture hippocampus shrinkage (which are key AD characteristics). DBCE [16] learns to deform a healthy brain image into an AD image. However, the resulting atrophy is not sufficiently distinct for AD characteristic visualization. Saliency maps must be further computed to reveal AD regions. Xia et al. [23] proposed a GAN for ageing brain synthesis. While the results clearly depict ventricle enlargement for advanced ages, other AD characteristics remain unclear as the images show significant anatomical inconsistencies.

Another limitation of the above methods is that disease features are intertwined with anatomical features. For disease modelling, a dedicated mechanism (such as a latent space) to store disease features in an explorable manner would be desirable. The objective to learn a dedicated disease latent space naturally aligns with style-based [4] frameworks, which have mostly been used for medical image translation [1, 11]. Style-based frameworks have the potential to enhance generative disease modelling. Fetty et al. [5] have shown that StyleGANs can learn a dedicated latent space (manifold) of medical image features. This latent space can be smoothly interpolated to manipulate synthesis. Other works like StarGAN [4] have shown that this latent can be disentangled from the content. Similarly, in medical image analysis, Chatsias et al. [3] have shown modality information can be disentangled from anatomical content for multi-modal image translation. These findings motivate DiDiGAN to use style to learn a manifold of disease features and then apply the manifold to “stylized” an input anatomical constraint into AD and CN images. However, StyleGAN3 [9] discovered that common style-based methods suffer from aliasing [24]. When interpolating the manifold, aliasing causes undesirable texture disruptions, and it threatens to disrupt anatomical structures if applied to disease studies. Thus, applying anti-aliasing (as per Shannon-Nyquist theorem) is critical for DiDiGAN.

## 3 Methods

DiDiGAN projects different disease states onto a common anatomical structure to form disease reference images, and the output images must maintain a consistent anatomical structure. A detailed architecture diagram is shown in Fig. 1.

**Disease Style  $w_c$  and Anatomical Constraint  $x_{AC}$ :** Let  $x_c$  denote the images in a dataset and  $c \in \{c_1, c_2 \dots c_n\}$  are the disease classifications for the images, the generator  $G$ ’s aim is to synthesise reference disease images  $x'_c$  that share a consistent anatomical structure. This is done by injecting disease styles  $w_c$  into



**Fig. 1.** Generator and discriminator architectures of DiDiGAN. Source image  $x_c$  is first down-sampled into anatomical constraint  $x_{AC}$ , then disease style codes  $w_c$  ( $c \in \{c_1, c_2 \dots c_n\}$ ) are injected into  $G$  to generate reference images  $x'_c$  for different disease states. All  $x'_c$ s maintain a consistent anatomical structure.

a shared anatomical constraint  $x_{AC}$  to produce the output  $x'_c = G(w_c, x_{AC})$ .  $w_c$  is created following  $w_c = M(emb(c), z)$  where  $emb$  is a learned 512-d embedding specific to each class,  $z$  is a 512-d  $N(0, 1)$  noise vector and  $M$  is a 2-layer mapping network made up of dense layers. Like StyleGAN, the style code space  $w_c$  is a learned manifold (of disease states), and it can be interpolated for smooth disease states transitions. The reason to include  $z$  is to introduce variability such that  $w_c$  forms a surface instead of a fixed point.  $x_{AC}$  is a  $4 \times$  down-sampled version of  $x_c$  which only enforces coarse anatomical structure similarities between the reference images, and it lacks resolution on purpose allowing room for disease features from  $w_c$  to manifest. As already discussed, full anatomical correspondence requires anti-aliasing to enforce.

**Alias-Free Style Generator  $G$ :** Generator  $G$  is alias-free as the feature maps are processed carefully following the Shannon-Nyquist sampling theorem. Like StyleGAN 3,  $G$ 's convolutional (AA Conv) layers apply anti-aliasing to the leaky-reLU activation function. This process involves first interpolating discrete feature maps to continuous domains. Then, low-pass filtering is used to remove offending frequencies. Finally, the activation function is applied, and the signal is re-sampled back to the discrete domain [9]. Architecture wise  $G$  is a progressively up-sampling decoder Convolutional Neural Network (CNN) with a starting resolution of  $4 \times 4$  and an output resolution of  $256 \times 256$ . There are 13 convolutional blocks that gradually perform upsampling and synthesis. All the convolutional layers are modulated convolutions [4] to inject disease features  $w_c$ . The anatomical constraint  $x_{AC}$  is introduced by concatenating it to the input of every block, and we observed more stable training using this method.

**Multi-head Discriminator  $D$ :** The discriminator  $D$  is an encoder architecture with consecutive convolution and down-sampling. There are three prediction heads  $D_{R/F}$ ,  $D_c$ , and  $D_{AC}$  which predict a realism logit, disease classification, and reconstruction of constraint  $x_{AC}$ , respectively. The main loss functions are these outputs are 1) the standard non-saturating GAN loss  $L_{R/F}$  for generating realistic images, 2) the adversarial classification loss  $L_c$  which supervises the

conditioning on  $c$  3) the anatomical reconstruction loss  $L_{AC}$  ensuring the generated reference images share the high-level structure of the input anatomical constraint. These losses are as follows:

$$\min_D \max_G L_{R/F} = \mathbb{E}[\log D_{R/F}(x_c)] - \mathbb{E}[\log(D_{R/F}(x'_c))] \quad (1)$$

$$\min_D \max_G L_c = - \sum_{c \in \{c_1, c_2, \dots, c_n\}} y_c \log(D_c(X)), \text{ where } X \in \{x'_c, x_c\} \quad (2)$$

$$L_{AC} = \|x_{AC} - D_{AC}(G(w_c, x_{AC}))\|^2 + \|x_{AC} - D_{AC}(x_c)\|^2 \quad (3)$$

where  $x_c$  is a real image of class  $c$ ,  $x'_c$  is a fake image of the same class,  $w_c = M(\text{emb}(c), z)$ , and  $x_{AC}$  is a down-sampled version of  $x_c$ .

There are also other standard regularization loss functions including the path-length regularization  $L_{pl}$  [9],  $R1$  gradient penalty  $L_{R1}$  [14] and an explicit diversification loss  $L_{div} = -\|G(w, c, z_1) - G(w, c, z_2)\|$  to prevent the manifold from converging to fixed points. The total loss  $L_{total}$  is then defined as

$$L_{total} = L_{R/F} + L_c + L_{AC} + L_{div} + L_{pl} + L_{R1} \quad (4)$$

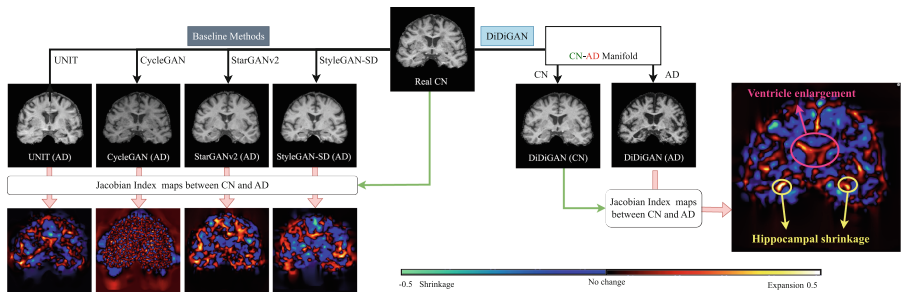
## 4 Experiment, Results and Discussion

**Data Preparation and Training.** DiDiGAN is applied to the ADNI dataset to visualize MRIs pathological changes from CN to AD. The dataset consists of T1-weighted 3D MRIs from the ADNI dataset totalling 1565 scans (680 patients) including 459 CN patients and 221 AD patients. The scans are divided into training, validation, and test sets at the patient level with a ratio of 0.7, 0.2, and 0.1. Standard preprocessing procedures (N4 bias-field correction [21], skull stripping using SPM [22]) were applied to the entire dataset. For each 3D scan, the center 40 coronal slices were extracted and zero-padded to  $256 \times 256$  pixels. Down-sampled versions of these slices ( $64 \times 64$ ) were used for  $X_{AC}$ . The network was implemented in Pytorch and trained for 60,000 steps (32 images per step) using Adam optimizer. The training hardware is a 32 GB Nvidia V100 graphics card to support a batch size of 32. The baseline methods include a range of style and non-style-based image translation methods. These methods can readily treat reference image generation as unpaired image translation where, for example, a CN brain is translated to an AD version of the same brain. All the methods were trained and evaluated on the exact same data and splits.

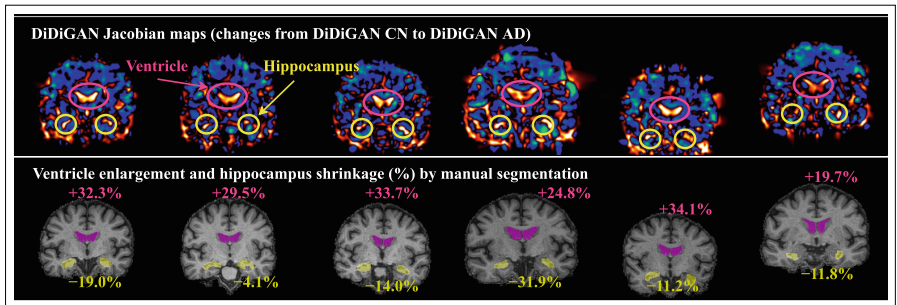
**Quantitative Image Quality Evaluation Using Perceptual Metrics.** The quality of the generated reference images (based on the test set, ignoring classes) from all the methods is assessed using Frechet Inception Distance (FID), Learned Perceptual Image Patch Similarity (LPIPS), and Kernel Inception Distance (KID) as they do not require paired data. As Table 1 shows, DiDiGAN generated the most realistic reference images compared to the baselines. MUNIT [6] and CUT [15] failed to reach convergence hence the metrics are omitted.

**Table 1.** Quality of reference images generated by DiDiGAN and the baselines measured using three popular perceptual metrics. All the baselines were run on the exact same data as DiDiGAN. DiDiGAN achieved the best image quality while CUT and MUNIT failed to converge despite all the training instructions being followed.

Method	FID↓	LPIPS↑	KID↓
CUT [15]	N/A	N/A	N/A
MUNIT [6]	N/A	N/A	N/A
CycleGAN [25]	32.52	0.04	0.041
UNIT [12]	34.93	0.01	0.039
StyleGAN-SD [10]	17.23	0.14	0.034
StarGANv2 [4]	17.82	0.07	0.037
<b>DiDiGAN</b>	<b>14.48</b>	<b>0.29</b>	<b>0.031</b>



**Fig. 2.** AD feature discovery results for DiDiGAN and the baselines. The baselines perform image translation from an input CN to a reference AD image while DiDiGAN uses its manifold to generate a pair of reference AD and CN images. Jacobian Index heat maps were computed to show shrinkage (cool) or expansion (hot) areas between the AD and CN images. DiDiGAN’s reference images reveal key AD characteristics including hippocampus shrinkage and ventricle enlargement. The baselines failed to highlight these characteristics due to anatomical disruptions.



**Fig. 3.** More example Jacobian maps for DiDiGAN’s reference AD and CN pairs. The hippocampus and ventricle areas are manually segmented with size changes labelled.

### AD Characteristic Extraction and Visualization Using DiDiGAN.

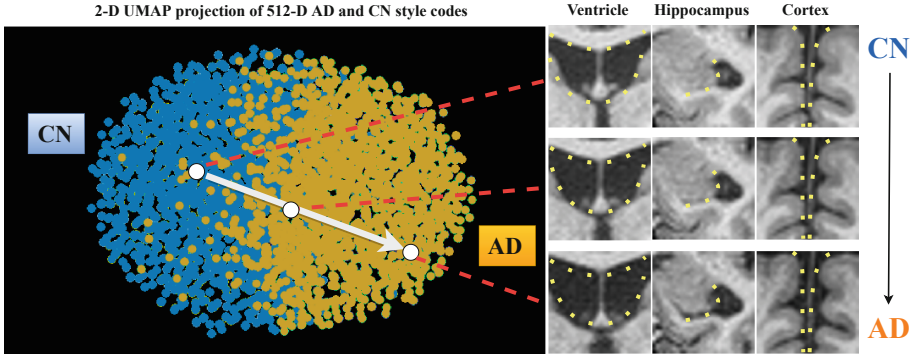
Figure 2 provides AD visualizations by generating reference images with DiDiGAN and the baseline methods. Since DiDiGAN’s uses stylization, it can freely generate correlated AD and CN pairs using the manifold. Comparison between the reference pairs reveals AD characteristics [19,20] such as hippocampus shrinkage, ventricle enlargement, and the thinning of cortical structures. The consistent anatomical structure also facilitates the computation of Jacobian (shrinkage and expansion) maps. For DiDiGAN, the bright regions in DiDiGAN’s Jacobian map more clearly highlight significant hippocampus shrinkage and ventricle enlargement from CN to AD. More examples are shown in Fig. 3. In these examples, we also manually segmented the ventricle and hippocampus, and the notable size changes marked in the figure. Anti-aliasing is crucial for maintaining anatomical correspondence across DiDiGAN’s reference images. As an ablation study, a DiDiGAN without anti-aliasing was trained. The generated AD and CN pairs exhibit significant anatomical disturbances (see S1 for examples).

**Comparing AD Characteristic Visualization with Baselines.** The baseline methods produce reference images via direct image translation from CN to an AD reference. The results in Fig. 2 suggest StarGANv2 [4], StyleGAN-SD [10] and UNIT [12] were unable to convey useful AD findings as they struggled to maintain anatomical correspondence. CycleGAN’s [25] simpler architecture helped avoid anatomical disruptions. However, the generated AD is almost pixel-identical to the input CN aside from the contrast difference (hence the high FID). The other two methods MUNIT [6] and CUT [15] could not reach convergence on the ADNI dataset despite the documentation being carefully followed.

**Disease Manifold Formation and Interpolation.** DiDiGAN’s disease manifold formation is visualized using UMAP [13] where style codes sampled from the manifold are reduced to 2-d from 512-d. Figure 4 shows a 2D cluster containing 10,000 AD and 10,000 CN style codes. While the input disease class is a discrete value, the manifold automatically maps it to a disease distribution with AD distributed on the left and CN on the right. The overlap between the two classes suggests AD is a progressive disease. The smoothness of the manifold facilitates smooth CN-AD transition animations by interpolating between an AD style code and a CN style code. Practically, this animation provides a more intuitive visualization of AD pathological changes.

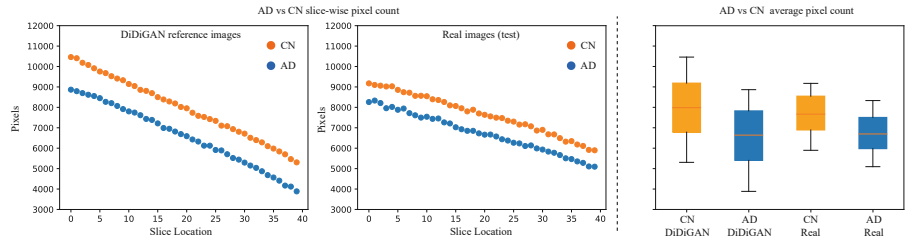
**Corroborating Tests:** Due to the lack of paired medical images as ground truths, we perform the following tests (I, II, III) to verify DiDiGAN’s findings. **I:** DiDiGAN was independently applied to the sagittal view slices, and similar hippocampus shrinkage and ventricle enlargement were clearly observed (see examples S2). These independent findings corroborate the coronal view visualizations produced by DiDiGAN’s reference images.





**Fig. 4.** 2D UMAP of 10000 CN and 10000 AD style codes sampled from the disease manifold. A linear path from CN to AD is traversed to generate an CN-AD animation. Changes in key affected regions (with reference points) are shown on the right. To help with visualization, we provide a video of several example animations.

**II:** SPM segmentation was applied to the test set and DiDiGAN’s reference images (generated based on the test set) to help compare the magnitudes of brain volume loss. For each image, the grey matter and white matter pixels were treated as brain mass. As Fig. 5 shows, DiDiGAN’s AD reference images consistently show systematic brain mass reduction across all 40 slice positions. This trend is consistent with that of the real data. The box plot suggests an average brain mass reduction of 17.3% compared to the 12.1% of the real data.



**Fig. 5.** Slice-wise brain mass difference (in pixels) between AD and CN as represented by DiDiGAN and real data. AD shows significant brain mass reductions in both cases.

**III:** After training DiDiGAN on ADNI for disease discovery, we fine-tuned  $D$ ’s classification head  $D_c$  for binary AD-CN classification and three-way AD-Mild Cognitive Impairment (MCI)-CN classification on the same dataset and splits. The test accuracies were 93.2% and 90.4% which are competitive with comparable methods in the literature (S3 shows a full comparison). Since  $D_c$  back-propagates disease class signals into  $G$ . This classification result suggests DiDiGAN successfully learned useful AD features.



**Limitations.** DiDiGAN is an initial step for style-based disease characteristic discovery. In the future, a more rigorous examination of the manifold is needed to fully understand the features and trends learned especially for clinical applications. For example, longitudinal trends and atrophy patterns among the AD population. Although DiDiGAN discovered brain tissue shrinkage as indicated by the SPM segmentation analysis, the learned magnitude is different from that of the real data. This is likely a limitation of GANs as the learnt distribution may not exactly match the real data in all aspects. Nonetheless, DiDiGAN's findings could serve as disease characteristic proposals. Additionally, more datasets and diseases should be tested to more thoroughly assess DiDiGAN's generalizability.

## 5 Conclusion

DiDiGAN demonstrated disease characteristic discovery by generating reference images that clearly depict relevant pathological features. The main technical novelties of DiDiGAN are i) the use of a learned disease manifold to manipulate disease states AD ii) the ability to interpolate the manifold to enhance visualization and iii) mechanisms including the structural constraint and anti-aliasing to maintain anatomical correspondence without direct registration. In the experiments involving the ADNI dataset, DiDiGAN discovered key AD features such as hippocampus shrinkage, ventricular enlargement, and cortex atrophy where other frameworks failed. DiDiGAN shows potential to aid disease characteristic discovery across time of other chronic diseases such as osteoarthritis.

## References

1. Armanious, K., et al.: MedGAN: medical image translation using GANs. *Comput. Med. Imaging Graph.* **79**, 101684 (2020). <https://doi.org/10.1016/j.compmedimag.2019.101684>
2. Bernal, J., Valverde, S., Kushibar, K., Cabezas, M., Oliver, A., Lladó, X.: Generating longitudinal atrophy evaluation datasets on brain magnetic resonance images using convolutional neural networks and segmentation priors. *Neuroinformatics* **19**(3), 477–492 (2021). <https://doi.org/10.1007/s12021-020-09499-z>
3. Chatsias, A., et al.: Disentangled representation learning in cardiac image analysis. *Med. Image Anal.* **58**, 101535 (2019). <https://doi.org/10.1016/j.media.2019.101535>
4. Choi, Y., Uh, Y., Yoo, J., Ha, J.W.: StarGAN v2: diverse image synthesis for multiple domains. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, June 2020. <https://doi.org/10.1109/cvpr42600.2020.00821>
5. Fetty, L., et al.: Latent space manipulation for high-resolution medical image synthesis via the StyleGAN. *Z. Med. Phys.* **30**(4), 305–314 (2020). <https://doi.org/10.1016/j.zemedi.2020.05.001>
6. Huang, X., Liu, M.-Y., Belongie, S., Kautz, J.: Multimodal unsupervised image-to-image translation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) *ECCV 2018*. LNCS, vol. 11207, pp. 179–196. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-01219-9\\_11](https://doi.org/10.1007/978-3-030-01219-9_11)

7. Jack, C.R., et al.: The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging* **27**(4), 685–691 (2008). <https://doi.org/10.1002/jmri.21049>
8. Jung, E., Luna, M., Park, S.H.: Conditional GAN with an attention-based generator and a 3D discriminator for 3D medical image generation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12906, pp. 318–328. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-87231-1\\_31](https://doi.org/10.1007/978-3-030-87231-1_31)
9. Karras, T., et al.: Alias-free generative adversarial networks. In: Proceedings NeurIPS (2021)
10. Kim, K.L., Park, S.Y., Jeon, E., Kim, T.H., Kim, D.: A style-aware discriminator for controllable image translation. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 18218–18227 (2022)
11. Liu, M., et al.: Style transfer using generative adversarial networks for multi-site MRI harmonization. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12903, pp. 313–322. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-87199-4\\_30](https://doi.org/10.1007/978-3-030-87199-4_30)
12. Liu, M.Y., Breuel, T., Kautz, J.: Unsupervised image-to-image translation networks. In: Guyon, I., et al. (eds.) Advances in Neural Information Processing Systems, vol. 30. Curran Associates, Inc. (2017)
13. McInnes, L., Healy, J., Saul, N., Grossberger, L.: UMAP: uniform manifold approximation and projection. *J. Open Source Softw.* **3**(29), 861 (2018)
14. Mescheder, L.M., Geiger, A., Nowozin, S.: Which training methods for GANs do actually converge? In: ICML (2018)
15. Park, T., Efros, A.A., Zhang, R., Zhu, J.-Y.: Contrastive learning for unpaired image-to-image translation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12354, pp. 319–345. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-58545-7\\_19](https://doi.org/10.1007/978-3-030-58545-7_19)
16. Peters, J., et al.: DBCE: a saliency method for medical deep learning through anatomically-consistent free-form deformations. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 1959–1969, January 2023. <https://doi.org/10.1109/WACV56688.2023.00200>
17. Ravi, D., Blumberg, S.B., Ingala, S., Barkhof, F., Alexander, D.C., Oxtoby, N.P.: Degenerative adversarial neuroimage nets for brain scan simulations: application in ageing and dementia, vol. 75, p. 102257. Elsevier BV, January 2022. <https://doi.org/10.1016/j.media.2021.102257>
18. Rusak, F., et al.: Quantifiable brain atrophy synthesis for benchmarking of cortical thickness estimation methods. *Med. Image Anal.* **82**, 102576 (2022). <https://doi.org/10.1016/j.media.2022.102576>
19. Sabuncu, M.R., et al.: The dynamics of cortical and hippocampal atrophy in Alzheimer disease. *Arch. Neurol.* **68**(8), 1040–1048 (2011). <https://doi.org/10.1001/archneurol.2011.167>
20. Schuff, N., et al.: MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. *Brain* **132**(4), 1067–1077 (2008). <https://doi.org/10.1093/brain/awp007>
21. Tustison, N.J., et al.: N4ITK: improved N3 bias correction **29**(6), 1310–1320 (2010). <https://doi.org/10.1109/tmi.2010.2046908>
22. Tzourio-Mazoyer, N., et al.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**(1), 273–289 (2002). <https://doi.org/10.1006/nimg.2001.0978>

23. Xia, T., Chartsias, A., Tsaftaris, S.A.: Consistent brain ageing synthesis. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11767, pp. 750–758. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-32251-9\\_82](https://doi.org/10.1007/978-3-030-32251-9_82)
24. Zhang, R.: Making convolutional networks shift-invariant again. In: ICML (2019)
25. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, October 2017. <https://doi.org/10.1109/iccv.2017.244>