Generating Realistic Brain MRIs via a Conditional Diffusion Probabilistic Model

Wei Peng[1], Ehsan Adeli[1], Tomas Bosschieter[1], Sang Hyun Park[2],
Qingyu Zhao[1], and Kilian M. Pohl[1(✉)]

[1] Stanford University, Stanford, CA 94305, USA
kpohl@stanford.edu
[2] Daegu Gyeongbuk Institute of Science and Technology, Daegu, South Korea

Abstract. As acquiring MRIs is expensive, neuroscience studies struggle to attain a sufficient number of them for properly training deep learning models. This challenge could be reduced by MRI synthesis, for which Generative Adversarial Networks (GANs) are popular. GANs, however, are commonly unstable and struggle with creating diverse and high-quality data. A more stable alternative is Diffusion Probabilistic Models (DPMs) with a fine-grained training strategy. To overcome their need for extensive computational resources, we propose a conditional DPM (cDPM) with a memory-efficient process that generates realistic-looking brain MRIs. To this end, we train a 2D cDPM to generate an MRI subvolume conditioned on another subset of slices from the same MRI. By generating slices using arbitrary combinations between condition and target slices, the model only requires limited computational resources to learn interdependencies between slices even if they are spatially far apart. After having learned these dependencies via an attention network, a new anatomy-consistent 3D brain MRI is generated by repeatedly applying the cDPM. Our experiments demonstrate that our method can generate high-quality 3D MRIs that share a similar distribution to real MRIs while still diversifying the training set. The code is available at https://github.com/xiaoiker/mask3DMRI_diffusion and also will be released as part of MONAI, at https://github.com/Project-MONAI/GenerativeModels.

1 Introduction

The synthesis of medical images has great potential in aiding tasks like improving image quality, imputing missing modalities [30], performing counterfactual analysis [17], and modeling disease progression [9,10,29]. However, synthesizing brain MRIs is non-trivial as they are of high dimension, yet the training data are relatively small in size (compared to 2D natural images). High-quality synthetic MRIs have been produced by conditional models based on real MRI of the same subject acquired with different MRI sequences [2,16,21,27]. However, such models require large data sets (which are difficult to get) and fail to significantly improve data diversity [11,26], i.e., producing MRIs substantially

deviating from those in the training data; data diversity is essential to the generalizability of large-scale models [26]. Unconditional models based on Generative Adversarial Networks (GANs) bypass this drawback by generating new, independent MRIs from random noise [1,7]. However, these models often produce lower quality MRIs as they currently can only be trained on lower resolution MRIs or 2D slices due to their computational needs [6,12]. Furthermore, GAN-based models are known to be unstable during training and even can suffer from mode collapse [4]. An alternative is diffusion probabilistic models (DPMs) [8,22], which formulate the fine-grained mapping between data distribution and Gaussian noise as a gradual process modeled within a Markov chain. Due to their multi-step, fine-grained training strategy, DPMs tend to be more stable during training than GANs and therefore are more accurate for certain medical imaging applications, such as segmentation and anomaly detection [13,25]. However, DPMs tend to be computationally too expensive to synthesize brain MRI at full image resolution [3,18]. We address this issue by proposing a memory-efficient 2D conditional DPM (cDPM) that relies on learning the interdependencies between 2D slices to produce high-quality 3D MRI volumes.
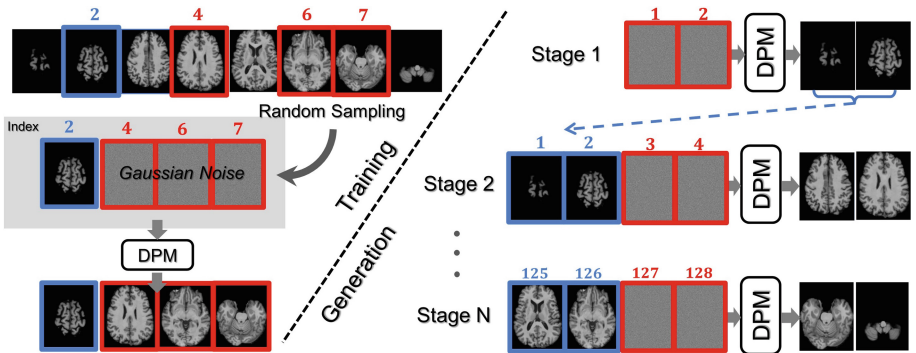
Fig. 1. A memory efficient DPM. **Left:** Based on 'condition' slices, cDPM learns to generate 'target' slices. **Right:** A new 3D MRI is created by repeatedly running the trained model to synthesize target slices conditioned on those it created in prior stages.

Unlike the sequence of 2D images defining a video, all 2D slices of an MRI are interconnected with each other as they define a 3D volume capturing brain anatomy. Our cDPM learns these interdependencies (even between distant slices) by training an attention network [24] on arbitrary combinations of condition and target slices. Once learned, the cDPM creates new samples while capturing brain anatomy in 3D. It does so by producing the first few slices from random noise and then using those slices to synthesize subsequent ones (see Fig. 1). We show that this computationally efficient conditional DPM can produce MRIs that are more realistic than those produced by GAN-based architectures. Furthermore, our experiments reveal that cDPM is able to generate synthetic MRIs, whose distribution matches that of the training data.

2   Methodology

We first review the basic DPM framework for data generation (Sect. 2.1). Then, we introduce our efficient strategy for generating 3D MRI slices (Sect. 2.2) and finally describe the neural architecture of cDPMs (Sect. 2.3).

2.1   Diffusion Probabilistic Model

The Diffusion Probabilistic Model (DPM) [8,22] generates MRIs from random noise by iterating between mapping 1) data gradually to noise (a.k.a., Forward Diffusion Process) and 2) noise back to data (a.k.a., Reverse Diffusion Process).

Forward Diffusion Process (FDP). Let real data $\mathbf{X_0} \sim \mathbf{q}$ sampled from the (real data) distribution q be the input to the FDP. FDP then simulates the diffusion process that turns X_0 after T perturbations into Gaussian noise $X_T \sim \mathcal{N}(0, I)$, where $\mathcal{N}$ is the Gaussian distribution with zero mean and the variance being the identity matrix I. This process is formulated as a Markov chain, whose transition kernel $q(X_t|X_{t-1})$ at time step $t \in \{0, \ldots, T\}$ is defined as

$$q(X_t|X_{t-1}) := \mathcal{N}(X_t; \sqrt{1 - \beta_t} \cdot X_{t-1}, \beta_t \cdot I). \tag{1}$$

The weight $\beta_t \in (0, 1)$ is changed so that the chain gradually enforces drift, i.e., adds Gaussian noise to the data. Let $\alpha_t := 1 - \beta_t$ and $\bar{\alpha}_t := \prod_{s=1}^{t}(1 - \beta_t)$, then X_t is a sample of the distribution conditioned on X_0 as

$$q(X_t|X_0) := \mathcal{N}(X_t; \sqrt{\bar{\alpha}_t} \cdot X_0, (1 - \bar{\alpha}_t) \cdot I). \tag{2}$$

Given this closed-form solution, we can sample X_t at any arbitrary time step t without needing to iterate through the entire Markov chain.

Reverse Diffusion Process (RDP). The RDP aims to generate realistic data from random noise X_T by approximating the posterior distribution $p(X_{t-1}|X_t)$. It does so by going through the entire Markov chain from time step T to 0, i.e.,

$$p(X_{0:T}) := p(X_T) \prod_{t=1}^{T} p_\theta(X_{t-1}|X_t). \tag{3}$$

Defining the conditional distribution $p_\theta(X_{t-1}|X_t) := \mathcal{N}(X_{t-1}; \mu_\theta(X_t, t), \Sigma)$ with fixed variance Σ, then (according to [8]) the mean can be rewritten as

$$\mu_\theta(X_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(X_t - \frac{\beta_t}{\sqrt{(1 - \bar{\alpha}_t)}} \epsilon_\theta(X_t, t) \right), \tag{4}$$

with $\epsilon_\theta(\cdot)$ being the estimate of a neural network defined by parameters θ. θ minimizes the reconstructing loss defined by the following expected value

$$\mathbb{E}_{\mathbf{X_0} \sim \mathbf{q}, \mathbf{t} \in [\mathbf{0}, \ldots, \mathbf{T}], \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} \left[||\epsilon - \epsilon_\theta(X_t, t)||_2^2 \right],$$

where $|| \cdot ||_2$ is the L2 norm, and X_t is inferred from Eq. (2) based on X_0.

2.2   Conditional Generation with DPM (cDPM)

To synthetically create high-resolution 3D MRI, we propose an efficient cDPM model that learns the interdependencies between 2D slices of an MRI so that it can generate slices based on another set of already synthesized ones (see Fig. 1).
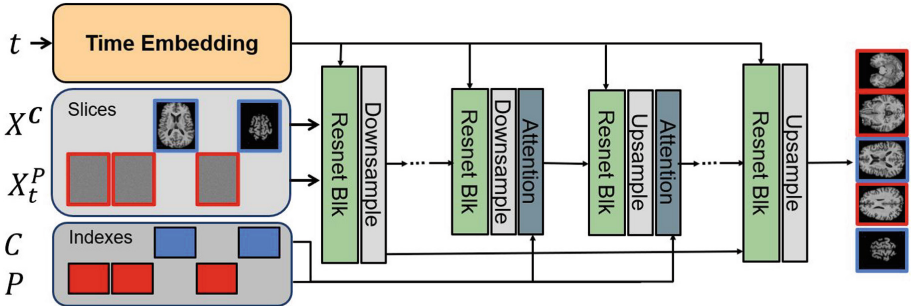
Specifically, given an MRI $X \in \mathbb{R}^{D \times H \times W}$, we randomly sample two sets of slice indexes: the condition set $\mathcal{C}$ and the target set $\mathcal{P}$. Let len($\cdot$) be the number of slices in a set, then the 'condition' slices are defined as $X^{\mathcal{C}} \in \mathbb{R}^{\text{len}(\mathcal{C}) \times H \times W}$ and the 'target' slices as $X^{\mathcal{P}} \in \mathbb{R}^{\text{len}(\mathcal{P}) \times H \times W}$ with len($\mathcal{P}$) $\geq$ 1. Confining the FDP of Sect. 2.1 just to the target $X^{\mathcal{P}}$, the RDP now aims to reconstruct $X_t^{\mathcal{P}}$ for each time $t = T, T-1, \ldots, 0$ starting from random noise at $t = T$ and conditioned on $X^{\mathcal{C}}$. Let $\widetilde{X}_t$ be the subvolume consisting of $X_t^{\mathcal{P}}$ and $X^{\mathcal{C}}$, then the joint distribution of the Markov chain defined by Eq. (3) now reads

$$p(X_{0:T}^{\mathcal{P}}) := p(X_T^{\mathcal{P}}) \prod_{t=1}^{T} p_\theta(X_{t-1}^{\mathcal{P}} | \widetilde{X}_t). \tag{5}$$

Observe that Eq. (5) is equal to Eq. (3) in case len($\mathcal{C}$) = 0.

To estimate $\mu_\theta(\widetilde{X}_t, t)$ as described in Eq. (4), we sample arbitrary index sets $\mathcal{C}$ and $\mathcal{P}$ so that len($\mathcal{C}$) + len($\mathcal{P}$) $\leq \tau_{\max}$, where $\tau_{\max}$ is the maximum number of slices based on the available resources. We then capture the dependencies across slices by feeding the index sets $\mathcal{C}$ and $\mathcal{P}$ and the corresponding slices (i.e., $X^{\mathcal{C}}$ and $X_t^{\mathcal{P}}$ built from $X_0 \sim q$) into an attention network [20]. The neural network aims to minimize the canonical loss function

$$\text{Loss}(\theta) := \mathbb{E}_{\mathbf{X_0} \sim \mathbf{q}, \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \mathcal{C} + \mathcal{P} \leq \tau_{\max}, \mathbf{t}} \left[||\epsilon - \epsilon_\theta(X_t^{\mathcal{P}}, X^{\mathcal{C}}, \mathcal{C}, \mathcal{P}, t)||_2^2 \right]. \tag{6}$$

Fig. 2. The architecture of cDPM is a U-shape neural network with skip connections and the input at step 't' are slice indexes $\{\mathcal{C}, \mathcal{P}\}$, condition sub-volume $X^{\mathcal{C}}$, and current target sub-volume $X_t^{\mathcal{P}}$.

As the neural network can now be trained on many different (arbitrary) slice combinations (defined by $\mathcal{C}$ and $\mathcal{P}$), the cDPM only requires a relatively

small number of MRIs for training. Furthermore, it will learn short- and long-range dependencies across slices as the spatial distance between slices from $\mathcal{C}$ and $\mathcal{P}$ varies. Learning these dependencies (after being trained for a sufficiently large number of iterations) enables cDPMs to produce 2D slices that, when put together, result in realistic looking, high-resolution 3D MRIs.

2.3   Network Architecture

As done by [8], cDPMs are implemented as a U-Net [19] with a time embedding module (see Fig. 2). We add a multi-head self-attention mechanism [24] to model the relationship between slices. After training the cDPM as in Fig. 1, a 3D MRI is generated in N stages. Specifically, the cDPM produces the initial set of slices of that MRI volume from random noise (i.e., unconditioned). Conditioned on those synthetic slices, the cDPM then runs again to produce a new set of slices. The process of synthetically creating slices based on ones generated during prior stages is repeated until an entire 3D MRI is produced.

3   Experiments

3.1   Data

We use 1262 t1-weighted brain MRIs of subjects from three different datasets: the Alzheimer's Disease Neuroimaging Initiative (ADNI-1), UCSF (PI: V. Valcour), and SRI International (PI: E.V. Sullivan and A. Pfefferbaum) [28]. Processing includes denoising, bias field correction, skull stripping, and affine registration to a template, and normalizing intensity values between 0 and 1. In addition, we padded and resized the MRIs to have dimensions $128 \times 128 \times 128$ resulting in a voxel resolution of $1.375\,mm \times 1.375\,mm \times 1.0\,mm$. Splitting the MRI along the axial direction results in 2D slices. Note, this could have also been done along the sagittal or coronal direction.

3.2   Implementation Details

Our experiments are conducted on an NVIDIA A100 GPU using the PyTorch framework. The model is trained using 200,000 iterations with the AdamW optimizer adopting a learning rate of 10^{-4} and a batch size of 3. $\tau_{\max}$ is set to 20. After the training, cDPM generates a synthetic MRI consisting of 128 slices by following the process outlined in Fig. 1 in N=13 stages. Each stage generates 10 slices starting with pure noise ($X^{\mathcal{C}} = \emptyset$) and (after the first stage) being conditioned on the 10 slices produced by the prior stage. After training on all real MRIs, we use the resulting conditional DPM to generate 500 synthetic MRIs.
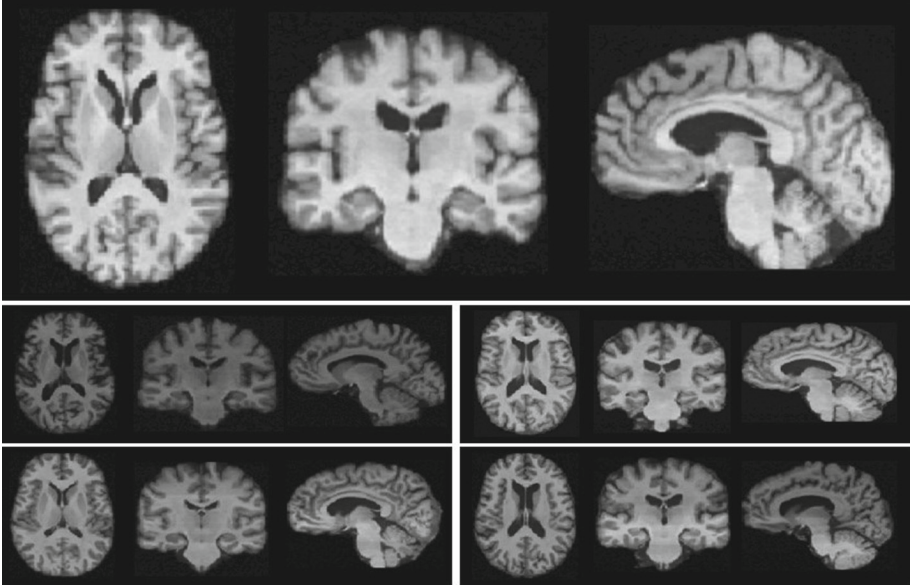
Fig. 3. 5 MRIs generated by our conditional DPM visualized in the axial, coronal, and sagittal plane. The example in the first row is enlarged to highlight the high quality of synthetic MRIs generated by our approach.

HA-GAN   CCE-GAN   3D-GAN-GP 3D-VAE-GAN 3D-α-WGAN   3D-DPM   **Ours(cDPM)**
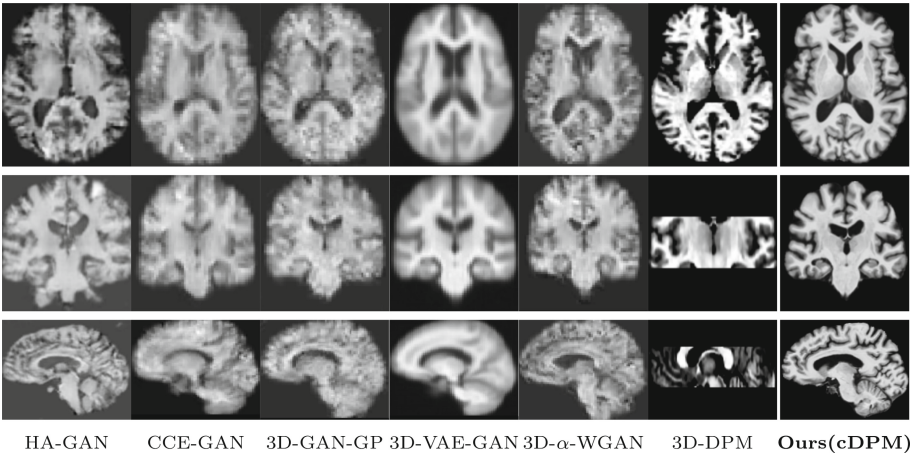
Fig. 4. 3 views of MRIs generated by 7 models. Compared to the MRIs produced by the other approaches, our cDPM model generates the most realistic MRI scans that provide more distinct gray matter boundaries and greater anatomical details. (Color figure online)

3.3   Quantitative Comparison

We evaluate the quality of synthetic MRIs based on 3 metrics: (i) computing the distance between synthetic and 500 randomly selected real MRIs via the Maximum-Mean Discrepancy (MMD) score [5], (ii) measuring the diversity of the synthetic MRIs via the pair-wise multi-scale Structure Similarity (MS-SSIM) [12], and (iii) comparing the distributions of synthetic to real MRIs with respect to the 3 views via the Frèchet Inception Distance (FID) [26] (a.k.a, FID-Axial, FID-Coronal, FID-Sagittal).

We compare those scores to ones produced by six recently published methods: (i) 3D-DPM [3], (ii) 3D-VAE-GAN [14], (iii) 3D-GAN-GP [6], (iv) 3D-α-WGAN [12], (v) CCE-GAN [26], and (vi) HA-GAN [23]. We needed to reimplement the first 5 methods and used the open-source code available for HA-GAN. 3D-DPM was only able to generate 32 slices at a time (due to GPU limitations) so that we computed its quality metrics by also cropping the corresponding real MRI to those 32 slices.
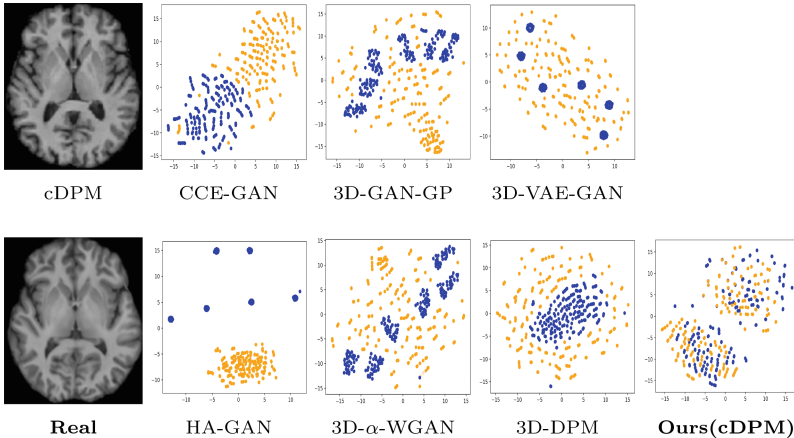
Fig. 5. Left: One MRI generated by our model and its closest real MRI based on MS-SSIM. **Right:** tSNE embedding of 200 generated samples (blue) of each model and their closest real MRIs (orange). Only our model generated independent and diverse samples as the data points overlay but are not identical to the training data. (Color figure online)

3.4   Results

Qualitative Results. The center of the axial, coronal, and sagittal views of five MRIs generated by cDPM shown in Fig. 3 look realistic. Compared to the MRIs produced by the other approaches other than 3D-DPM (see Fig. 4), the MRIs of cDPM are sharper; specifically, the gray matter boundaries are more distinct and the scan provides greater anatomical details. As expected, 3D-DPM

produced synthetic slices of similar quality as cDPM but failed to do so for the entire MRI.

The synthetic MRIs of cDPM shown in Fig. 3 are also substantially different from each other, suggesting that our method could be used to create an augmented data set that is anatomically diverse. Figure 5 further substantiates the claim, which plots the t-SNE embedding [15] of 200 synthetic MRIs (blue) and their closest real counterpart (orange) according to MS-SSIM for each method. Note, matching of all 500 synthetic MRIs was computationally too expensive to perform (takes days to complete per method). Based on those plots, cDPM is the only approach able to generate MRIs, whose distribution resembled that of the real MRIs. This finding is somewhat surprising given that the MRI subvolumes generated by 3D-DPM looked real. Unlike the real data, however, their distributions are clustered around the average. Thus, 3D-DPM fails to diversify the data set even if (in the future) more computational resources would allow the method to generate a complete 3D MRI.

Table 1. Measuring the quality of 500 synthetic MRIs. '( )' contains the absolute difference to the MS-SSIM score of the real MRIs, which was 0.792. In bold are the optimal scores among methods that generate the entire volume. Scores denoted with an asterisk '*' are only computed on 32 slices.

	MS-SSIM (%)	MMD↓ (10^3)	FID-A ↓	FID-C ↓	FID-S ↓
3D-VAE-GAN [14]	88.3 (9.1)	5.15	320	247	398
3D-GAN-GP [6]	81.0 (1.8)	15.7	141	127	281
3D-α-WGAN [12]	82.6 (3.4)	13.2	121	116	193
CCE-GAN [26]	81.5 (2.3)	3.54	69.4	869	191
HA-GAN [23]	36.8 (42.4)	226	477	1090	554
3D-DPM [3]	79.7 (0.5)*	15.2*	188	–	–
Ours (cDPM)	**78.6 (0.6)**	**3.14**	**32.4**	**45.8**	**91.1**

Quantitative Results. Table 1 lists the average scores of MS-SSIM, MMD, and FID for each method. Among all models that generated complete MRI volumes, cDPM performed best. Only the absolute difference between the MS-SSIM score of 3D-DPM and the real MRIs was slightly lower (i.e., 0.005) than the absolute difference for cDPM (i.e., 0.006). This comparison, however, is not fair as the MS-SSIM score for 3D-DPM was only computed on 32 slices. Further supporting this argument is that FID-A (the only score computed for the same slice across all methods) was almost 5 times worse for 3D-DPM than cDPM.

4   Conclusion

We propose a novel conditional DPM (cDPM) for efficiently generating 3D brain MRIs. Starting with random noise, our model can progressively generate MRI slices based on previously generated slices. This conditional scheme enables training the cDPM with limited computational resources and training data. Qualitative and quantitative results demonstrate that the model is able to produce high-fidelity 3D MRIs and outperform popular and recent generative models such as the CCE-GAN and 3D-DPM. Our framework can easily be extended to other imaging modalities and can potentially assist in training deep learning models on a small number of samples.

References

1. Bermudez, C., Plassard, A.J., Davis, L.T., Newton, A.T., Resnick, S.M., Landman, B.A.: Learning implicit brain MRI manifolds with deep learning. In: Medical Imaging 2018: Image Processing, vol. 10574, pp. 408–414. SPIE (2018)
2. Dar, S.U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., Çukur, T.: Image synthesis in multi-contrast MRI with conditional generative adversarial networks. IEEE Trans. Med. Imaging **38**(10), 2375–2388 (2019)
3. Dorjsembe, Z., Odonchimed, S., Xiao, F.: Three-dimensional medical image synthesis with denoising diffusion probabilistic models. In: Medical Imaging with Deep Learning (2022). https://openreview.net/forum?id=Oz7lKWVh45H
4. Goodfellow, I., et al.: Generative adversarial networks. Commun. ACM **63**(11), 139–144 (2020)
5. Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A.: A kernel two-sample test. J. Mach. Learn. Res. **13**(1), 723–773 (2012)
6. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of wasserstein GANs. In: NIPS, vol. 30, pp. 5769–5779 (2017)
7. Han, C., et al.: GAN-based synthetic brain MR image generation. In: IEEE International Symposium on Biomedical Imaging, pp. 734–738 (2018)
8. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Adv. Neural Inf. Process. Syst. **33**, 6840–6851 (2020)
9. 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
10. Jung, E., Luna, M., Park, S.H.: Conditional GAN with 3D discriminator for MRI generation of Alzheimer's disease progression. Pattern Recogn. **133**, 109061 (2023)
11. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4401–4410 (2019)

12. Kwon, G., Han, C., Kim, D.: Generation of 3D brain MRI using auto-encoding generative adversarial networks. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11766, pp. 118–126. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32248-9_14

13. La Barbera, G., et al.: Anatomically constrained CT image translation for heterogeneous blood vessel segmentation. In: British Machine Vision Virtual Conference, p. 776 (2022)

14. Larsen, A.B.L., Sønderby, S.K., Larochelle, H., Winther, O.: Autoencoding beyond pixels using a learned similarity metric. In: International Conference on Machine Learning, pp. 1558–1566. Proceedings of Machine Learning Research (2016)

15. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**(11) (2008)

16. Ouyang, J., Adeli, E., Pohl, K.M., Zhao, Q., Zaharchuk, G.: Representation disentanglement for multi-modal brain MRI Analysis. In: Feragen, A., Sommer, S., Schnabel, J., Nielsen, M. (eds.) IPMI 2021. LNCS, vol. 12729, pp. 321–333. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78191-0_25

17. Pawlowski, N., Coelho de Castro, D., Glocker, B.: Deep structural causal models for tractable counterfactual inference. Adv. Neural Inf. Process. Syst. **33**, 857–869 (2020)

18. Pinaya, W.H., et al.: Brain imaging generation with latent diffusion models. In: Deep Generative Models: DGM4MICCAI 2022, pp. 117–126 (2022)

19. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, vol. 9351, pp. 234–241 (2015)

20. Shaw, P., Uszkoreit, J., Vaswani, A.: Self-attention with relative position representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 2, pp. 464–468. Association for Computational Linguistics, New Orleans (2018). https://doi.org/10.18653/v1/N18-2074, https://aclanthology.org/N18-2074

21. Shin, H.C., et al.: Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: International Workshop on Simulation and Synthesis in Medical Imaging, vol. 11037 (2018)

22. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: International Conference on Machine Learning, pp. 2256–2265. PMLR (2015)

23. Sun, L., Chen, J., Xu, Y., Gong, M., Yu, K., Batmanghelich, K.: Hierarchical amortized GAN for 3D high resolution medical image synthesis. IEEE J. Biomed. Health Inf. **26**(8), 3966–3975 (2022)

24. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30 (2017)

25. Wolleb, J., Bieder, F., Sandkühler, R., Cattin, P.C.: Diffusion models for medical anomaly detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, vol. 13438, pp. 35–45 (2022)

26. Xing, S., Sinha, H., Hwang, S.J.: Cycle consistent embedding of 3D brains with auto-encoding generative adversarial networks. In: Medical Imaging with Deep Learning (2021)

27. Yu, B., Zhou, L., Wang, L., Fripp, J., Bourgeat, P.: 3D cGAN based cross-modality MR image synthesis for brain tumor segmentation. In: IEEE 15th International Symposium on Biomedical Imaging, pp. 626–630 (2018)

28. Zhang, J., et al.: Multi-label, multi-domain learning identifies compounding effects of HIV and cognitive impairment. Med. Image Anal. **75**, 102246 (2022)

29. Zhao, Q., Liu, Z., Adeli, E., Pohl, K.M.: Longitudinal self-supervised learning. Med. Image Anal. **71**, 102051 (2021)
30. Zheng, S., Charoenphakdee, N.: Diffusion models for missing value imputation in tabular data. In: NeurIPS Table Representation Learning (TRL) Workshop (2022)