MedGen3D: A Deep Generative Framework for Paired 3D Image and Mask Generation

Kun Han[1(✉)], Yifeng Xiong[1], Chenyu You[2], Pooya Khosravi[1], Shanlin Sun[1], Xiangyi Yan[1], James S. Duncan[2], and Xiaohui Xie[1]

[1] University of California, Irvine, USA
khan7@uci.edu
[2] Yale University, New Haven, USA

Abstract. Acquiring and annotating sufficient labeled data is crucial in developing accurate and robust learning-based models, but obtaining such data can be challenging in many medical image segmentation tasks. One promising solution is to synthesize realistic data with ground-truth mask annotations. However, no prior studies have explored generating complete 3D volumetric images with masks. In this paper, we present MedGen3D, a deep generative framework that can generate paired 3D medical images and masks. First, we represent the 3D medical data as 2D sequences and propose the Multi-Condition Diffusion Probabilistic Model (MC-DPM) to generate multi-label mask sequences adhering to anatomical geometry. Then, we use an image sequence generator and semantic diffusion refiner conditioned on the generated mask sequences to produce realistic 3D medical images that align with the generated masks. Our proposed framework guarantees accurate alignment between synthetic images and segmentation maps. Experiments on 3D thoracic CT and brain MRI datasets show that our synthetic data is both diverse and faithful to the original data, and demonstrate the benefits for downstream segmentation tasks. We anticipate that MedGen3D's ability to synthesize paired 3D medical images and masks will prove valuable in training deep learning models for medical imaging tasks.

Keywords: Deep Generative Framework · 3D Volumetric Images with Masks · Fidelity and Diversity · Segmentation

1 Introduction

In medical image analysis, the availability of a substantial quantity of accurately annotated 3D data is a prerequisite for achieving high performance in tasks like segmentation and detection [7,15,23,26,28–36]. This, in turn, leads to more

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43907-0_72.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14220, pp. 759–769, 2023.
https://doi.org/10.1007/978-3-031-43907-0_72

precise diagnoses and treatment plans. However, obtaining and annotating such data presents many challenges, including the complexity of medical images, the requirement for specialized expertise, and privacy concerns.

Generating realistic synthetic data presents a promising solution to the above challenges as it eliminates the need for manual annotation and alleviates privacy risks. However, most prior studies [4,5,14,25] have focused on 2D image synthesis, with only a few generating corresponding segmentation masks. For instance, [13] uses dual generative adversarial networks (GAN) [12,34] to synthesize 2D labeled retina fundus images, while [10] combines a label generator [22] with an image generator [21] to generate 2D brain MRI data. More recently, [24] uses WGAN [3] to generate small 3D patches and corresponding vessel segmentations.

However, there has been no prior research on generating whole 3D volumetric images with the corresponding segmentation masks. Generating 3D volumetric images with corresponding segmentation masks faces two major obstacles. First, directly feeding entire 3D volumes to neural networks is impractical due to GPU memory constraints, and downsizing the resolution may compromise the quality of the synthetic data. Second, treating the entire 3D volume as a single data point during training is suboptimal because of the limited availability of annotated 3D data. Thus, innovative methods are required to overcome these challenges and generate high-quality synthetic 3D volumetric data with corresponding segmentation masks.

We propose MedGen3D, a novel diffusion-based deep generative framework that generates paired 3D volumetric medical images and multi-label masks. Our approach treats 3D medical data as sequences of slices and employs an autoregressive process to sequentially generate 3D masks and images. In the first stage, a Multi-Condition Diffusion Probabilistic Model (MC-DPM) generates mask sequences by combining conditional and unconditional generation processes. Specifically, the MC-DPM generates mask subsequences (i.e., several consecutive slices) at any position directly from random noise or by conditioning on existing slices to generate subsequences forward or backward. Given that medical images have similar anatomical structures, slice indices serve as additional conditions to aid the mask subsequence generation. In the second stage, we introduce a conditional image generator with a seq-to-seq model from [27] and a semantic diffusion refiner. By conditioning on the mask sequences generated in the first stage, our image generator synthesizes realistic medical images aligned with masks while preserving spatial consistency across adjacent slices.

The main contributions of our work are as follows: 1) Our proposed framework is the *first* to address the challenge of synthesizing complete 3D volumetric medical images with their corresponding masks; 2) we introduce a multi-condition diffusion probabilistic model for generating 3D anatomical masks with high fidelity and diversity; 3) we leverage the generated masks to condition an image sequence generator and a semantic diffusion refiner, which produces realistic medical images that align accurately with the generated masks; and 4) we present experimental results that demonstrate the fidelity and diversity of the generated 3D multi-label medical images, highlighting their potential benefits for downstream segmentation tasks.

2 Preliminary

2.1 Diffusion Probabilistic Model

A diffusion probabilistic model (DPM) [16] is a parameterized Markov chain of length T, which is designed to learn the data distribution $p(X)$. DPM builds the Forward Diffusion Process (FDP) to get the diffused data point X_t at any time step t by $q\left(X_t \mid X_{t-1}\right)=\mathcal{N}\left(X_t ; \sqrt{1-\beta_t} X_{t-1}, \beta_t I\right)$, with $X_0 \sim q(X_0)$ and $p(X_T)=\mathcal{N}\left(X_T ; 0, I\right)$. Let $\alpha_t=1-\beta_t$ and $\bar{\alpha}_t=\prod_{s=1}^{t}\left(1-\beta_s\right)$, Reverse Diffusion Process (RDP) is trained to predict the noise added in the FDP by minimizing:

$$Loss(\theta)=\mathbb{E}_{X_0 \sim q(X_0), \epsilon \sim \mathcal{N}(0, I), t}\left[\left\|\epsilon-\epsilon_\theta\left(\sqrt{\bar{\alpha}_t} X_0+\sqrt{1-\bar{\alpha}_t} \epsilon, t\right)\right\|^2\right], \quad (1)$$

where ϵ_θ is predicted noise and θ is the model parameters.

2.2 Classifier-Free Guidance

Samples from conditional diffusion models can be improved with classifier-free guidance [17] by setting the condition c as $\emptyset$ with probability p. During sampling, the output of the model is extrapolated further in the direction of $\epsilon_\theta\left(X_t \mid c\right)$ and away from $\epsilon_\theta\left(X_t \mid \emptyset\right)$ as follows:

$$\hat{\epsilon}_\theta\left(X_t \mid c\right)=\epsilon_\theta\left(X_t \mid \emptyset\right)+s \cdot\left(\epsilon_\theta\left(X_t \mid c\right)-\epsilon_\theta\left(X_t \mid \emptyset\right)\right), \quad (2)$$

where $\emptyset$ represents a null condition and $s \geq 1$ is the guidance scale.

3 Methodology

We propose a sequential process to generate complex 3D volumetric images with masks, as illustrated in Fig. 1. The first stage generates multi-label segmentation, and the second stage performs conditional medical image generation. The details will be presented in the following sections.

3.1 3D Mask Generator

Due to the limited annotated real data and GPU memory constraints, directly feeding the entire 3D volume to the network is impractical. Instead, we treat 3D

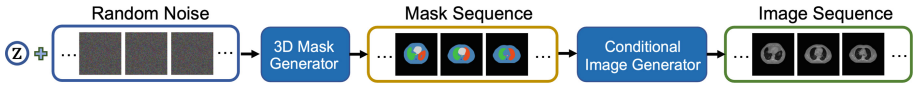

Fig. 1. Overview of the proposed **MedGen3D**, including a 3D mask generator to autoregressively generate the mask sequences starting from a random position z, and a conditional image generator to generate 3D images conditioned on generated masks.

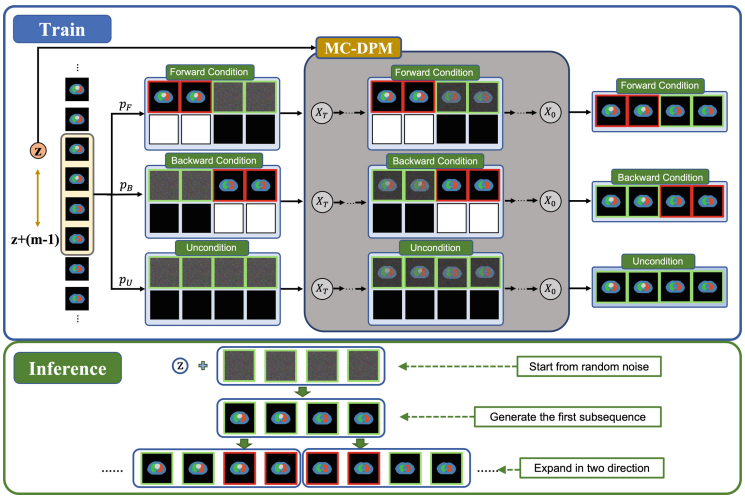

Fig. 2. Proposed 3D mask generator. Given target position z, MC-DPM is designed to generate mask subsequences (length of m) for specific region, unconditionally or conditioning on first or last n slices, according to the pre-defined probability $p^C \in \{p_F, p_B, p_U\}$. Binary indicators are assigned to slices to signify the conditional slices. We ignore the binary indicators in the inference process for clear visualization with red outline denoting the conditional slices and green outline denoting the generated slices.

medical data as a series of subsequences. To generate an entire mask sequence, an initial subsequence of m consecutive slices is **unconditionally** generated from random noise. Then the subsequence is expanded **forward** and **backward** in an autoregressive manner, conditioned on existing slices.

Inspired by classifier-free guidance in Sect. 2.2, we propose a general Multi-Condition Diffusion Probabilistic Model (MC-DPM) to unify all three conditional generations (unconditional, forward, and backward). As shown in Fig. 2, MC-DPM is able to generate mask sequences directly from random noise or conditioning on existing slices.

Furthermore, as 3D medical data typically have similar anatomical structures, slices with the same relative position roughly correspond to the same anatomical regions. Therefore, we can utilize the relative position of slices as conditions to guide the MC-DPM in generating subsequences of the target region and control the length of generated sequences.

Train: For a given 3D multi-label mask $M \in \mathbb{R}^{D\times H\times W}$, subsequneces of m consecutive slices are selected as $\{M_z, M_{z+1}, \ldots, M_{z+(m-1)}\}$, with z as the randomly selected starting indices. For each subsequence, we determine the conditional slices $X^C \in \{\mathbb{R}^{n\times H\times W}, \emptyset\}$ by selecting either the first or the last n slices, or no slice, based on a probability $p^C \in \{p_{Forward}, p_{Backward}, p_{Uncondition}\}$. The objective of the MC-DPM is to generate the remaining slices, denoted as $X^P \in \mathbb{R}^{(m-\text{len}(X^C))\times H\times W}$.

To incorporate the position condition, we utilize the relative position of the subsequence $\tilde{z} = z/D$, where z is the index of the subsequence's starting slice. Then we embed the position condition and concatenate it with the time embedding to aid the generation process. We also utilize a binary indicator for each slice in the subsequence to signify the existence of conditional slices.

The joint distribution of reverse diffusion process (RDP) with the conditional slices X^C can be written as:

$$p_\theta(X^P_{0:T}|X^C, \tilde{z}) = p(X^P_T) \prod_{t=1}^{T} p_\theta(X^P_{t-1} \mid X^P_t, X^C, \tilde{z}). \tag{3}$$

where $p(X^P_T) = \mathcal{N}\left(X^P_T; 0, I\right)$, $\tilde{z} = z/D$ and p_θ is the distribution parameterized by the model.

Overall, the model will be trained by minimizing the following loss function, with $X^P_t = \sqrt{\bar{\alpha}_t} X^P_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$:

$$\text{Loss}(\theta) = \mathbb{E}_{X_0 \sim q(X_0), \epsilon \sim \mathcal{N}(0,I), p^C, z, t}\left[\left\|\epsilon - \epsilon_\theta\left(X^P_t, X^C, z, t\right)\right\|^2\right]. \tag{4}$$

Inference: During inference, MC-DPM first generates a subsequence of m slices from random noise given a random location z. The entire mask sequence can then be generated autoregressively by expanding in both directions, conditioned on the existing slices, as shown in Fig. 2. Please refer to the **Supplementary** for a detailed generation process and network structure.

3.2 Conditional Image Generator

In the second step, we employ a sequence-to-sequence method to generate medical images conditioned on masks, as shown in Fig. 3.

Image Sequence Generator: In the sequence-to-sequence generation task, new slice is the combination of the warped previous slice and newly generated texture, weighted by a continuous mask [27]. We utilize Vid2Vid [27] as our image sequence generator. We train Vid2Vid with its original loss, which includes GAN loss on multi-scale images and video discriminators, flow estimation loss, and feature matching loss.

Semantic Diffusion Refiner: Despite the high cross-slice consistency and spatial continuity achieved by vid2vid, issues such as blocking, blurriness and suboptimal texture generation persist. Given that diffusion models have been shown to generate superior images [9], we propose a semantic diffusion refiner utilizing a diffusion probabilistic model to refine the previously generated images.

For each of the 3 different views, we train a semantic diffusion model (SDM), which takes 2D masks and noisy images as inputs to generate images aligned with input masks. During inference, we only apply small noising steps (10 steps)

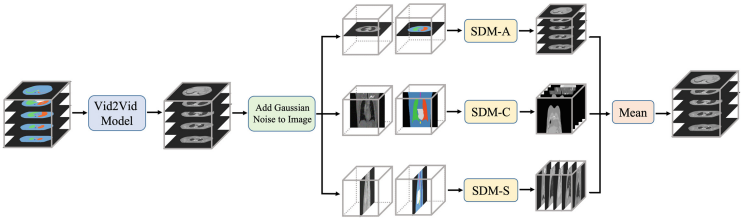

Fig. 3. Image Sequence Generator. Given the generated 3D mask, the initial image is generated by Vid2Vid model sequentially. To utilize the semantic diffusion model (SDM) to refine the initial result, we first apply small steps (10 steps) noise, and then use three SDMs to refine. The final result is the mean 3D images from 3 different views (Axial, Coronal, and Sagittal), yielding significant improvements over the initially generated image.

to the generated images so that the overall anatomical structure and spatial continuity are preserved. After that, we refine the images using the pre-trained semantic diffusion model. The final refined 3D images are the mean results from 3 views. Experimental results show an evident improvement in the quality of generated images with the help of semantic diffusion refiner.

4 Experiments and Results

4.1 Datasets and Setups

Datasets: We conducted experiments on the thoracic site using three thoracic CT datasets and the brain site with two brain MRI datasets. For both generative models and downstream segmentation tasks, we utilized the following datasets:

- SegTHOR [19]: 3D thorax CT scans (25 training, 5 validation, 10 testing);
- OASIS [20]: 3D brain MRI T1 scans (40 training, 10 validation, 10 testing);

For the downstream segmentation task only and the transfer learning, we utilized 10 fine-tuning, 5 validation, and 10 testing scans from each of the 3D thorax CT datasets of StructSeg-Thorax [2] and Public-Thor [7], as well as the 3D brain MRI T1 dataset from ADNI [1].

Implementation: For thoracic datasets, we crop and pad CT scans to $(96 \times 320 \times 320)$. The annotations of six organs (left lung, right lung, spinal cord, esophagus, heart, and trachea) are examined by an experienced radiation oncologist. We also include a body mask to aid in the image generation of body regions. For brain MRI datasets, we use Freesurfer [11] to get segmentations of four regions (cortex, subcortical gray matter, white matter, and CSF), and then crop the volume to $(192 \times 160 \times 160)$. We assign discrete values to masks of different regions or organs for both thoracic and brain datasets and then combine them into one 3D volume. When synthesizing mask sequences, we resize the

width and height of the masks to 128×128 and set the length of the subsequence m to 6. We use official segmentation models provided by MONAI [6] along with standard data augmentations, including spatial and color transformations.

Setup: We compare the synthetic image quality with DDPM [16], 3D-α-WGAN [18] and Vid2Vid [27], and utilize four segmentation models with different training strategies to demonstrate the benefit for the downstream task.

4.2 Evaluate the Quality of Synthetic Image.

Synthetic Dataset: To address the limited availability of annotated 3D medical data, we used only 30 CT scans from SegTHOR (25 for training and 5 for validation) and 50 MRI scans from OASIS (40 for training and 10 for validation) to generate 110 3D thoracic CT scans and 110 3D brain MRI scans, respectively (Fig. 4).

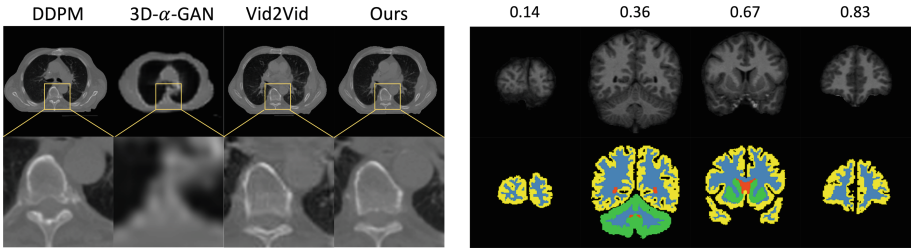

Fig. 4. Our proposed method produces more anatomically accurate images compared to 3D-α-WGAN and vid2vid, as demonstrated by the clearer organ boundaries and more realistic textures. Left: Qualitative comparison between different generative models. Right: Visualization of synthetic 3D brain MRI slices at different relative positions.

We compare the fidelity and diversity of our synthetic data with DDPM [16] (train 3 for different views), 3D-α-WGAN [18], and vid2vid [27] by calculating the mean Frèchet Inception Distance (FID) and Learned Perceptual Image Patch Similarity (LPIPS) from 3 different views.

According to Table 1, our proposed method has a slightly lower FID score but a similar LPIPS score compared to DDPM which directly generates 2D images from noise. We speculate that this is because DDPM is trained on 2D images without explicit anatomical constraints and only generates 2D images. On the other hand, 3D-α-WGAN [18], which uses much larger 3D training data (146 for thorax and 414 for brain), has significantly worse FID and LPIPS scores than our method. Moreover, our proposed method outperforms Vid2Vid, showing the effectiveness of our semantic diffusion refiner.

Table 1. Synthetic image quality comparison between baselines and ours.

	Thoracic CT		Brain MRI	
	FID ↓	LPIPS ↑	FID ↓	LPIPS ↑
DDPM [16]	**35.2**	**0.316**	**34.9**	0.298
3D-α-WGAN [18]	136.2	0.286	136.4	0.289
Vid2Vid [27]	47.3	0.300	48.2	0.324
Ours	39.6	0.305	40.3	**0.326**

Table 2. Experiment 2: DSC of different thoracic segmentation models. There are 5 training strategies, namely: **E2-1:** Training with real SegTHOR training data; **E2-2:** Training with synthetic data; **E2-3:** Training with both synthetic and real data; **E2-4:** Finetuning model from E2-2 using real training data; and **E2-5:** finetuning model from E2-3 using real training data. (* denotes the training data source.)

	SegTHOR*				StructSeg-Thorax				Public-Thor			
	Unet 2D	Unet 3D	UNETR	Swin UNETR	Unet 2D	Unet 3D	UNETR	Swin UNETR	Unet 2D	Unet 3D	UNETR	Swin UNETR
E2-1	0.817	0.873	0.867	0.878	0.722	0.793	0.789	0.810	0.822	0.837	0.836	0.847
E2-2	0.815	0.846	0.845	0.854	0.736	0.788	0.788	0.803	0.786	0.838	0.814	0.842
E2-3	0.845	0.881	0.886	0.886	0.772	0.827	0.824	0.827	0.812	0.856	0.853	0.856
E2-4	**0.855**	0.887	**0.894**	**0.899**	0.775	**0.833**	**0.825**	0.833	**0.824**	0.861	0.852	**0.867**
E2-5	0.847	**0.891**	0.890	0.897	**0.783**	**0.833**	0.823	**0.835**	0.818	**0.864**	**0.858**	**0.867**

4.3 Evaluate the Benefits for Segmentation Task

We explore the benefits of synthetic data for downstream segmentation tasks by comparing Sørensen-Dice coefficient (DSC) of 4 segmentation models, including Unet2D [23], UNet3D [8], UNETR [15], and Swin-UNETR [26]. In Table 2 and 3, we utilize real training data (from SegTHOR and OASIS) and synthetic data to train the segmentation models with 5 different strategies, and test on all 3 thoracic CT datasets and 2 brain MRI datasets. In Table 4, we aim to demonstrate whether the synthetic data can aid transfer learning with limited real finetuning data from each of the testing datasets (StructSeg-Thorax, Public-Thor and ADNI) with four training strategies.

According to Table 2 and Table 3, the significant DSC difference between 2D and 3D segmentation models underlines the crucial role of 3D annotated data. While purely synthetic data (**E2-2**) fails to achieve the same performance as real training data (**E2-1**), the combination of real and synthetic data (**E2-3**) improves model performance in most cases, except for Unet2D on the Public-Thor dataset. Furthermore, fine-tuning the pre-trained model with real data (**E2-4** and **E2-5**) consistently outperforms the model trained only with real data. Please refer to **Supplementary** for organ-level DSC comparisons of the Swin-UNETR model with more details.

According to Table 4, for transfer learning, utilizing the pre-trained model (**E3-2**) leads to better performance compared to training from scratch (**E3-1**).

Table 3. Experiment 2: DSC of brain segmentation models. Please refer to Table 2 for detailed training strategies. (* denotes the training data source.)

	OASIS*				ADNI			
	Unet 2D	Unet 3D	UNETR	Swin UNETR	Unet 2D	Unet 3D	UNETR	Swin UNETR
E2-1	0.930	0.951	0.952	0.954	0.815	0.826	0.880	0.894
E2-2	0.905	0.936	0.935	0.934	0.759	0.825	0.828	0.854
E2-3	0.938	0.953	0.953	0.955	0.818	0.888	0.898	**0.906**
E2-4	**0.940**	**0.955**	**0.954**	**0.956**	**0.819**	0.891	**0.903**	0.903
E2-5	**0.940**	0.954	**0.954**	**0.956**	**0.819**	**0.894**	0.902	**0.906**

Table 4. Experiment 3: DSC of Swin-UNETR finetuned with real dataset. There are 4 training strategies: **E3-1:** Training from scratch for each dataset using limited finetuning data; **E3-2** Finetuning the model E2-1 from experiment 2; **E3-3** Finetuning the model E2-4 from experiment 2; and **E3-4** Finetuning the model E2-5 from experiment 2. (* denotes the finetuning data source.)

	Thoracic CT		Brain MRI
	StructSeg-Thorax*	Public-Thor*	ADNI*
E3-1	0.845	0.897	0.946
E3-2	0.865	0.901	0.948
E3-3	0.878	0.913	**0.949**
E3-4	**0.882**	**0.914**	**0.949**

Additionally, pretraining the model with synthetic data (**E3-3** and **E3-4**) can facilitate transfer learning to a new dataset with limited annotated data.

We have included video demonstrations of the generated 3D volumetric images in the **supplementary material**, which offer a more comprehensive representation of the generated image's quality.

5 Conclusion

This paper introduces MedGen3D, a new framework for synthesizing 3D medical mask-image pairs. Our experiments demonstrate its potential in realistic data generation and downstream segmentation tasks with limited annotated data. Future work includes merging the image sequence generator and semantic diffusion refiner for end-to-end training and extending the framework to synthesize 3D medical images across modalities. Overall, we believe that our work opens up new possibilities for generating 3D high-quality medical images paired with masks, and look forward to future developments in this field.

References

1. https://adni.loni.usc.edu/
2. https://structseg2019.grand-challenge.org/dataset/
3. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein gan. arXiv preprint arXiv: Arxiv-1701.07875 (2017)
4. Baur, C., Albarqouni, S., Navab, N.: Melanogans: high resolution skin lesion synthesis with gans. arXiv preprint arXiv:1804.04338 (2018)
5. 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: Image Processing. SPIE (2018)
6. Cardoso, M.J., et al.: Monai: an open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
7. Chen, X., et al.: A deep learning-based auto-segmentation system for organs-at-risk on whole-body computed tomography images for radiation therapy. Radiother. Oncol. **160**, 175–184 (2021)
8. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
9. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. In: NeurIPS (2021)
10. Fernandez, V., et al.: Can segmentation models be trained with fully synthetically generated data? In: Zhao, C., Svoboda, D., Wolterink, J.M., Escobar, M. (eds.) MICCAI Workshop. SASHIMI 2022, vol. 13570, pp. 79–90. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-16980-9_8
11. Fischl, B.: Freesurfer. In: Neuroimage (2012)
12. Goodfellow, I., et al.: Generative adversarial networks. Commun. ACM **63**, 139–144 (2020)
13. Guibas, J.T., Virdi, T.S., Li, P.S.: Synthetic medical images from dual generative adversarial networks. arXiv preprint arXiv:1709.01872 (2017)
14. Han, C., et al.: Gan-based synthetic brain MR image generation. In: ISBI. IEEE (2018)
15. Hatamizadeh, A., et al.: Unetr: transformers for 3d medical image segmentation. In: WACV (2022)
16. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: NeurIPS (2020)
17. Ho, J., Salimans, T.: Classifier-free diffusion guidance. arXiv preprint arXiv: Arxiv-2207.12598 (2022)
18. 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
19. Lambert, Z., Petitjean, C., Dubray, B., Kuan, S.: Segthor: segmentation of thoracic organs at risk in ct images. In: IPTA. IEEE (2020)
20. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. J. Cogn. Neurosci. **19**, 1498–1507 (2007)

21. Park, T., Liu, M.Y., Wang, T.C., Zhu, J.Y.: Semantic image synthesis with spatially-adaptive normalization. In: CVPR (2019)
22. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: CVPR (2022)
23. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
24. Subramaniam, P., Kossen, T., et al.: Generating 3d tof-mra volumes and segmentation labels using generative adversarial networks. Med. Image Anal. **78**, 102396 (2022)
25. 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. **28**, 3966–3975 (2022)
26. Tang, Y., et al.: Self-supervised pre-training of swin transformers for 3d medical image analysis. In: CVPR (2022)
27. Wang, T.C., et al.: Video-to-video synthesis. arXiv preprint arXiv:1808.06601 (2018)
28. Yan, X., Tang, H., Sun, S., Ma, H., Kong, D., Xie, X.: After-unet: axial fusion transformer unet for medical image segmentation. In: WACV (2022)
29. You, C., et al.: Mine your own anatomy: revisiting medical image segmentation with extremely limited labels. arXiv preprint arXiv:2209.13476 (2022)
30. You, C., et al.: Rethinking semi-supervised medical image segmentation: a variance-reduction perspective. arXiv preprint arXiv:2302.01735 (2023)
31. You, C., Dai, W., Min, Y., Staib, L., Duncan, J.S.: Implicit anatomical rendering for medical image segmentation with stochastic experts. arXiv preprint arXiv:2304.03209 (2023)
32. You, C., Dai, W., Min, Y., Staib, L., Sekhon, J., Duncan, J.S.: Action++: improving semi-supervised medical image segmentation with adaptive anatomical contrast. arXiv preprint arXiv:2304.02689 (2023)
33. You, C., Dai, W., Staib, L., Duncan, J.S.: Bootstrapping semi-supervised medical image segmentation with anatomical-aware contrastive distillation. arXiv preprint arXiv:2206.02307 (2022)
34. You, C., et al.: Class-aware adversarial transformers for medical image segmentation. In: NeurIPS (2022)
35. You, C., Zhao, R., Staib, L.H., Duncan, J.S.: Momentum contrastive voxel-wise representation learning for semi-supervised volumetric medical image segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI, vol. 13434, pp. 639–652. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-16440-8_61
36. You, C., Zhou, Y., Zhao, R., Staib, L., Duncan, J.S.: Simcvd: simple contrastive voxel-wise representation distillation for semi-supervised medical image segmentation. IEEE Trans. Med.Imaging **41**, 2228–2237 (2022)