

Phase IV-AI LAND: Lung and Nodule Diffusion for 3D Chest CT Synthesis with Anatomical Guidance



CIDAI-POC-2026-01

Some rights reserved. This work is available under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International ([CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)).

Under the terms of this licence, you may copy, redistribute and adapt the work for non-commercial purposes, provided the work is appropriately cited, as indicated below.

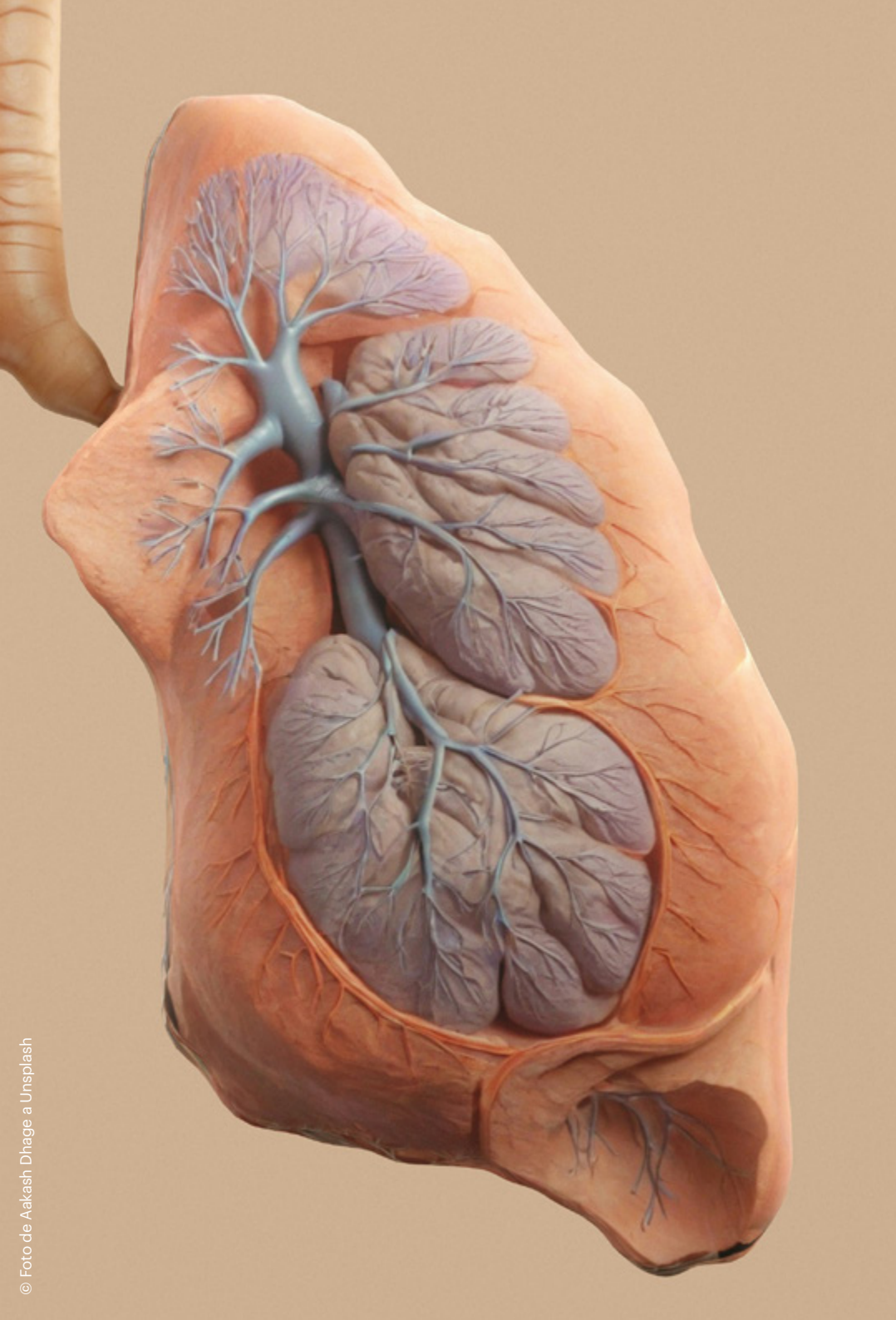
In any use of this work, there should be no suggestion that CIDAI endorses any specific organization, products or services. The use of the CIDAI logo is not permitted.

If you adapt the work, then you must license your work under the same or equivalent Creative Commons licence.

If you create a translation of this work, you should add the following disclaimer along with the suggested citation: “This translation was not created by the Centre of Innovation for Data tech and Artificial Intelligence (CIDAI). CIDAI is not responsible for the content or accuracy of this translation. The original English edition shall be the binding and authentic edition”.

Any mediation relating to disputes arising under the licence shall be conducted in accordance with the mediation rules of the [World Intellectual Property Organization](https://www.wipo.int/mediation/).

Suggested citation. CIDAI-POC-2026-01 // Phase IV-AI - Lung and Nodule Diffusion for 3D Chest CT Synthesis with Anatomical Guidance by Anna Oliveras Tous i Oriol Guardiola Olivella - Eurecat, CIDAI, 2026. Licence: [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)



1.	ABSTRACT	4
2.	DESCRIPTION OF THE PROBLEM SOLVED	6
3.	METHODOLOGY	8
4.	EXPERIMENTS	8
5.	CONCLUSIONS AND FUTURE WORK	15
6.	REFERENCES	16



1. Abstract

1.1. Objectives

The objective of this work is to develop an efficient latent diffusion framework for the conditional generation of high-quality, anatomically consistent 3D chest CT volumes from 3D anatomical masks. The model aims to improve synthesis quality as measured by established generative metrics—such as the Fréchet Inception Distance (FID) and structural similarity metrics—while substantially reducing computational requirements. A key performance target is to enable training and inference on a single mid-range GPU without compromising spatial resolution or anatomical fidelity.

1.2. Vision

This work seeks to enable scalable, controllable, and resource-efficient synthesis of realistic 3D chest CT scans. By lowering hardware barriers and improving anatomical controllability, the proposed approach supports data augmentation for AI model training, simulation of rare pathological cases, and educational applications in medical imaging, even in computationally constrained research or clinical environments.

1.3. Problem Summary

Three-dimensional medical image synthesis has become an important tool for addressing data scarcity, privacy concerns, and class imbalance in medical AI. However, existing state-of-the-art approaches typically rely on computationally intensive 3D GANs or diffusion models that require multiple high-end GPUs and long training times, limiting their accessibility and scalability. Moreover, conditional generation is often based on limited anatomical inputs, such as lesion or nodule masks alone, which can lead to anatomically implausible or globally inconsistent chest structures. These limitations reduce the clinical realism and downstream utility of synthesized CT scans for training, validation, and educational use.

1.4. Solution

We propose a latent diffusion model for 3D chest CT synthesis that generates full-resolution $256 \times 256 \times 256$ volumes at 1 mm isotropic spacing while operating on a single mid-range GPU. The model is conditioned on 3D anatomical masks representing both global lung anatomy and local nodule regions, enabling precise and anatomically consistent control over generated structures. Compared to existing 3D generative approaches, the proposed method significantly reduces memory and computational demands while maintaining competitive or improved generative quality, as demonstrated by quantitative metrics and qualitative evaluation. Experimental results highlight the importance of incorporating global lung structure to achieve anatomically coherent synthesis and show that the model can generate diverse CT volumes with and without pulmonary nodules of varying sizes, locations, and characteristics. This approach offers a practical and efficient alternative to current state-of-the-art methods for 3D medical image generation.

2. Description of the problem solved

2.1. Background

Synthetic medical images have emerged as an essential tool for addressing the limited availability of high-quality training data, one of the primary challenges in medical AI [1]. Deep learning has demonstrated outstanding performance in domains with abundant data, but its application in medical imaging is often hindered by the unavailability of large, diverse, and expertly annotated datasets. This limitation arises from several factors, including the costs of data acquisition, privacy regulations, and the time-consuming nature of expert annotation.

Deep generative models are today the most effective tool for synthesizing realistic, high-quality images. Among these, diffusion models [2] have gained special relevance, surpassing previous methods such as VAEs [4] and GANs [10] in terms of sample realism, variety, and training stability [5, 6]. Despite their generative capabilities, diffusion models struggle with high-resolution 3D synthesis due to computational demands, limiting their applicability to volumetric medical imaging modalities like Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) [1].

Among medical imaging modalities, chest CT is particularly relevant for the detection of pulmonary diseases, especially lung cancer, the leading cause of cancer-related deaths worldwide [3]. A major challenge in this domain is the significant variability in the size, shape, texture, and location of lung nodules [1, 7].



2.2. Related work

Several strategies aim to reduce computational cost while maintaining synthesis quality in diffusion models [8]. Among them, Latent Diffusion Models (LDMs) have become a cornerstone technique [15]. Rather than operating in the high-dimensional image space, LDMs perform the diffusion process in a compressed latent space learned through pretrained autoencoders. [4]. By capturing the semantic structure of the data, the latent space helps the model focus on generating high-level features instead of raw pixels.

Medical Diffusion [22] and other LDMs [16, 17] synthesize 3D MRI and CT scans, but are limited to resolutions at least $\times 3$ lower than LAND. To boost efficiency, 3D diffusion models like PatchDDM [19] and WDM [9] use alternative representations without autoencoders: PatchDDM generates subvolumes by conditioning on a spatial coordinate grid, while WDM uses wavelet decomposition to reduce the spatial dimension of the input volumes. While these approaches circumvent the need for autoencoders, they still use $\times 2$ the GPU memory of LAND during training for $256 \times 256 \times 256$ outputs.

Another challenge added to the computational cost is generating diverse samples with anatomically meaningful variations, such as detailed lung nodules. The incorporation of conditioning information into the diffusion process is a well-established strategy for controlling the characteristics of synthetic data [15]. In medical imaging, conditional diffusion models can generate clinically meaningful variations of scans. Conditioning inputs may include segmentation masks [27, 11, 14], numerical data [7, 16, 17], text prompts [12, 13, 25]

or multi-modal encodings [23, 39]. NVIDIA recently introduced the MAISI model [11], a conditional LDM capable of generating CT volumes up to $512 \times 512 \times 768$ voxels.

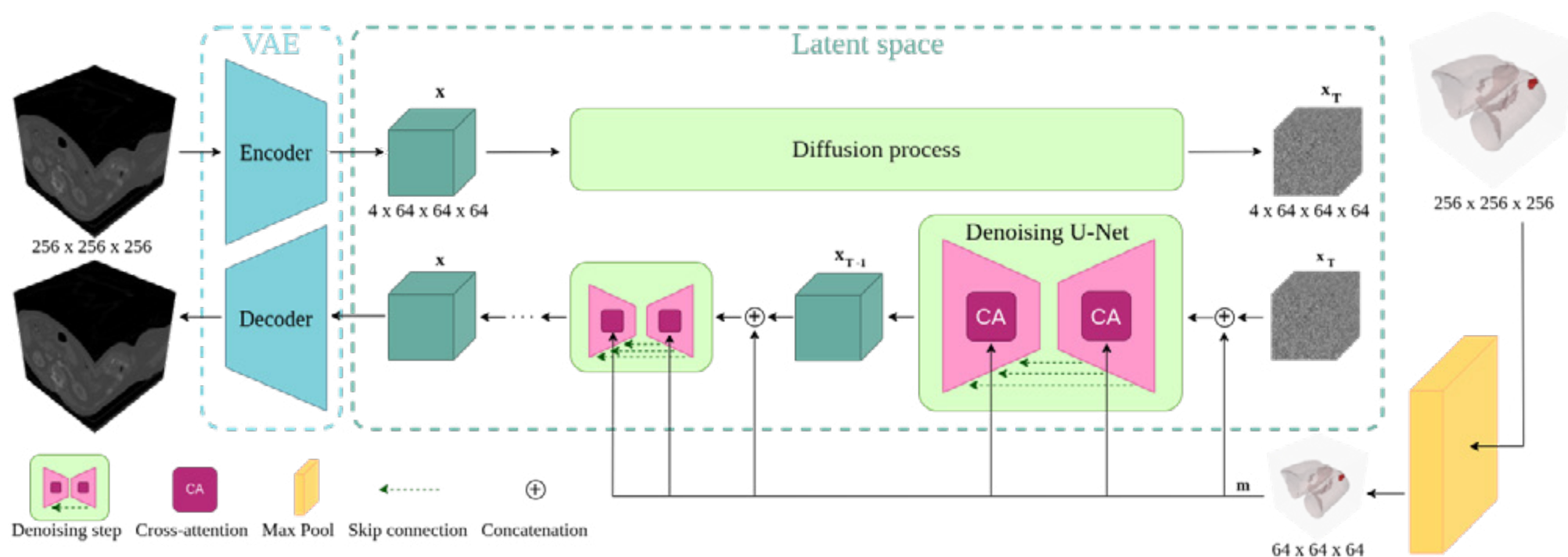
To the best of our knowledge, MAISI represents the highest reported resolution for synthetic 3D medical volumes, highlighting the scalability and potential of latent diffusion. However, such level of detail requires 58 GB of GPU memory, which remains a barrier for researchers with limited resource computation.

3. Methodology

We introduce LAND (Lung-And-Nodule-Diffusion), a latent diffusion model for generating high-quality 3D chest CT scans conditioned on 3D anatomical masks. An overview of the method is shown in Fig. 1. LAND consists of a 3D Variational Autoencoder (VAE) and

a 3D U-Net. We first train the 3D VAE to learn the mapping of 3D volumes into a latent space. The 3D U-Net is trained afterwards to learn the denoising generative process in the latent space, with the VAE weights kept frozen.

Fig. 1. Overview of the LAND pipeline. CT volumes are encoded into latent space using a 3D VAE. A 3D U-Net performs latent diffusion, optionally conditioned on an anatomical mask m with lung/nodule regions. Masks are subsampled to the latent space, concatenated with the latent samples x , and injected into cross-attention layers.



3.1. Latent Diffusion Architecture

3D VAE A fully 3D VAE encodes input CT images into latent representations compressing $\times 4$ the spatial resolution and expanding $\times 4$ the feature dimensionality, i.e., each $256 \times 256 \times 256$ volume is encoded as a $64 \times 64 \times 64 \times 4$ latent sample. We adopt a lightweight variant of the MAISI architecture [11], using 3 resolution levels with one residual block per level and the same number of channels in both the encoder and the decoder. The VAE is trained using a cost function comprising an L1 loss (MAE), a perceptual similarity loss (LPIPS), an adversarial loss and a Kullback-Leibler term.

3D U-Net The denoising network is a 3D U-Net with 5 resolution levels and two residual blocks per level. Similarly to [19], additive skip connections are used to reduce memory load while preserving spatial information flow. To enhance conditional generation, we follow [7, 15] and add cross-attention modules to the U-Net for re-injecting the conditioning masks (if any) at different resolution levels. Velocity prediction [34] is used to train the U-Net, enabling it to learn the denoising process by estimating a linear combination of clean latent and added noise. This formulation helps to improve training stability and sample quality in high-resolution synthesis [18, 34]. A linear noise schedule is used for the timesteps of the diffusion process, and training follows a Min-SNR- γ loss weighting to balance the contribution of each timestep t according to signal-to-noise ratios [33].

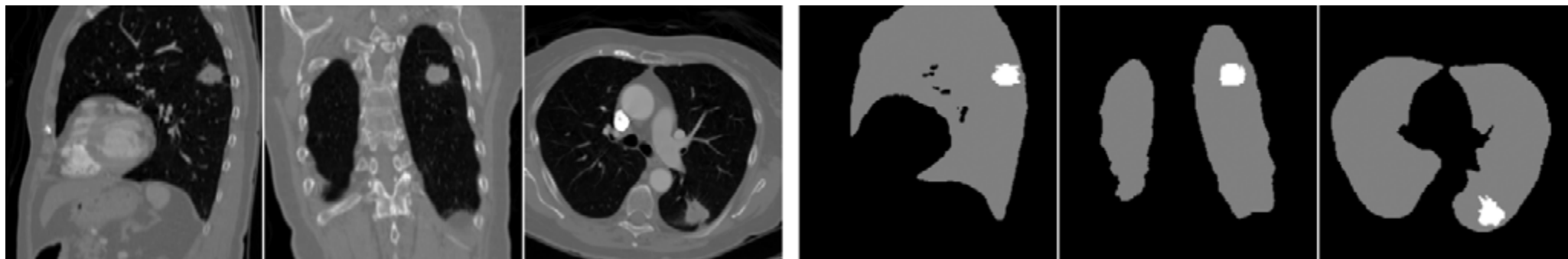


3.2. Lung and Nodule-Conditioned Diffusion

This section explains how to build the anatomical masks that condition the synthesis. Previous work [7] has shown that Conditioning solely on nodule masks often leads to anatomically implausible nodule placements. To address this, we use conditioning masks including lung and nodule regions (Fig. 2). To encode textural and spatial information, lung regions are assigned a value of 0.5, while

nodule regions receive texture values from 1 (Non-Solid) to 5 (Solid). The normalized mask (between [0,1]) is concatenated with the noisy latent sample feeding the U-Net [7, 15] and injected into the cross-attention layers. The masks are downsampled with 3D max pooling of factor 4 to match the latent resolution, preserving all nodules despite spatial collapse.

Fig. 2. Sagittal, coronal, and axial slices of a chest CT volume (left) and corresponding anatomical mask slices (right), including lung (grey) and nodule (white) regions.



4. Implementation

4.1. Experimental Settings

Datasets and Preprocessing Two publicly available datasets were used: LIDC-IDRI [21] and NLST [30]. The LIDC-IDRI data consists of 1,010 CT volumes annotated by four experienced radiologists with nodule masks and attribute ratings. For this study, nodule texture scores were considered, ranging from 1 to 5: 1 (Non-Solid), 2 (Non-Solid/Mixed), 3 (Part-Solid), 4 (Solid/Mixed), and 5 (Solid). The NLST CT scan collection contains over 25,000 volumes, of which we selected a subset of 881 with at least one nodule detection from [28] and generated the nodule masks using an ad-hoc U-Net segmenter. Lung areas were segmented in both datasets using a pre-trained open-source U-Net [20]. All CT scans were pre-processed similarly to [9]: intensities were clipped between -1000 and the 0.01 upper percentile to remove air and outliers, followed by resampling to 1 mm isotropic voxel size, center-cropping of $256 \times 256 \times 256$, and normalization to $[0, 1]$. The LIDC-IDRI data was used for training the LAND pipeline, while the NLST data was only used for inference and evaluation, to feed the model with realistic and unseen anatomical masks.

Metrics The Fréchet Inception Distance (FID) [37] is used to evaluate the quality of the synthetic 3D CT scans, while the Multi-Scale Structural Similarity Index Measure (MS-SSIM) [38] is used to assess their diversity. All FID scores were computed using 881 real and synthetic samples, with a ResNet-50 feature extractor pretrained on 23 medical image datasets [24], following the implementation in [9]. Lower FID indicates that the distribution of generated images closely matches that of real samples, while lower MS-SSIM corresponds to greater sample diversity. To calculate MS-SSIM, 10k pairs of synthetic

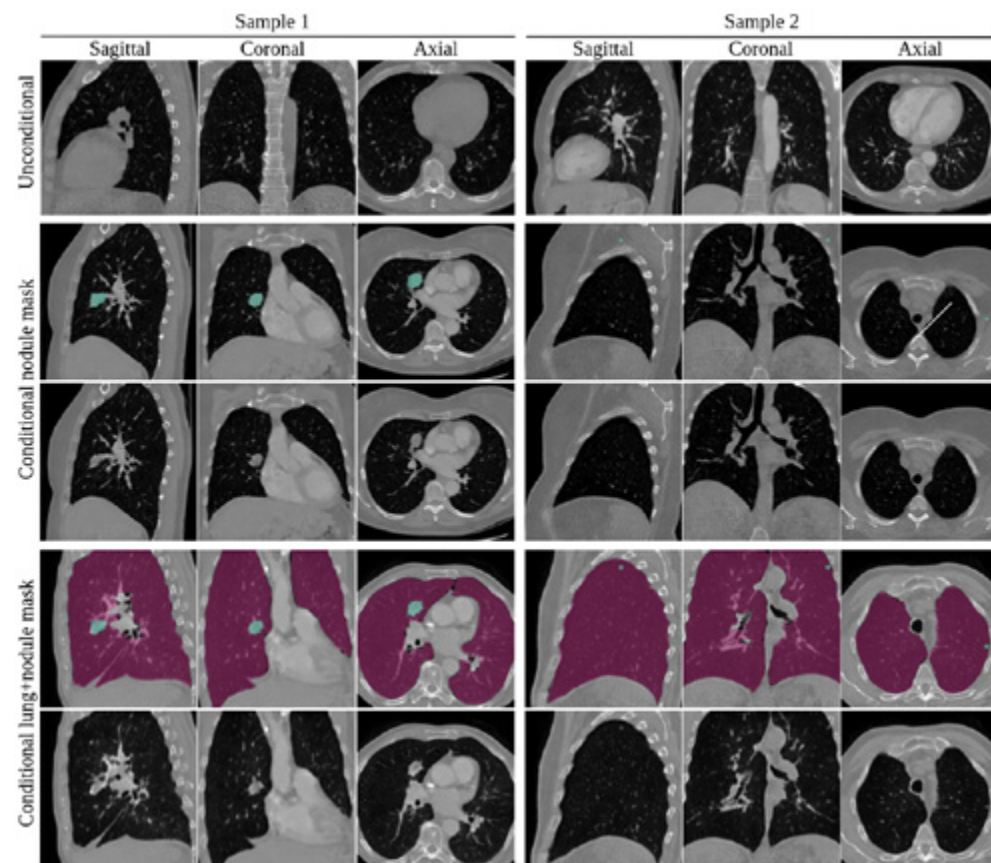
samples were used. Finally, radiomic features (intensity and GLCM texture descriptors [32]) are compared between real and synthetic nodules across different texture scores in crops of size $32 \times 32 \times 32$ masked by the corresponding nodule segmentation.

Implementation Details The diffusion process is performed along $T = 1000$ timesteps and a linear noise schedule ranging from $\beta_1 = 1 \times 10^{-4}$ to $\beta_T = 0.02$. The separate trainings of the 3D VAE and the 3D U-Net were performed on a single NVIDIA GRID A100-20C (20 GB) GPU with batch size 1, AdamW optimizer, and a learning rate of 1×10^{-4} and 1×10^{-5} , respectively. All experiments were trained for 500k optimization steps.

4.2. Experimental Results

Unconditional Synthesis We compare the unconditional LAND pipeline with state-of-the-art methods WDM [9] and PatchDDM [19]. WDM used pre-trained weights, whereas PatchDDM was trained from scratch, as it was originally designed for brain MRI. Table 1 shows that LAND achieves lower FID, likely due to its VAE-based latent compression capturing richer semantic features more aligned with the generative task. Although WDM achieves slightly lower MS-SSIM, LAND's superior FID suggests it produces more realistic outputs without compromising diversity. PatchDDM yields the weakest performance. While trained for 500k steps to match LAND's setup, the 17.5 training days reported in [19] suggest additional iterations. PatchDDM may also suffer from unaligned volumes, where spatial coordinates often map to varying anatomical structures.

Fig. 3. Two examples of synthetic chest CT generated with LAND and masks from NLST. For conditional settings, CT slices are shown with and without mask overlays (nodules in cyan, lungs in magenta). Note how the nodule-only conditional approach generates an anatomically incorrect nodule outside the lung area (right column).

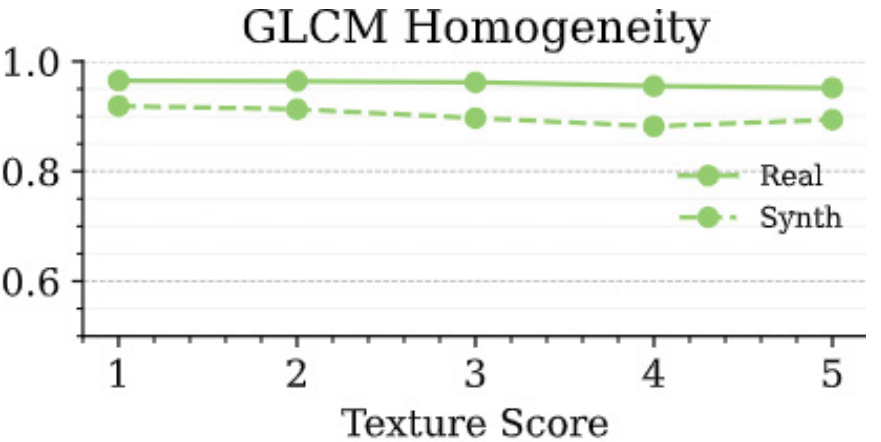
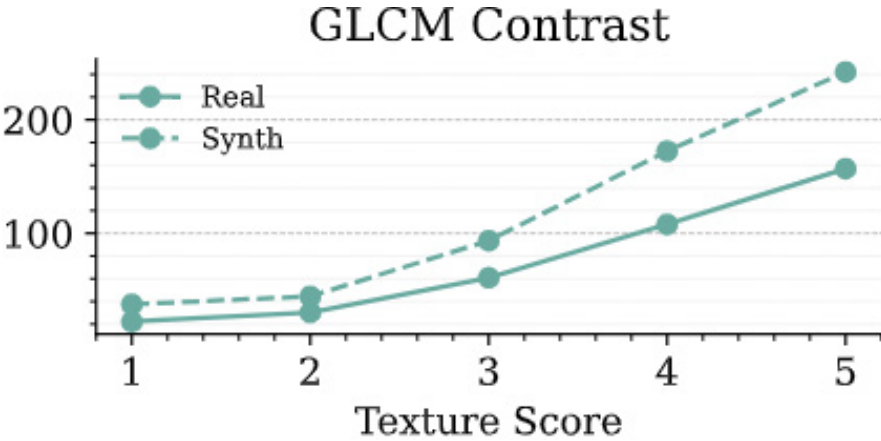
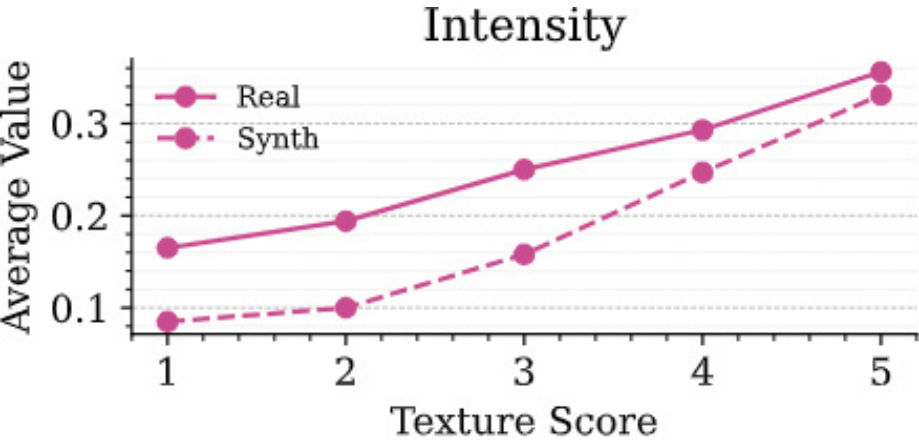


Conditional Synthesis We compare incremental conditional configurations of LAND using anatomical masks: (1) nodules only, (2) lungs and nodules, and (3) lungs and nodules with texture-based nodule values. We run two separate inferences, one with LIDC masks and one with NLST masks, and compute FID using the corresponding dataset. Note that NLST masks were not used during training and MS-SSIM values are computed based on this case. As shown qualitatively in Fig. 3, LAND learns to generate nodules according to the location and size defined by the conditional mask, although nodule shape distortions may arise from mask downsampling. Notably, the synthetic lung shapes closely align with the input mask, and nodules are correctly placed consistently within the lungs only when lung areas are also present in the mask. Incorporating lung areas into the anatomical mask led to improvements in FID scores (Table 1), highlighting the importance of global structural context. In contrast, nodules provide only localized information within the full volume and do not always result in improved FID scores. The FID improvement for NLST with lung+nodule masks suggests that the model has not memorized the LIDC masks used for training and is capable of adapting to novel lung contours. Additionally, MS-SSIM values are similar across LAND configurations, suggesting comparable sample diversity.

Table 1. Evaluation metrics for the proposed experiments (left-to-right): FID scores on the LIDC-IDRI and NLST datasets, average MS-SSIM across 10k pairs of synthetic samples, and inference memory in GB. All FID values are multiplied by 10^3 .

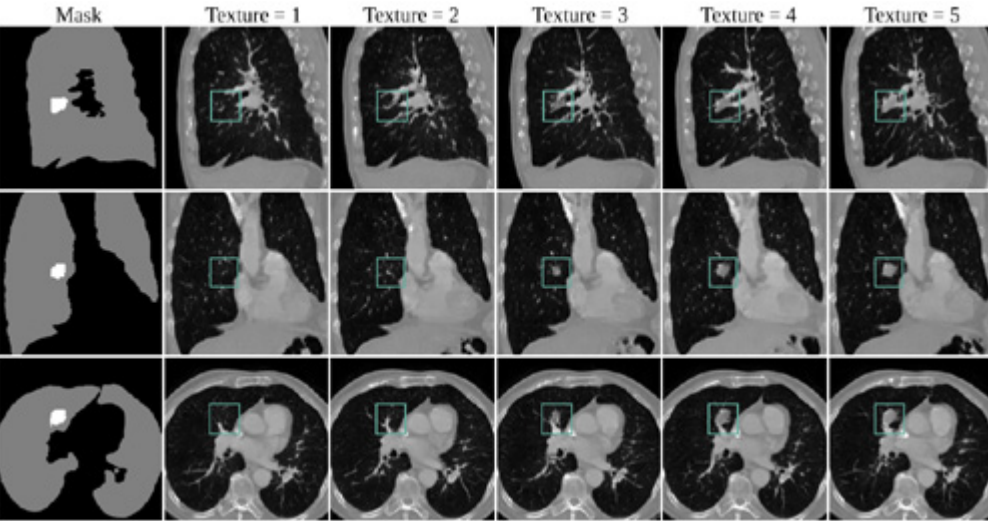
Method		FID↓ (LIDC)	FID↓ (NLST)	MS-SSIM↓	Mem GB↓
Unconditional	PatchDDM [2]	317.534	376.397	0.390	19.61
	WDM [9]	15.236	32.662	0.270	7.27
	LAND Unconditional	5.062	4.759	0.294	7.38
Conditional	LAND nodule mask	4.518	5.821	0.297	7.52
	LAND lung+nodule mask	4.475	3.373	0.290	7.52
	LAND lung+nodule+texture mask	4.603	3.865	0.289	7.52

Fig. 4. Average intensity and GLCM texture features comparing real and synthetic nodules (generated with NLST masks) across different texture scores.



As shown in Fig. 4, comparing local intensity and GLCM texture features between real and synthetic nodule crops reveals that synthetic nodules have higher contrast and slightly lower homogeneity and intensity. However, both sets follow similar trends, where contrast and intensity increase along the texture score and homogeneity remains quite stable, indicating that the model learns to capture expected texture characteristics. To qualitatively evaluate the texture synthesis, multiple samples were generated by varying nodule texture scores on the same anatomical mask. Fig. 5 shows that LAND produces increasingly solid nodules according to the texture score.

Fig. 5. LAND enables the synthesis of nodules with varying texture, with texture scores ranging from 1 (non-solid) to 5 (solid). Bounding boxes indicate the target nodule location as specified by the anatomical mask.

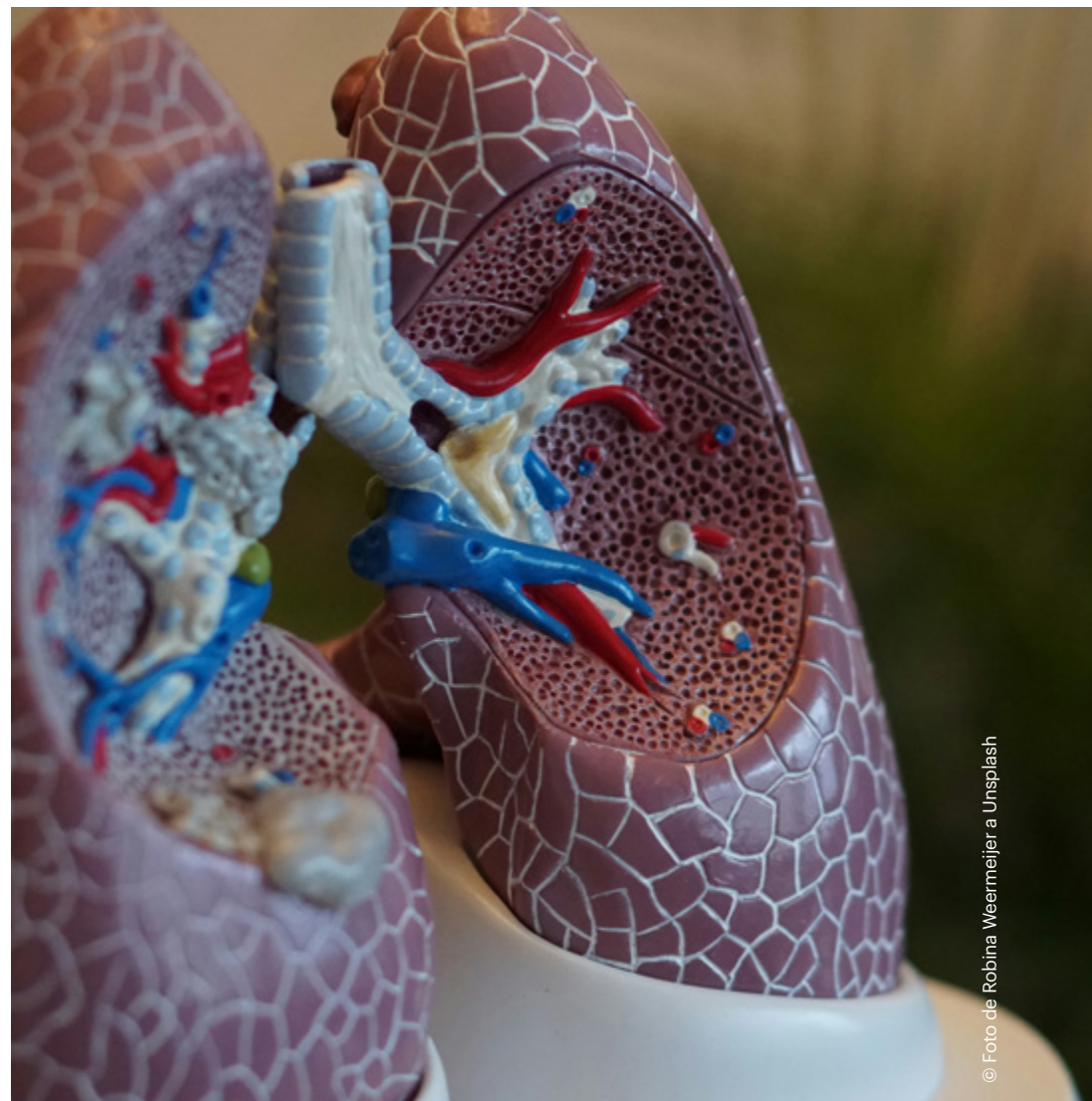


Additional Remarks and Limitations While WDM and LAND demand similar inference memory (Table 1), only LAND trains within a 20 GB budget. Note that WDM performance in Table 1 differs from [9], which may be related to differences in the number of samples used for FID computation. Therefore, relative score differences are likely more informative than absolute FID values. As for limitations, LAND is currently restricted to chest CT and does not support multi-modal or multi-organ synthesis. It also uses 1000 timesteps for inference (same as training), leading to synthesis times on the order of minutes.

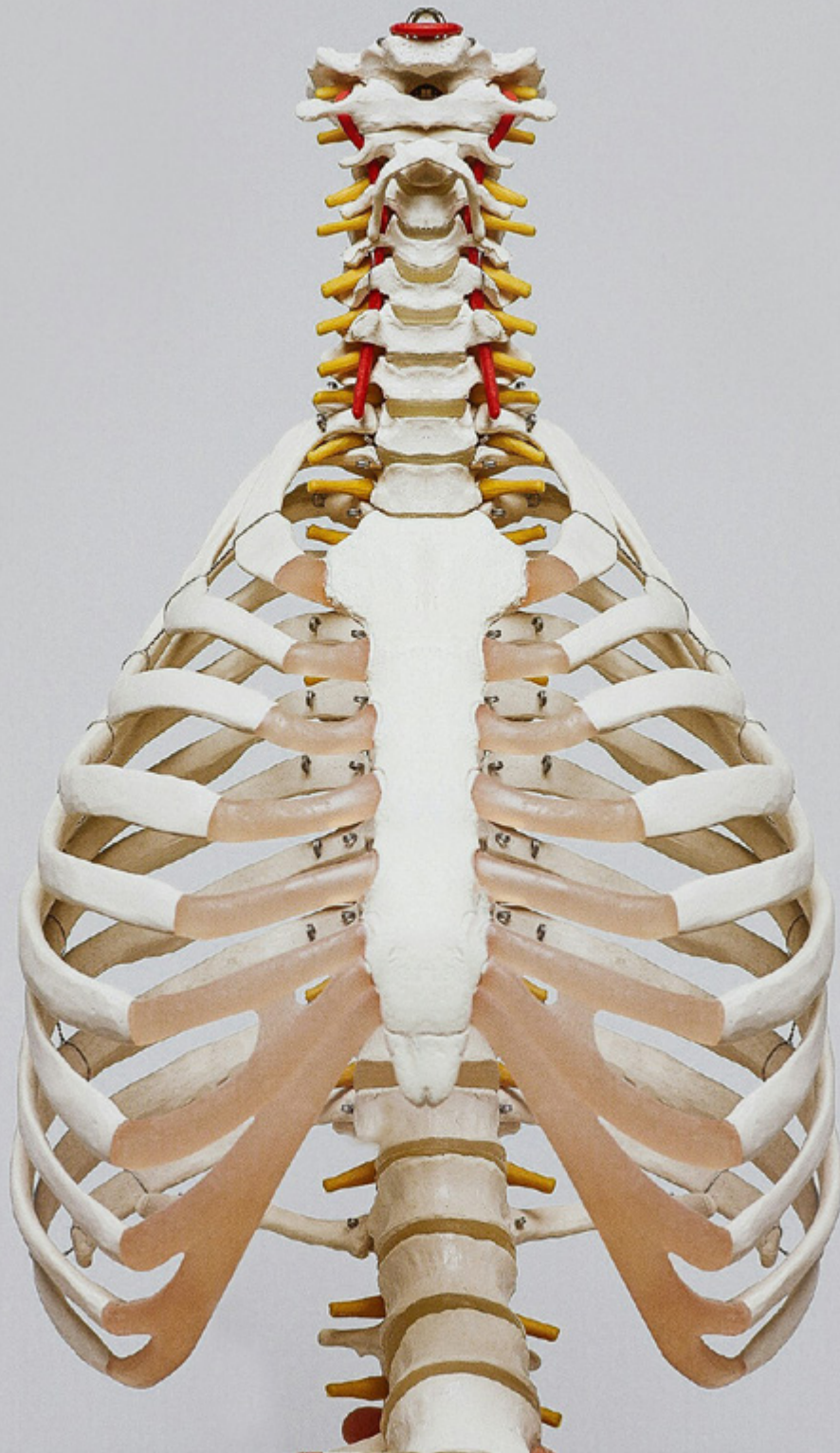
5. Conclusions and Future Work

This work introduces LAND, a latent diffusion model for synthesizing high-quality chest CT volumes conditioned on anatomical masks. LAND enables fine-grained control over lung and nodules, including location and texture, and operates efficiently on a single mid-range GPU, significantly lowering the computational training cost of concurrent methods. By enabling the generation of diverse and realistic medical imaging data, our method holds promise for advancing AI-driven research and clinical applications in which real data is limited.

Future work will evaluate the effectiveness of synthetic images in downstream tasks such as classification and segmentation of chest CT scans. We also aim to extend LAND with clinically relevant nodule features beyond texture—such as spicularity, sphericity and margins—and incorporate a mask generation module to enable more diverse anatomical variations and reduce dependence on predefined masks.



© Foto de Robina Weermeijer a Unsplash



6. References

1. Kazerouni, A., Aghdam, E.K., Heidari, M., et al.: Diffusion models in Medical imaging: A comprehensive survey. *Medical Image Analysis* p. 102846 (2023)
2. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. *NeurIPS* 33,6840–6851 (2020)
3. Bray, F., Laversanne, M., Sung, H., et al.: Global cancer statistics 2022: GLOBO-CAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* 74(3), 229–263 (2024)
4. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013)
5. Müller-Franzes, G., Niehues, J.M., Khader, F., et al.: A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis. *Scientific Reports* 13(1), 12098 (2023)
6. Dhariwal, P., Nichol, A.: Diffusion models beat GANs on image synthesis. *NeurIPS* 34, 8780–8794 (2021)
7. Marí Molas, R., Subías-Beltrán, P., Pitarch Abaigar, C., Galofré Cardo, M., Redondo Tejedor, R.: Characterization of synthetic lung nodules in conditional latent diffusion of chest CT scans. In: *Artificial Intelligence Research and Development*, pp. 44–51 (2024)
8. Ma, Z., Zhang, Y., Jia, G., et al.: Efficient diffusion models: A comprehensive survey from principles to practices. *TPAMI* (2025)
9. Friedrich, P., Wolleb, J., Bieder, F., Durrer, A., Cattin, P.C.: WDM: 3D wavelet diffusion models for high-resolution medical image synthesis. In: *MICCAI Workshops*. pp. 11–21 (2024)

10. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. *NeurIPS* 27 (2014)
11. Guo, P., Zhao, C., Yang, D., et al.: MAISI: Medical AI for synthetic imaging. In: *WACV*. pp. 4430–4441 (2025)
12. Guo, P., Zhao, C., Yang, D., et al.: Text2CT: Towards 3D CT volume generation from free-text descriptions using diffusion model. *arXiv preprint arXiv:2505.04522* (2025)
13. Hamamci, I.E., Er, S., Sekuboyina, A., et al.: GenerateCT: Text-conditional generation of 3D chest CT volumes. In: *ECCV*. pp. 126–143 (2024)
14. Han, K., Xiong, Y., You, C., Khosravi, P., Sun, S., Yan, X., Duncan, J.S., Xie, X.: MedGen3D: A deep generative framework for paired 3D image and mask generation. In: *MICCAI*. pp. 759–769 (2023)
15. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: *CVPR*. pp. 10684–10695 (2022)
16. Pinaya, W.H., Tudosiu, P.D., Dafflon, J., Da Costa, P.F., Fernandez, V., Nachev, P., Ourselin, S., Cardoso, M.J.: Brain imaging generation with latent diffusion models. In: *MICCAI Workshops*. pp. 117–126 (2022)
17. Wang, J., Levman, J., Pinaya, W.H.L., Tudosiu, P.D., Cardoso, M.J., Marinescu, R.: InverseSR: 3D brain MRI super-resolution using a latent diffusion model. In: *MICCAI*. pp. 438–447 (2023)
18. Ho, J., Chan, W., Saharia, C., et al.: Imagen Video: High definition video generation with diffusion models. *arXiv preprint arXiv:2210.02303* (2022)
19. Bieder, F., Wolleb, J., Durrer, A., Sandkuehler, R., Cattin, P.C.: Memory-efficient 3D denoising diffusion models for medical image processing. In: *Medical Imaging with Deep Learning*. pp. 552–567 (2024)
20. Hofmanninger, J., Prayer, F., Pan, J., Röhrich, S., Prosch, H., Langs, G.: Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. *European Radiology Experimental* 4, 1–13 (2020)
21. Armato III, Samuel G., McLennan, G., Bidaut, L., et al.: Data from LIDC-IDRI (2015), <https://www.cancerimagingarchive.net/collection/lidc-idri/>
22. Khader, F., Müller-Franzes, G., Tayebi Arasteh, S., et al.: Denoising diffusion probabilistic models for 3D medical image generation. *Scientific Reports* 13(1) (2023)
23. Kim, J., Park, H.: Adaptive latent diffusion model for 3D medical image to image translation: Multi-modal magnetic resonance imaging study. In: *WACV*. pp. 7604–7613 (2024)
24. Chen, S., Ma, K., Zheng, Y.: Med3D: Transfer learning for 3D medical image analysis. *arXiv preprint arXiv:1904.00625* (2019)
25. Liu, C., Yuan, X., Yu, Z., Wang, Y.: TexDC: Text-driven disease-aware 4D cardiac cine MRI images generation. In: *ACCV*. pp. 3005–3021 (2024)

26. Durrer, A., Wolleb, J., Bieder, F., et al.: Denoising diffusion models for 3D healthy brain tissue inpainting. In: MICCAI Workshops. pp. 87–97 (2024)
27. Dorjsembe, Z., Pao, H.K., Odonchimed, S., Xiao, F.: Conditional diffusion models for semantic 3D brain MRI synthesis. JBHI (2024)
28. Mikhael, P.G., Wohlwend, J., Yala, A., et al.: Sybil: A validated deep learning model to predict future lung cancer risk from a single low-dose chest computed tomography. Journal of Clinical Oncology pp. JCO-22 (2023)
29. Chung, H., Ryu, D., McCann, M.T., Klasky, M.L., Ye, J.C.: Solving 3D inverse problems using pre-trained 2D diffusion models. In: CVPR. pp. 22542–22551 (2023)
30. National Lung Screening Trial Research Team: Data from the National Lung Screening Trial (NLST) (2013), <https://www.cancerimagingarchive.net/collection/nlst/>
31. Pinaya, W.H., Graham, M.S., Kerfoot, E., et al.: Generative AI for medical imaging: Extending the MONAI framework. arXiv preprint arXiv:2307.15208 (2023)
32. Haralick, R.M., Shanmugam, K., Dinstein, I.H.: Textural features for image classification. IEEE Transactions on systems, man, and cybernetics (6), 610–621 (2007)
33. Hang, T., Gu, S., Li, C., Bao, J., Chen, D., Hu, H., Geng, X., Guo, B.: Efficient diffusion training via Min-SNR weighting strategy. In: ICCV. pp. 7441–7451 (2023)
34. Salimans, T., Ho, J.: Progressive distillation for fast sampling of diffusion models. arXiv preprint arXiv:2202.00512 (2022)
35. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: ICML. pp. 2256–2265 (2015)
36. Volokitin, A., Erdil, E., Karani, N., et al.: Modelling the distribution of 3D brain MRI using a 2D slice VAE. In: MICCAI. pp. 657–666 (2020)
37. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs trained by a two time-scale update rule converge to a local nash equilibrium. NeurIPS 30 (2017)
38. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: ACSSC. vol. 2, pp. 1398–1402 (2003)
39. Xing, Z., Yang, S., Chen, S., et al.: Cross-conditioned diffusion model for Medical image to image translation. In: MICCAI. pp. 201–211 (2024)

Socis



Membres

