

Anatomy-compliant medical image synthesis by latent diffusion models

Nuno Capitão¹

Yidong Zhao¹

Yi Zhang¹

Nathan Geerts¹

João V. Lopes²

Qian Tao^{*1}

N.M.NUNOMIGUELFERREIRACAPITAO@TUDELFT.NL

Y.ZHAO-8@TUDELFT.NL

Y.ZHANG-43@TUDELFT.NL

N.GEERTS@STUDENT.TUDELFT.NL

JLOPES@FC.UP.PT

Q.TAO@TUDELFT.NL

¹ *Delft University of Technology, Lorentzweg 1, Delft, The Netherlands*

² *Department of Physics and Astronomy, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal*

Abstract

Data scarcity presents a significant challenge for achieving optimal model performance in medical imaging, due to the limited availability of high-quality data. One potential solution to address this issue is to synthesize medical images using powerful generative models with conditioning prior. However, obtaining full anatomical annotations of all organs for anatomical conditioning is impractical, resulting in synthetic images with incoherent or hallucinated anatomy. In this paper, we propose an innovative medical image generation method based on state-of-the-art latent diffusion models (LDM). To tackle the anatomy compliance challenge, we leverage both the anatomical mask, which is specific to the organ of interest, and the edge information, which is general and easy to compute in the full field of view (FOV), as dual conditioning. Our method does not require extra annotations to achieve anatomy compliance. Our method was evaluated on the ACDC dataset and compared with GAN baselines. Results demonstrate that incorporating edge-based conditioning strongly complements image semantics, leading to high-quality, anatomy-compliant medical image generation.

Keywords: Latent diffusion models (LDM), Medical image synthesis, Edge.

1. Introduction

Learning-based approaches have gained prominence in many medical image processing tasks (Zhou et al., 2021). However, they rely heavily on high-quality labeled data, which is frequently unavailable given the high acquisition costs, rare pathologies, privacy concerns, and lack of annotation expertise (Castro et al., 2020). Generative AI offers a promising solution to the data scarcity problem (Dorjsembe et al., 2024). Previous medical image synthesis studies predominantly leveraged the SPADE (Park et al., 2019) framework which translates the semantic map to an image with Generative Adversarial Networks (GAN) (Goodfellow et al., 2014). Built upon SPADE, (Abbasi-Sureshjani et al., 2020) proposed generating cardiac magnetic resonance (CMR) images given heart segmentation maps. However, GAN-based generation can have degenerated quality (Müller-Franzes et al., 2023; Skandarani et al., 2021) and has recently been outperformed by diffusion-based generation (Ho et al., 2020; Dhariwal and Nichol, 2021; Rombach et al., 2022). Additionally and importantly, it

is often overlooked that the available semantic maps usually only cover specific organs (i.e. heart) (Abbasi-Sureshjani et al., 2020), leading to spurious and hallucinated generation of surrounding organs in the thorax and abdomen (i.e. lung, liver).

Image edges can delineate anatomical structures with little computational effort and can be used to guide image generation with GANs (Luo et al., 2021; Yu et al., 2019). Nevertheless, its use has not been explored in state-of-the-art diffusion models which promise superior quality to GANs. In this paper, we propose a novel LDM-based generative model conditioned on both incomplete organ annotations and image edges in the full field of view as dual anatomy conditioning. Our method ensures anatomical compliance in medical image generation, which has been understudied but is important for downstream tasks such as training AI models or educating radiologists (Skandarani et al., 2023).

2. Methods and Materials

We propose an LDM-based conditional generator \mathcal{G} (Rombach et al., 2022) to translate anatomy into realistic images, which performs the diffusion process in an auto-encoded (AE) latent space. The anatomy is modeled by combining the partly labeled tissue map y and the image edge e_x . The edge e of an image x is retrieved by Canny detector (Canny, 2009). The conditions are integrated into \mathcal{G} through a binary boundary map derived from the image gradients, concatenated with a one-hot tensor representation of the semantic labels map, as represented in Figure 1.

We trained our models using the publicly available Automated Cardiac Diagnosis Challenge (ACDC) dataset (Bernard et al., 2018), and compared the proposed model to GAN-based conditional generation with SPADE (Park et al., 2019). To validate the utility of the edge as an additional condition, we assessed both models’ performance with and without edge guidance, and evaluate quantitatively using metrics of FID, SSIM, NMSE. Additionally, we compared the Fourier space of generated images by different models.

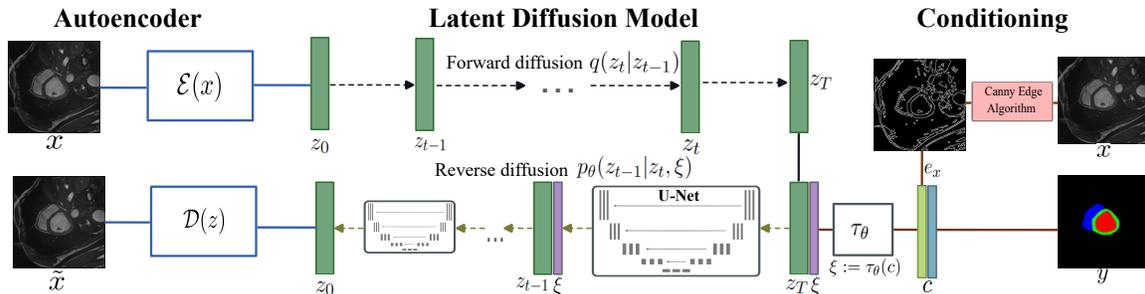


Figure 1: Illustration of the LDM-based edge conditioning pipeline.

3. Results and Conclusion

The quantitative evaluation results are reported in Table 1. The proposed LDM-based generation achieved the best performance in terms of FID (61.9/98.1) compared to the best SPADE variant (101.9/161.4), with or without edge guidance. Our LDM has a lower SSIM

than the SPADE-based model without edge guidance. However, the edge-guided LDM achieved an SSIM of 0.608 which is much higher than the SPADE variant (0.556). Using edge guidance also significantly improved the generation quality for SPADE variants.

Figure 2 presents some examples of generated CMR images and their Fourier frequency spectra. For LDM, we observe that the generated images are closest to the ground truth and correctly delineate organ details that the other models fail to capture. Nevertheless, we note that all edge-informed models show similar improvement. Moreover, we observe that for the edge-informed models, LDM aligns closely with the ground truth in both low and high-frequency ranges, whereas SPADE models only achieve closer alignment in higher frequencies.

Table 1: Image generation performance measured by FID, SSIM and NMSE.

Method	FID ↓	w/ edge conditioning		FID ↓	w/o edge conditioning	
		SSIM ↑	NMSE ↓		SSIM ↑	NMSE ↓
LDM	61.9	0.608 ± 0.123	0.071 ± 0.035	98.1	0.283 ± 0.094	0.136 ± 0.080
SPADE	109.8	0.556 ± 0.119	0.117 ± 0.041	160.3	0.328 ± 0.093	0.132 ± 0.077
SPADE w/ VAE	101.9	0.556 ± 0.128	0.104 ± 0.034	161.4	0.338 ± 0.090	0.145 ± 0.093

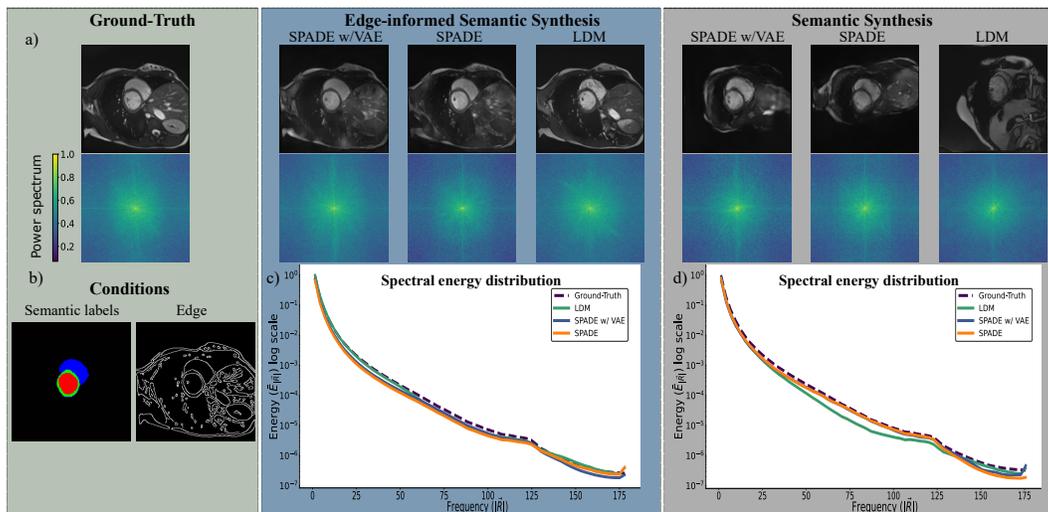


Figure 2: Qualitative results: (a) The ground truth and its Fourier spectrum. (b) Left: Segmentation map. Right: Edge anatomical boundaries by the Canny operator. (c),(d) Comparison of the average model spectra as a function of $|\vec{\mathbf{R}}| = \sqrt{x^2 + y^2}$, representing a radius distance from the 2D grid origin.

In conclusion, we proposed a novel medical image generation method using LDM with dual conditioning by both semantic labels and Canny edges. Our experiments showed that our method improved image synthesis quality and anatomy compliance, without requiring additional exhaustive annotations. Furthermore, we have shown that LDM outperforms GAN-based models in both the image and frequency space.

References

- Samaneh Abbasi-Sureshjani, Sina Amirrajab, Cristian Lorenz, Juergen Weese, Josien Pluim, and Marcel Breeuwer. 4d semantic cardiac magnetic resonance image synthesis on xcat anatomical model. In Tal Arbel, Ismail Ben Ayed, Marleen de Bruijne, Maxime Descoteaux, Herve Lombaert, and Christopher Pal, editors, *Proceedings of the Third Conference on Medical Imaging with Deep Learning*, number 121 in Proceedings of Machine Learning Research, pages 6–18. PMLR, February 2020. Third Conference on Medical Imaging with Deep Learning ; Conference date: 06-07-2020 Through 08-07-2020.
- Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, Gerard Sanroma, Sandy Napel, Steffen Petersen, Georgios Tziritas, Elias Grinias, Mahendra Khened, Varghese Alex Kollerathu, Ganapathy Krishnamurthi, Marc-Michel Rohe, Xavier Pennec, Maxime Sermesant, Fabian Isensee, Paul Jager, Klaus H Maier-Hein, Peter M Full, Ivo Wolf, Sandy Engelhardt, Christian F Baumgartner, Lisa M Koch, Jelmer M Wolterink, Ivana Isgum, Yeonggul Jang, Yoonmi Hong, Jay Patravali, Shubham Jain, Olivier Humbert, and Pierre-Marc Jodoin. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Trans. Med. Imaging*, 37(11):2514–2525, November 2018.
- John F. Canny. Canny edge detection. 2009. URL <https://api.semanticscholar.org/CorpusID:47899497>.
- Daniel C. Castro, Ian Walker, and Ben Glocker. Causality matters in medical imaging. *Nature Communications*, 11(1):3673, Jul 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-17478-w. URL <https://doi.org/10.1038/s41467-020-17478-w>.
- Prafulla Dhariwal and Alex Nichol. Diffusion models beat gans on image synthesis, 2021.
- Zolnamar Dorjsembe, Hsing-Kuo Pao, Sotavilan Odonchimed, and Furen Xiao. Conditional diffusion models for semantic 3d brain mri synthesis, 2024.
- Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.
- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020.
- Yanmei Luo, Dong Nie, Bo Zhan, Zhiang Li, Xi Wu, Jiliu Zhou, Yan Wang, and Ding-gang Shen. Edge-preserving mri image synthesis via adversarial network with iterative multi-scale fusion. *Neurocomputing*, 452:63–77, 2021. ISSN 0925-2312. doi: <https://doi.org/10.1016/j.neucom.2021.04.060>. URL <https://www.sciencedirect.com/science/article/pii/S0925231221006020>.
- Gustav Müller-Franzes, Jan Moritz Niehues, Firas Khader, Soroosh Tayebi Arasteh, Christoph Haarbuerger, Christiane Kuhl, Tianci Wang, Tianyu Han, Teresa Nolte, Sven Nebelung, 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.

- Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization, 2019.
- Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, 2022.
- Youssef Skandarani, Pierre-Marc Jodoin, and Alain Lalande. Gans for medical image synthesis: An empirical study, 2021.
- Youssef Skandarani, Pierre-Marc Jodoin, and Alain Lalande. Gans for medical image synthesis: An empirical study. *Journal of Imaging*, 9(3):69, March 2023. ISSN 2313-433X. doi: 10.3390/jimaging9030069.
- Biting Yu, Luping Zhou, Lei Wang, Yinghuan Shi, Jurgen Fripp, and Pierrick Bourgeat. Eagan: Edge-aware generative adversarial networks for cross-modality mr image synthesis. *IEEE Transactions on Medical Imaging*, 38(7):1750–1762, 2019. doi: 10.1109/TMI.2019.2895894.
- S. Kevin Zhou, Hayit Greenspan, Christos Davatzikos, James S. Duncan, Bram Van Ginneken, Anant Madabhushi, Jerry L. Prince, Daniel Rueckert, and Ronald M. Summers. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, 109(5): 820–838, 2021. doi: 10.1109/JPROC.2021.3054390.