

Deep Generative Models for Medical Image Synthesis

Charlotte Bennett

A thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
of University College London

Department of Medical Physics and Biomedical Engineering
University College London

2026

Declaration

I, Charlotte Bennett, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Medical imaging datasets are small, unevenly labelled and hard to share. Generative models that synthesize realistic images could augment training sets and translate between imaging modalities, but only if the synthetic images preserve the anatomy that clinical decisions rely on. This thesis develops and evaluates generative models for magnetic resonance imaging with that requirement in mind. It proposes a diffusion model conditioned on segmentation maps, so that synthetic anatomy can be specified rather than hoped for, and introduces an evaluation protocol that measures anatomical fidelity with a segmentation network rather than with image statistics alone. Augmenting a small training set with conditioned synthetic images raises the Dice score of a downstream tumour segmentation model from 0.71 to 0.78.

Impact Statement

Hospitals rarely have enough labelled scans to train reliable software, and privacy rules make pooling scans between hospitals difficult. The methods in this thesis create realistic synthetic scans whose anatomy is controlled, so that small datasets can be enlarged without sharing patient data. Within academia, the evaluation protocol gives other researchers a way to check whether synthetic images are anatomically faithful, not merely convincing. Outside academia, the approach could shorten the time needed to build segmentation tools for rare conditions, where scans are scarcest. Any clinical use would first need prospective validation and regulatory review.

Acknowledgements

I thank my supervisors for their support and their careful reading of every chapter. I am grateful to my colleagues in the medical imaging group and to the clinical collaborators whose questions kept this work pointed at real problems.

Contents

Abstract	iii
Impact Statement	iv
Acknowledgements	v
1 Introduction	1
2 Background	2
3 Anatomy-Conditioned Diffusion	3
4 Evaluation	4
5 Conclusion	5
A Scanner and Preprocessing Details	6

List of Figures

3.1	Generation is conditioned on anatomy and evaluated on anatomy.	3
-----	--	---

List of Tables

4.1 Tumour segmentation Dice on held-out real scans.	4
--	---

Chapter 1

Introduction

Deep learning has transformed medical image analysis, but it needs labelled data that medicine struggles to provide. Generative models offer a partial remedy. Adversarial networks [1] and, more recently, denoising diffusion models [2] can synthesize images that are hard to tell from real ones.

Looking real is not enough in medicine. A synthetic brain scan with a plausible but wrong ventricle shape could teach a model the wrong anatomy. This thesis therefore conditions generation on anatomy and evaluates it on anatomy.

Chapter 2

Background

A denoising diffusion model learns to reverse a fixed noising process that turns an image x_0 into Gaussian noise over T steps. It is trained by predicting the added noise,

$$\mathcal{L} = \mathbb{E}_{t, x_0, \epsilon} \left\| \epsilon - \epsilon_\theta(x_t, t, c) \right\|^2, \quad (2.1)$$

where c is an optional condition [2]. The denoiser is typically a U-Net [3], the same architecture widely used for medical image segmentation.

Chapter 3

Anatomy-Conditioned Diffusion

We condition the denoiser on a segmentation map c injected at every resolution through spatially adaptive normalization, so that each synthetic image follows the given anatomy. Label maps can be taken from real scans or edited to create new configurations, for example to move or resize a tumour.

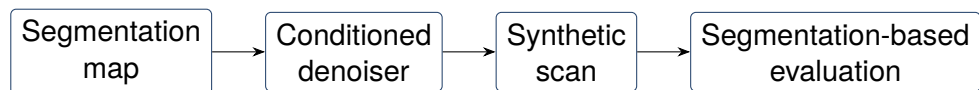


Figure 3.1: Generation is conditioned on anatomy and evaluated on anatomy.

Chapter 4

Evaluation

We trained a tumour segmentation network on 40 real scans alone and with added synthetic scans. Table 4.1 reports the Dice score on 120 held-out real scans.

Table 4.1: Tumour segmentation Dice on held-out real scans.

Training data	Dice
40 real scans	0.71
+ 400 unconditioned synthetic scans	0.72
+ 400 anatomy-conditioned synthetic scans	0.78

Unconditioned synthesis barely helps: its images are realistic but their tumours do not match any label. Conditioned synthesis supplies correctly labelled variety.

Chapter 5

Conclusion

For medical images, generative models are most useful when anatomy is an input rather than an accident. Conditioning on segmentation maps makes synthetic data trustworthy enough to help a downstream model, and evaluating with segmentation catches failures that image statistics miss.

Appendix A

Scanner and Preprocessing Details

All scans were resampled to 1 mm isotropic resolution, intensity-normalized per volume, and cropped to a common field of view before training.

Bibliography

- [1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, volume 27, 2014.
- [2] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851, 2020.
- [3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015.