Fast and Clinically Relevant Radiotherapy Dose Estimation with Deep Learning

Joseph Weibel

Supervisors: Prof. Dr. Mauricio Reyes, Amith Kamath PhD
Institution: University of Bern, ARTORG Center for Biomedical Engineering Research

Introduction

Radiotherapy (RT) treatment planning is a labour-intensive, iterative process requiring hours of manual optimisation. Faster dose estimation would enable real-time feedback during contouring, reduce planning bottlenecks, and is essential for adaptive RT, where plans must be regenerated in response to anatomical changes. Deep learning offers a promising route to clinically deployable inference speeds, yet existing literature conflates two distinct tasks, dose prediction and dose calculation, applying uniform evaluation protocols inappropriate to each task's distinct requirements.

Materials and Methods

This work presents a systematic benchmarking of deep learning architectures for RT dose estimation on the GDP-HMM head-and-neck (HaN) Volumetric Modulated Arc Therapy dataset, comparing U-Net variants, Denoising Diffusion Probabilistic Models (DDPM), Rectified Flow Matching (RFM), and Inversion by Direct Iteration. A central contribution is the proposal of separate, clinically grounded evaluation frameworks for dose prediction and dose calculation, developed in collaboration with medical physicists, using metrics tailored to each task's requirements.

Results

The benchmark demonstrates that diffusion-based models are strong performers across both tasks, with BeamDiff emerging as the strongest overall model and SP-DiffDose excelling particularly for dose calculation. RFM is found to be superior to DDPM for both tasks, offering better quality at fewer sampling steps and converging to a single-step solution for dose calculation under strong fluence conditioning. The analysis reveals that BeamDiff's dedicated beam conditioning pathway is ineffective in the HaN setting studied, and that the absence of meaningful inter-sample variance suggests current models do not fully capture the stochastic structure of dose prediction, likely attributable to single-plan supervision.

A systematic exploration of acceleration strategies yields clinically viable inference speeds: RFM at a single sampling step achieves a 3464x speedup for dose calculation, reducing root mean squared error from 2.52 Gy to 2.44 Gy compared to the unaccelerated DDPM baseline. Hybrid approaches combining U-Net initialisation with RFM prove particularly effective for dose prediction, with Warm-Start RFM at two sampling steps combined with architecture optimisation, caching, and half-precision sampling achieving sub-0.5-second per-slice inference without sacrificing clinical metric performance for the majority of patients. Expert evaluation confirms that predicted distributions are sufficient for early-stage clinical decision-making, with organs-at-risk dose-volume histogram curves rated as closely matching ground truth.

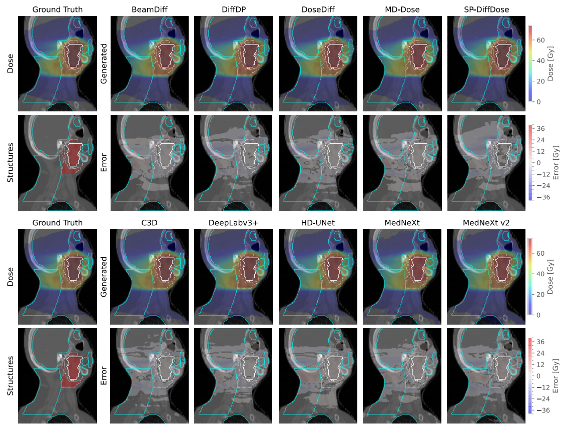

Fig. 1 Qualitative comparison of predicted dose distributions and voxel-wise error maps. Each panel shows the generated dose overlaid on CT with the PTV (white) and OAR (cyan) contours in top rows and the corresponding error relative to ground truth in bottom rows, with discrete colours blue and red indicating over- and under-prediction, respectively, no shade indicating no difference. DDPMs are shown in the top panel and U-Nets in the bottom panel.

Discussion

These results establish flow-based generative models as strong performers for RT dose estimation, demonstrate that clinically viable inference speeds are achievable through different acceleration techniques including hybrid U-Net and RFM architectures, and provide a principled evaluation framework that distinguishes between tasks whose distinct clinical requirements have previously been conflated in literature.

UNIVERSITÄT BERN