Helmholtz Imaging Annual Conference 2025



Contribution ID: 57

Type: not specified

Unsupervised Temporal Diffusion-Based Interpolation for 4D CT

Unsupervised Temporal Diffusion-Based Interpolation for 4D CT

Authors:

Zeyad Mahmoud¹, Anna Reithmeir¹²⁴, Julia A. Schnabel¹²³⁴, Daniel M. Lang¹²

Affiliations:

¹ Institute of Machine Learning in Biomedical Imaging, Helmholtz Center Munich, Munich, Germany

² School of Computation, Information and Technology, Technical University of Munich, Munich, Germany

³ School of Biomedical Engineering & Imaging Sciences, King's College London, London, UK

⁴ Munich Center for Machine Learning, Munich, Germany

Abstract

4D Spatio-temporal computed tomography (CT) plays a critical role in radiation therapy planning by capturing patient-specific respiratory motion that is essential to minimize doses to healthy tissues and organs at risk. However, repeated temporal image acquisition increases radiation exposure to patients and features high costs and acquisition times. Learning-based methods offer a promising solution to reduce the number of acquisitions needed to capture dynamic patient motion.

However, synthesizing realistic 4D CT sequences remains highly challenging, as the task demands strict preservation of anatomical structures and temporal coherence across consecutive time-points. To achieve this, we aim to synthesize images from intermediate time points through diffusion-based generative modeling. We extend the diffusion-based MAISI framework of Guo et al., which was originally designed to generate static CT images conditioned solely on segmentation masks, to 4D CT generation. Specifically, we propose a novel architecture based on a modified ControlNet structure that incorporates dual-conditioning, leveraging both the segmentation mask at time t_2 and the CT scan from a preceding time point t_1 . The architecture is trained on the LIDC-IDRI dataset, segmentation masks were generated using the TotalSegmentator and VISTA3D frameworks, and synthetic elastic deformations were applied to simulate motion. This results in a training dataset composed of triples (CT at t_1 , segmentation mask at t_2 , CT at t_2), enabling the framework to generate the CT scan at t_2 while effectively capturing temporal dynamics and preserving anatomical integrity across the 4D sequence. The dual-conditioning approach significantly improves anatomical accuracy and ensures smoother temporal transitions between time-points.

Experimental results demonstrate superior performance of the proposed method over the baseline MAISI implementation. Quantitatively, we observe a substantial reduction in normalized mean squared error (NMSE \downarrow) from 0.039 to 0.034, an increase in structural similarity index measure (SSIM \uparrow) from 0.70 to 0.73, an improvement in peak signal-to-noise ratio (PSNR \uparrow) from 16.8 to 17.5, and a decrease in learned perceptual image patch similarity (LPIPS \downarrow) from 0.19 to 0.16. Qualitatively, the generated CT volumes exhibit enhanced anatomical realism, with smoother motion transitions and improved consistency between consecutive time-points.

In conclusion, this study shows the effectiveness of dual-conditioning with CT scans and segmentation masks for dynamic 4D CT generation. The proposed enhancements deliver superior anatomical fidelity and temporal coherence compared to previous approaches. In future work, we aim to extend this framework to more complex motion patterns and evaluate its robustness across broader medical imaging modalities. Scaling the model with larger datasets could further strengthen its performance and generalizability.

Keywords

AI for image analysis

Primary authors: Ms REITHMEIR, Anna (TUM); LANG, Daniel (Technical University of Munich); Prof. SCHNABEL, Julia (Helmholtz); MAHMOUD, Zeyad Abdelfattah Zaki

Session Classification: AI for Image Analysis Part 1