

The Development of a label-to-image latent diffusion model for patient-derived digital twin generation using pelvis radiotherapy CT simulation images

René E. Rodríguez Pérez¹, Gerardo U. Pérez Rojas¹, Adair Maldonado Luna¹, Benito de Celis Alonso¹, Marvin Kinz², Kamal Singhrao^{2*}

(1) Faculty of Mathematical & Physical Sciences, Meritorious Autonomous University of Puebla, Puebla, Puebla, Mexico.

(2) Department of Radiation Oncology, Brigham and Women's Hospital, Dana Farber Cancer Institute, Harvard Medical School, Brookline, Massachusetts, USA.

* Corresponding author. Email: ksinghrao@bwh.harvard.edu

Purpose

Anatomic motion and deformation modeling using digital twins (DTs) hold much promise for radiotherapy applications¹. However, clinical implementation is limited by availability of high-quality data, computationally intensive and unphysical anatomical simulations.

To address these limitations, we propose a modality-agnostic latent diffusion model for synthetic DT image generation conditioned on semantic label maps to preserve geometric consistency is proposed. Validated by CT-alike coronal pelvic slices, generated from bone masks, confirming geometric consistency.

This study serves as an approach to create geometry-constrained generative twins, potentially surrogating a complex biomechanical simulation, restrained by patient-specific anatomical tissue geometry. This might reduce time for fundamental tasks such as contour propagation in online adaptive radiotherapy and allowing to mitigate chronic issues such as data scarcity.

Methods

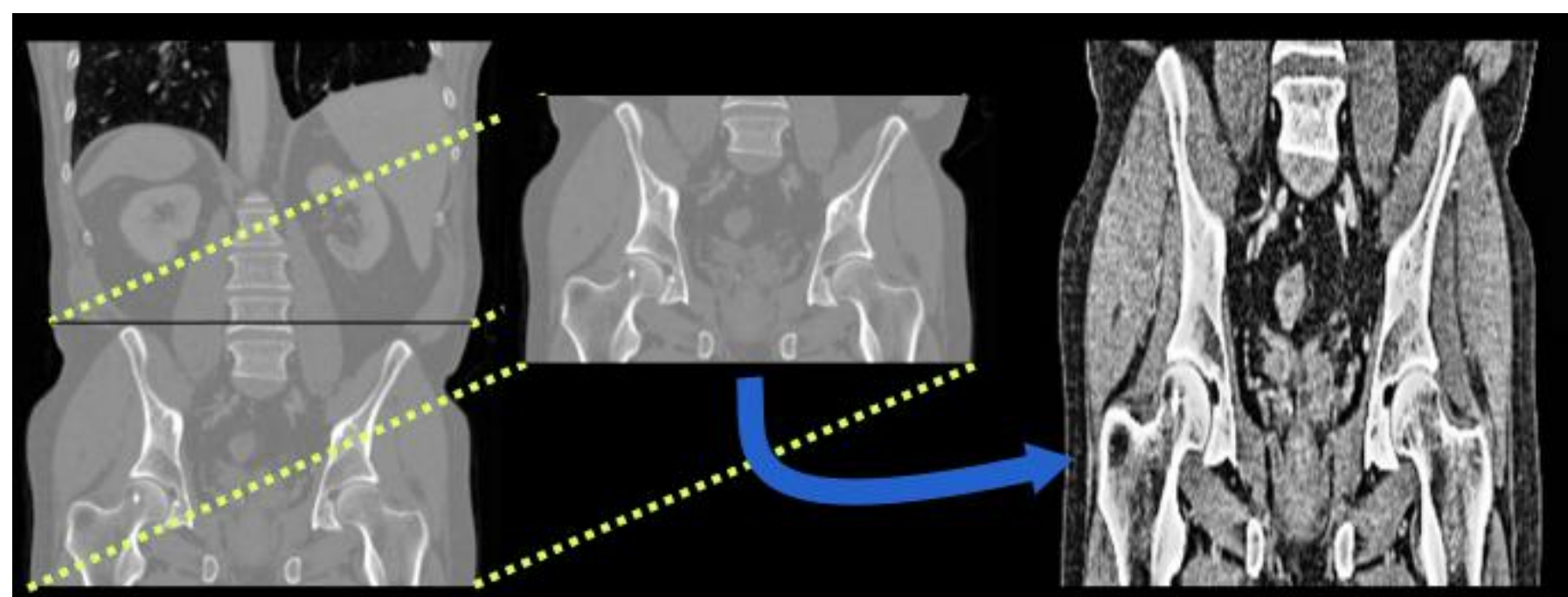
A publicly available dataset of 87 patient-derived CT scans (76/11 split) was processed using a fused window approach. Where a double window was applied to CT volumes: One for bone (W:1800, L:400) and another for soft tissue (W:350, L:50) to optimize Hounsfield units.

Coronal slices were extracted from the pelvic bone mask-defined region for both CT and mask volumes. Model preprocessing included CT image resampling and renormalization.

CT-Mask pairs were used to fine-tune a hybrid Stable Diffusion v1.5-ControlNet model for 16 epochs. ControlNet was initialized from a pre-trained Canny edge model and fully fine-tuned using bone masks to enforce geometric constraints. Cross-attention layers and UNet up-blocks were adjusted to achieve texture fidelity.

Forty-four synthetic images from different slices and positions were generated. Bones were manually segmented without knowledge of the conditioning mask. Geometric consistency was evaluated using the Dice coefficient between ground-truth and segmented bone masks.

Figure 1. Fused window result. Comparison between initial and preprocessed slice to train the Stable Diffusion model after applying the fused window approach, optimizing the Hounsfield unit range.



Results

The study found a mean Dice coefficient of 0.78 ± 0.06 across 44 generated coronal slices with a maximum Dice coefficient value of 0.90. Slice generation took an average of 253.4 seconds employing 35 inference steps.

Generated slices demonstrated morphological and spatial preservation for constraint tissues due to ControlNet intervention regulating hallucination from Stable Diffusion. Proving the adaptability of general-purpose models to medical applications with limited datasets by fine-tuning specific layers.

The generative model can create synthetic slices in a couple of minutes, in contrast to standard biomechanical simulations that can last from hours to days.

Dice coefficient value	0.78 ± 0.06
Maximum Dice coefficient value	0.90
Generation time per slice	253.4 s

Table 1. Results summary. Dice coefficient results from generated slices. Generation time per slice is included.

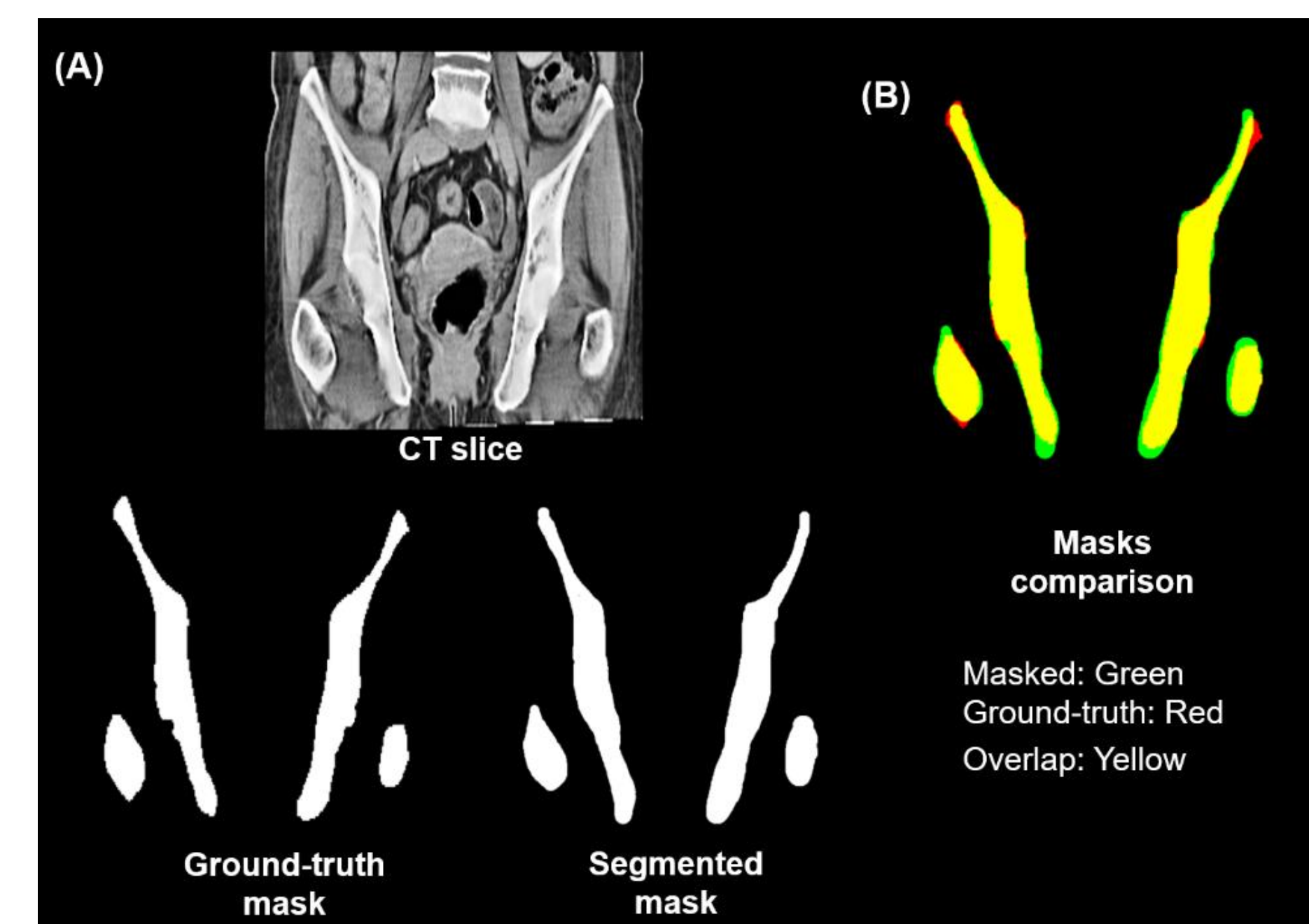


Figure 2. Visual comparison of results. A) Comparative of CT image with its corresponding ground-truth mask and the segmented bone mask from generated image. B) Overlap representation between masked and ground-truth bones.

Conclusions

The model has proven to preserve bone structure, replicate CT noise patterns and its capability to create slice images faster than BM model approaches. Future work will incorporate a multi-label map to control soft tissue consistency and sCT texture will be refined.

Citations

- [1] Chaudhuri, A., Pash, G., Hormuth, D. A., Lorenzo, G., Kapteyn, M., Wu, C., Lima, E. A. B., Yankeelov, T. E., & Willcox, K. (2023). Predictive digital twin for optimizing patient-specific radiotherapy regimens under uncertainty in high-grade gliomas. *Frontiers In Artificial Intelligence*, 6, 1222612. <https://doi.org/10.3389/frai.2023.1222612>
- [2] Meijer, C., Uh, H., & Bouhaddani, S. E. (2023). Digital Twins in Healthcare: Methodological Challenges and Opportunities. *Journal Of Personalized Medicine*, 13(10), 1522. <https://doi.org/10.3390/jpm13101522>
- [3] Wasserthal, J. (2023). Dataset with segmentations of 117 important anatomical structures in 1228 CT images [Conjunto de datos]. En Zenodo (CERN European Organization for Nuclear Research). <https://doi.org/10.5281/zenodo.10047292>

Acknowledgements

The authors would like to thank the Harvard Joint Center for Radiation Therapy Foundation Grant which was used in part for the completion of this work.