CBCT Liver and Lesion Segmentation Using CT Fine-Tuned Foundation Model for CBCT-guided Radiotherapy

Ryan Sanford[1,2], Mojtaba Safari[1,4], Justin Roper[1], Chih-Wei Chang[1], Pretesh Patel[1], Richard Qiu[1], Yue Chen[3], Jun Zhou[1] and Xiaofeng Yang[1,2,3,4]

[1]Department of Radiation Oncology and Winship Cancer Institute, Emory University, Atlanta, GA USA
[2]Medical Physics Graduate Program, Georgia Institute of Technology, Atlanta USA
[3]Biomedical Engineering, Georgia Institute of Technology / Emory University, Atlanta USA
[4]Department of Radiation and Cellular Oncology, The University of Chicago, Chicago, IL USA

2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

Introduction

- CBCT is integral to image-guided radiation therapy and adaptive radiotherapy (ART) workflows, whose constrained timeframe makes on-treatment auto-segmentation especially valuable.[1]
- Poor soft-tissue contrast and imaging artifacts inherent to CBCT limit its quantitative use.
- Models trained on conventional CT transfer poorly to CBCT due to domain mismatch.
- Vision foundation models offer superior performance and generalizability in the medical imaging domain yet remain largely unexplored for CBCT.
- **This work** investigates a CT-fine-tuned foundation model[2] as an initializer for CBCT segmentation, and whether domain-aligned fine-tuning improves downstream performance.

Materials and Methods

Experimental Models

This work investigates the effects of domain alignment on downstream segmentation task performance through a comparison study of three models representing varying levels of domain proximity.

- **nnU-Netv2[3]** Randomly initialized supervised baseline; represents zero/no domain alignment.
- **DINOv3[4]** Natural-image pretrained foundation model baseline initialized from frozen DINOv3 encoder; represents low-domain alignment.
- **CBCT-DINOv3** CT fine-tuned foundation model[2] initialized from frozen CT-DINOv3 encoder; represents high-domain alignment.

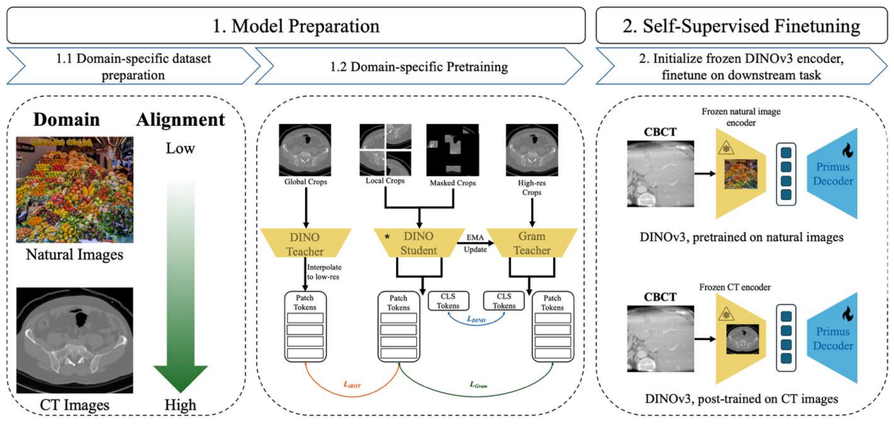

Figure 1: Schematic workflow for the investigation into the effects of domain-alignment on downstream segmentation task performance. Stage 1 (Model Preparation): Two pretraining datasets were constructed with different levels of domain-alignment to CBCT, and the corresponding DINOv3 encoders were initialized from natural image pretraining or from further CT fine-tuning. Stage 2 (Self-supervised Fine-tuning): The DINOv3 encoders are frozen (denoted with ice symbol) and integrated with the (trainable) Primus decoder (denoted with fire symbol); the models are then fine-tuned on the CBCTLiTS dataset to assess downstream segmentation task performance and investigate the impact of pretraining domain-alignment on performance.

Dataset Descriptions

Three datasets were used to prepare and evaluate the three models; a large million-scale CT dataset for domain-specific pretraining, and two CBCT datasets for task-specific finetuning and model evaluation.

- **CT-Pretraining[2]** 3.87 M axial CT slices aggregated from 16 publicly available datasets; spans abdominal, thoracic, and pelvic regions for broad anatomical coverage; 3D volumes resampled to 2D in-plane isotropic spacing of 0.45 mm^2 with uniform slice resolution of 256×256.
- **CBCTLiTS[5]** 201 synthetic CBCT scans generated from contrast-enhanced abdominal CT, generated at five varying quality levels based on number of projections used during reconstruction (5 – 490, 4 – 256, 3 – 128, 2 – 64, 1 – 32), total 1,005 total scans; synthetic CBCTs reconstructed with in-plane isotropic spacing of 0.69 mm^2 and 1.032 mm slice thickness.
- **Institutional (Zero-shot)** 10 clinical abdominal CBCT scans with liver annotations; Liver and lesion contours propagated from planning CT to fractional CBCTs with expert revision; resulting CBCT volumes exported and resampled to in-plane isotropic spacing of 0.69 mm^2 and 1.032 mm slice thickness.

Evaluation

Five-fold cross validation was used to evaluate task-specific model finetuning. Model performance was evaluated through liver and lesion Dice scores, liver 95th percentile Hausdorff Distance, and lesion centroid distance. Additional evaluation through:

- **Lesion Volume stratification** Lesions stratified by volume into small (<5cc), medium (5 – 30cc), and large (>30cc) lesions.
- **Scan Quality stratification** Model performance was evaluated across all five quality levels to provide insight into model performance under varying CBCT image quality.
- **Zero-Shot evaluation** Evaluates model performance in transferring from synthetic to clinical CBCT

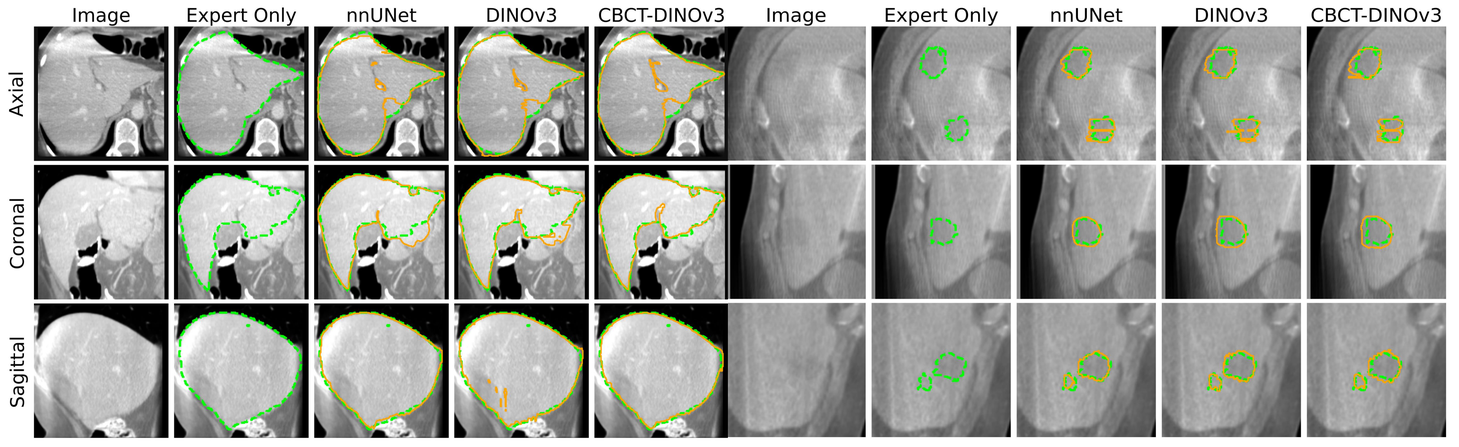

Figure 2: Representative cases for liver (left) and lesion (right) segmentation performance. Cases were chosen based on median performance across models. Each major row is an anatomical view, while each column shows the models predicted segmentation (orange) compared against the ground truth (green), displayed individually in column two.

		nnUNet		DINOv3		CBCT-DINOv3	
		Mean	Median	Mean	Median	Mean	Median
Liver	DSC ↑	$0.933_{(0.058)}$	0.953	$0.935_{(0.054)}$	0.953	**$0.939_{(0.049)}$**	**0.955**
	HD_{95} (mm) ↓	$11.72_{(13.09)}$	7.21	$12.23_{(16.56)}$	6.89	**$9.70_{(13.45)}$**	**5.88**
Lesion	DSC ↑	$0.406_{(0.304)}$	0.378	$0.404_{(0.310)}$	0.383	**$0.420_{(0.310)}$**	**0.441**
	Centroid Dist. (mm) ↓	$33.25_{(36.26)}$	19.01	**$32.82_{(38.42)}$**	**15.78**	$34.07_{(39.44)}$	18.88

Table 1: Quantitative results for CBCT liver and lesion segmentation. Performance is reported for Dice (DSC), 95th percentile Hausdorff Distance (HD_{95}) (mm) and centroid distance (mm), as mean (standard deviation) and median. Best performance for each metric highlighted in **bold**. Arrows indicate direction of better results.

Table 2: Quantitative results for CBCT lesion segmentation, stratified by lesion volume. Lesion volume categorized as small (<5 cc), medium (5 cc – 30 cc), and large (>30 cc). Performance is reported for two metrics, DSC and centroid distance (mm), as mean (standard deviation) and median. Best performance for each metric highlighted in **bold**. Arrows indicate direction of better results.

		nnUNet		DINOv3		CBCT-DINOv3	
		Mean	Median	Mean	Median	Mean	Median
DSC	Small	**$0.217_{(.254)}$**	**0.129**	$0.165_{(.236)}$	0.060	$.214_{(0.273)}$	0.060
	Med.	$0.477_{(0.268)}$	0.475	$0.503_{(0.251)}$	0.499	**$0.512_{(0.257)}$**	**0.519**
	Large	$0.511_{(0.300)}$	0.556	**$0.529_{(0.297)}$**	**0.622**	$0.503_{(0.316)}$	0.570
Centroid Dist. (mm)	Small	**$55.21_{(44.16)}$**	48.74	$58.45_{(46.24)}$	49.08	$59.75_{(48.97)}$	**45.94**
	Med.	$21.78_{(22.72)}$	13.01	**$15.54_{(14.29)}$**	**9.00**	$17.12_{(14.94)}$	15.83
	Large	**$23.66_{(29.12)}$**	14.03	$24.42_{(32.93)}$	**12.77**	$26.52_{(33.89)}$	16.36

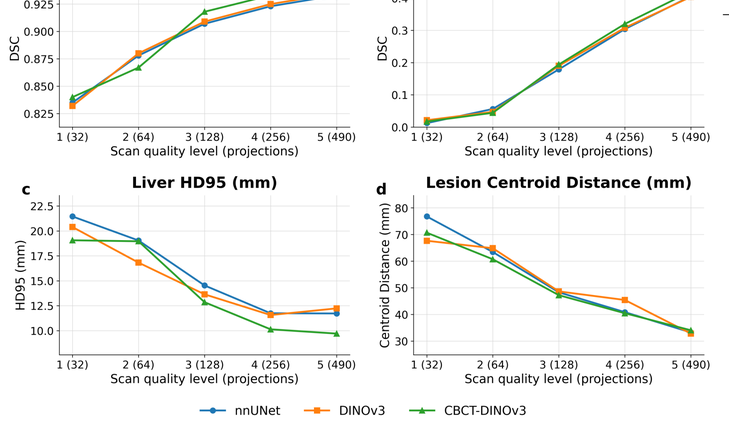

Figure 3: Quantitative results of model performance, stratified by scan quality. Scan quality determined by the number of projections used in CBCT generation. Scans categorized at five scan quality levels, with 5 (490 projections) being the highest and 1 (32 projections) being the lowest. Performance is reported as the mean for four metrics, Dice (DSC) (a, b), 95th percentile Hausdorff Distance (HD_{95}) (mm) (c) and centroid distance (mm) (d) at each of the scan quality levels. Results for nnUNet are shown in blue, in orange for DINOv3, and in green for CBCT-DINOv3

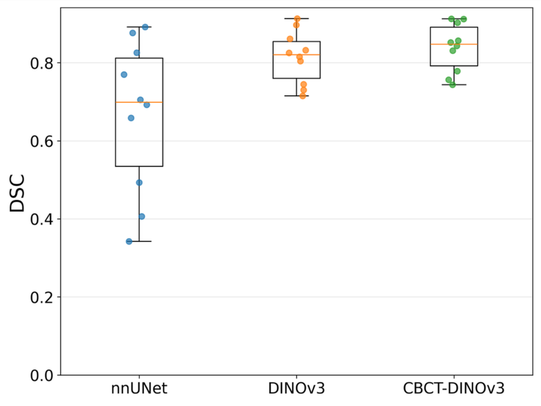

Figure 4: Distribution of zero-shot liver segmentation performance across ten clinical cases. Median DSC is shown as an orange line, with individual cases represented as data points. Boxes represent interquartile range and whiskers denote the full range of the data. CBCT-DINOv3 results are shown in green, DINOv3 results are shown in orange, and nnUNet results are shown in blue.

Results

Overall, CBCT-DINOv3 showed less false positive extension and better ground-truth alignment for liver segmentation and demonstrated comparable performance in lesion segmentation (Figure 2) compared to the baselines. Additionally, CBCT-DINOv3 demonstrated superior DSC scores in both liver and lesion segmentation and HD_{95} scores in liver segmentation (Table 1). On the zero-shot cohort, CBCT-DINOv3 demonstrated less performance variability across the cases and reported higher mean liver DSC scores (Figure 4). Finally, results were somewhat mixed when stratified by scan quality (Figure 3) and lesion volume (Table 2), with CBCT-DINOv3 sometimes outperforming baselines, and demonstrating comparable performance at others.

Discussion

CBCT-DINOv3's performance compared to baselines suggests a potentially feasible pathway toward developing large-scale CBCT foundation models, however limitations such as limited large-scale data availability and computational limitations provide directions for future work.

Acknowledgements

This research was supported by the National Institutes of Health under award numbers R01CA272991, R01EB032680, U54CA274513, and NSF 0000084444.

Contact

Corresponding Author: Xiaofeng Yang, PhD, DABR, FSPIE
Email: xfyang@uchicago.edu

References

1. Jaffray DA. Image-guided radiotherapy: from current concept to future perspectives. *Nat Rev Clin Oncol.* 2012;9(12):688-699. doi:10.1038/nrclinonc.2012.194
2. Li Y, Wu Y, Lai Y, Hu M, Yang X. MedDINOv3: how to adapt vision foundation models for medical image segmentation? *arXiv* [preprint]. Published 2025. doi:10.48550/arXiv.2509.02379
3. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods.* 2021;18(2):203-211. doi:10.1038/s41592-020-01008-z
4. Siméoni O, Vo HV, Seitzer M, et al. DINOv3. *arXiv* [preprint]. Published August 2025. doi:10.48550/arXiv.2508.10104
5. Tschuchnig ME, Steininger P, Gadermayr M. CBCTLiTS: a synthetic, paired CBCT/CT dataset for segmentation and style transfer. *arXiv* [preprint]. Published July 20, 2024. doi:10.48550/arXiv.2407.14853