High-Resolution CT Imaging Using a Few-Shot Foundation Model

Alexandru Florea[1,2], Mojtaba Safari[3], Shansong Wang[3], Mengdi Li[1], Beth Ghavidel[1], Huiqiao Xie[1] and Xiaofeng Yang[1,2,3]

[1]*Department of Radiation Oncology, Winship Cancer Institute, Emory University School of Medicine, Atlanta, GA , USA*
[2]*Department of Biomedical Engineering, Emory University and Georgia Institute of Technology, Atlanta, GA , USA*
[3]*Department of Radiation and Cellular Oncology, The University of Chicago, Chicago, IL, USA*

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

DINOv3-based Super Resolution (SR) models reproduce ground-truth HU values more faithfully and with more realistic anatomical reconstruction. SR-DINOv3 achieved the best balance of improved through-plane detail and intensity fidelity, supporting its potential utility for radiotherapy treatment planning.

Introduction

- **Low Resolution CT** is commonly used in radiotherapy planning to reduce imaging dose and decreases contouring or treatment planning workload due to fewer slices.
- **High Resolution CT** can improve through-plane resolution and reduce partial-volume effects, however, acquisitions are often constrained by radiation dose, motion artifacts, and scanner limitations.
- **SR-DINOv3** aims to leverage the rich latent space of DINOv3 models to generate high quality synthetic HR CT from LR inputs.

Methods

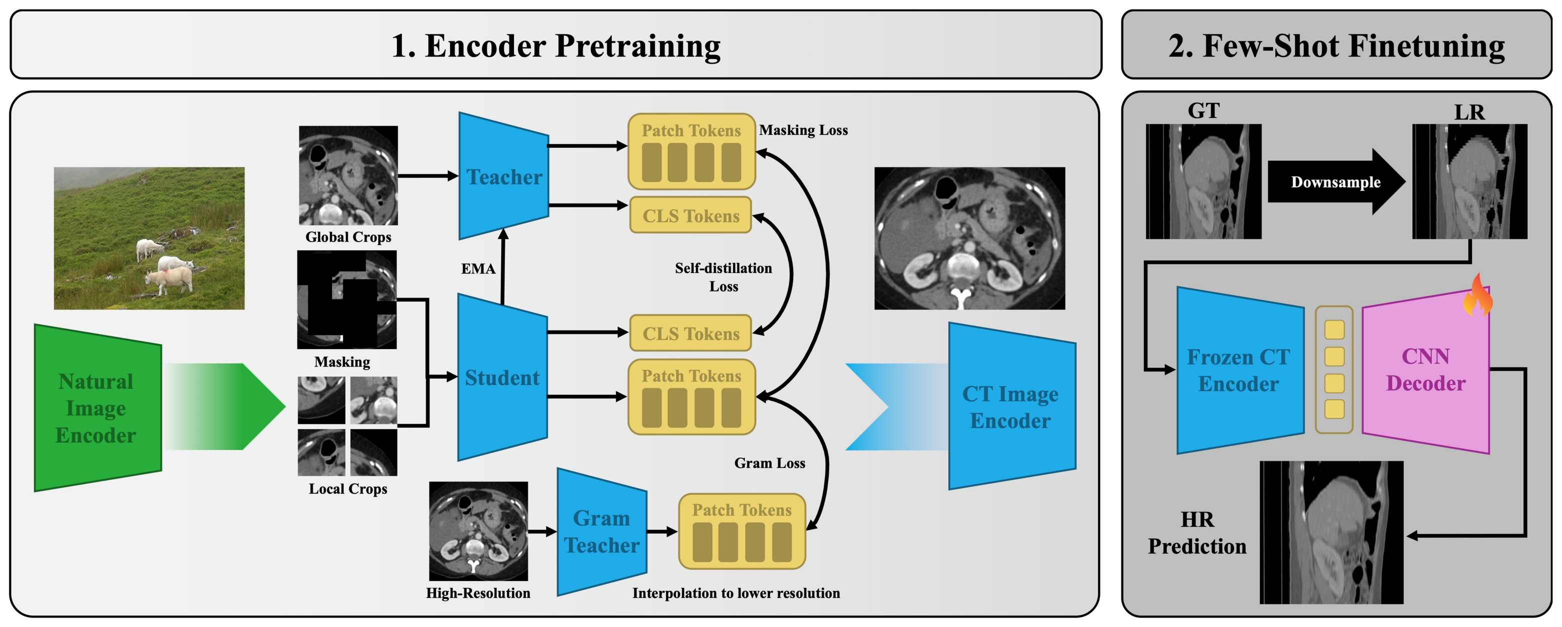


Figure 1. Workflow diagram used to investigate a domain-aligned DINOv3 model on downstream CT image super-resolution tasks. In Step 1 (Encoder Preparation), we prepare the domain-aligned SR-DINOv3 encoder. A generic DINOv3 encoder, whose weights are initialized on natural images, is pretrained on axial CT images using a curated dataset containing 3.87M axial CT slices. In Step 2 (Few-Shot Finetuning), the pretrained DINOv3 encoder is frozen and coupled to an up-sampling Convolutional Neural Network (CNN) decoder. The model is then trained for 300 epochs using Low Rank Adaptation (LoRA).

Datasets

- **CRLM** is a collection of CT images acquired from 197 patients with colorectal liver metastases. Low resolution images were synthetically generated using nearest neighbor interpolation and a downsample factor of 3x.
- **RPLHR** contains real-paired low (5 mm) and high (1 mm) resolution chest CT scans from 250 patients. To ensure uniformity, low resolution images were upsampled by a factor of 5x and dynamically cropped along the slice direction using a sliding window-based approach.

References

1. Xie, Huiqiao, Yang Lei, Tonghe Wang, et al. "High Through-Plane Resolution CT Imaging with Self-Supervised Deep Learning." Physics in Medicine and Biology 66, no. 14 (2021): 10.1088/1361-6560/ac0684. https://doi.org/10.1088/1361-6560/ac0684.
2. Li, Yuheng, Yizhou Wu, Yuxiang Lai, Mingzhe Hu, and Xiaofeng Yang. "MedDINOv3: How to Adapt Vision Foundation Models for Medical Image Segmentation?" arXiv:2509.02379. Preprint, arXiv, September 3, 2025. https://doi.org/10.48550/arXiv.2509.02379.
3. Simpson, Amber L., Jacob Peoples, John M. Creasy, et al. "Preoperative CT and Survival Data for Patients Undergoing Resection of Colorectal Liver Metastases." Scientific Data 11, no. 1 (2024): 172. https://doi.org/10.1038/s41597-024-02981-2.
4. Yu, Pengxin, Haoyue Zhang, Han Kang, Wen Tang, Corey W. Arnold, and Rongguo Zhang. RPLHR-CT Dataset and Transformer Baseline for Volumetric Super-Resolution from CT Scans. Vol. 13436. 2022. https://doi.org/10.1007/978-3-031-16446-0_33.

Results

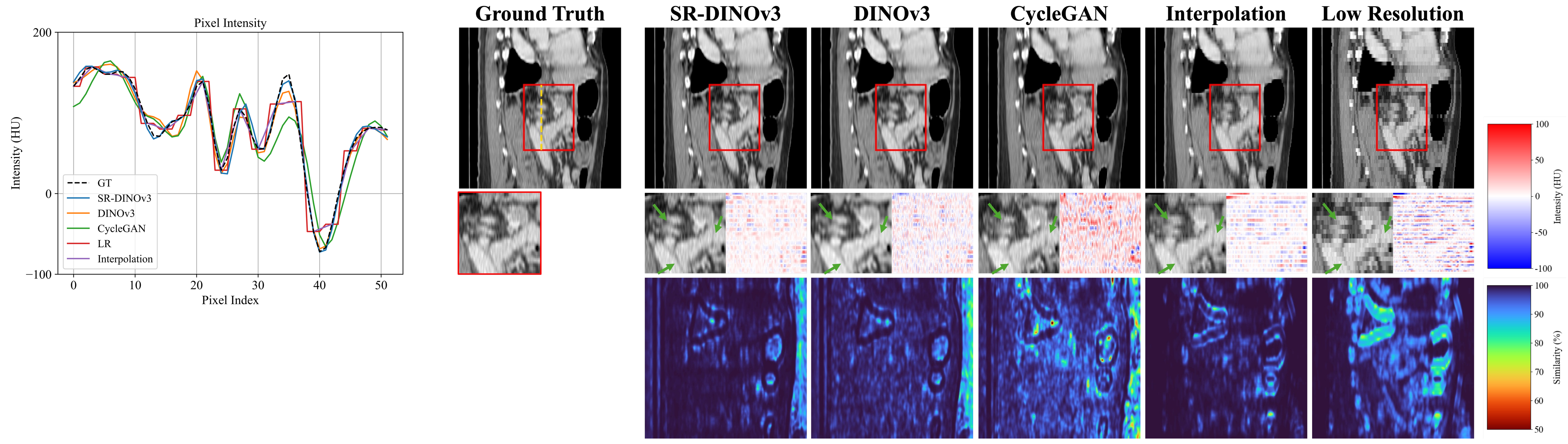
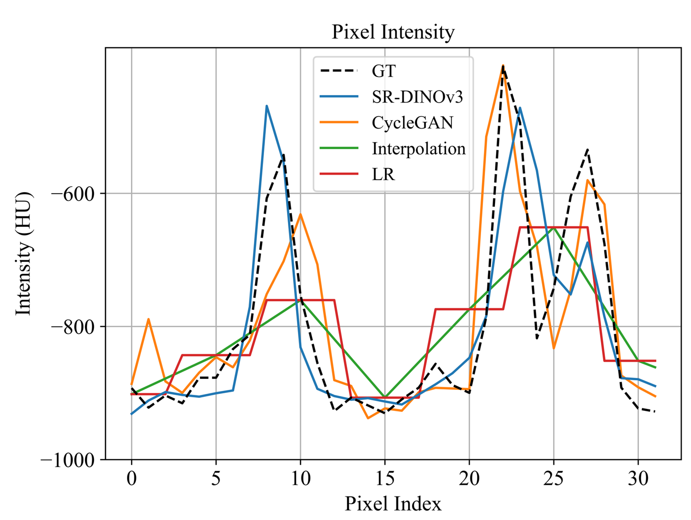
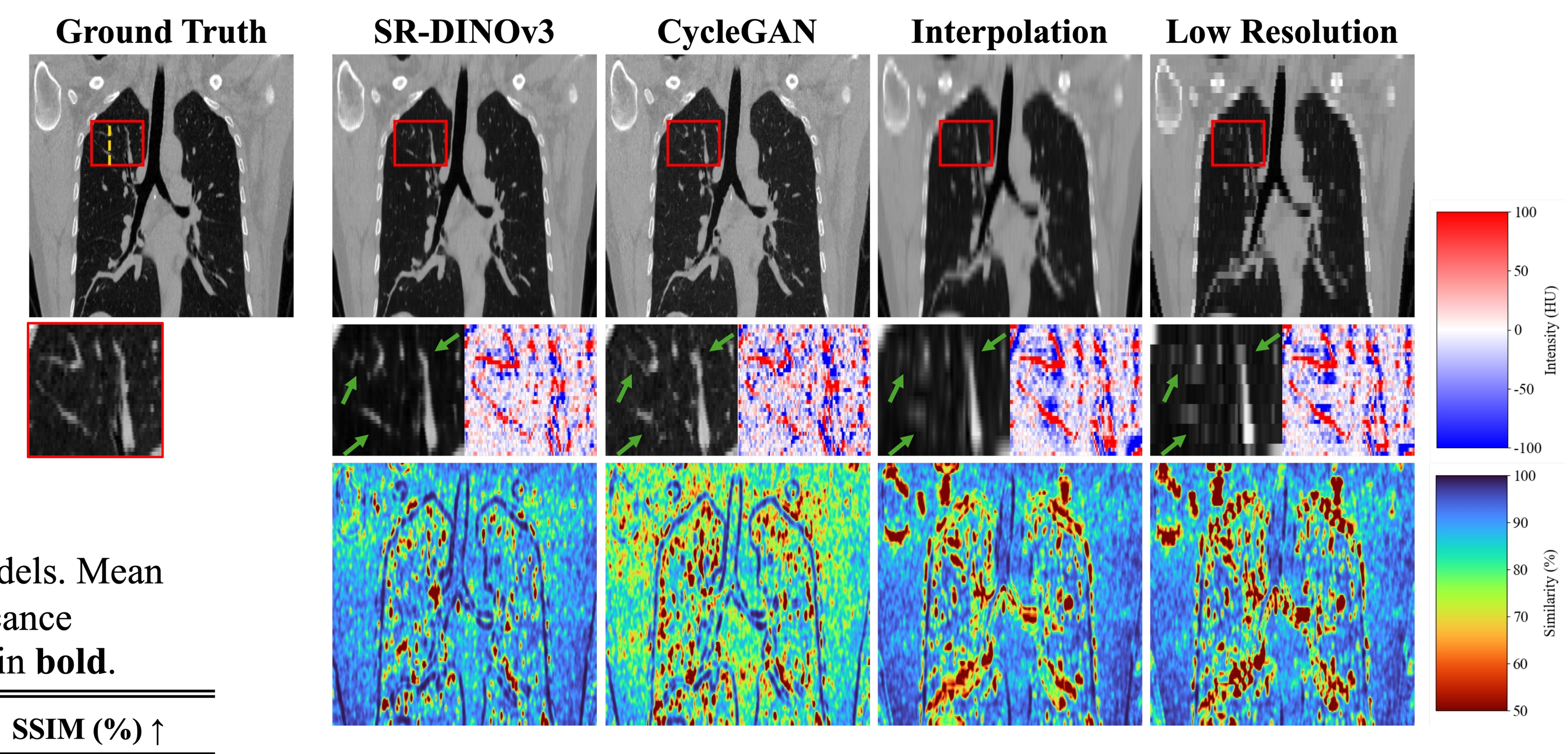


Figure 2. (Right) Comparison of the super-resolution results on the Colorectal Liver Metastases (CRLM) dataset. From left to right: Ground Truth (GT), proposed SR-DINOv3, DINOv3, CycleGAN, Linear Interpolation, Low Resolution input. As expected, resolution improves along the superior–inferior axis, while resolution in the anterior–posterior direction remained unchanged since it was not undersampled. Regions of Interest (ROI) are highlighted with a red bounding box and magnified beneath, with paired difference maps between GT and different SR-CT reconstruction techniques measured in absolute HU difference presented side-by-side. **(Below)** Structural Similarity Index Measure (SSIM) maps measuring similarity between GT and different SR-CT reconstructions. **(Left)** Signal intensity profile comparison corresponding to pixel intensities (measured in HU) along the yellow dashed line in the ROI of the GT image. SR-DINOv3 shows improved anatomical reconstruction detail indicated by the green arrows, higher overall SSIM, and better HU fidelity.





Tabel 1. Average reconstruction performance for super-resolution models. Mean value with one standard deviation above and below. Statistical significance evaluated using two-sided Wilcoxon signed-rank test. Best values are in **bold**.

Dataset	Model	MAE (HU) ↓	PSNR (dB) ↑	SSIM (%) ↑
CRLM	SR-DINOv3	11.29 ± 4.48	**41.47 ± 4.20**	97.40 ± 2.02
	DINOv3	12.71 ± 4.34*	40.30 ± 3.42*	96.38 ± 2.31*
	CycleGAN	21.07 ± 13.82*	36.50 ± 3.90*	93.37 ± 3.90*
	Interpolation	**8.69 ± 4.26***	40.15 ± 3.42*	**97.45 ± 2.13***
RPLHR	SR-DINOv3	**23.09 ± 6.07**	**35.14 ± 2.66**	**88.13 ± 3.68**
	CycleGAN	32.14 ± 16.33*	32.75 ± 2.70*	81.35 ± 6.69*
	Interpolation	27.82 ± 6.89*	32.63 ± 2.61*	86.17 ± 4.27*

***Denotes $p < 0.05$ compared to SR-DINOv3**

Figure 3. (Right) Comparison of the super-resolution results on the Real-Paired Low- and High-Resolution (RPLHR) dataset. From left to right: Ground Truth (GT), proposed SR-DINOv3, CycleGAN, Linear Interpolation, Low Resolution input. Regions of Interest (ROI) are highlighted with a red bounding box and magnified beneath, with paired difference maps between GT and different SR-CT reconstruction techniques measured in absolute HU difference presented side-by-side. **(Below)** Structural Similarity Index Measure (SSIM) maps measuring similarity between GT and different SR-CT reconstructions. **(Left)** Signal intensity profile comparison corresponding to pixel intensities (measured in HU) along the yellow dashed line in the ROI of the GT image. SR-DINOv3 shows improved anatomical reconstruction detail indicated by the green arrows, higher overall SSIM, and better HU fidelity.

Results

The **CRLM dataset** highlights performance of larger anatomical structure and various soft tissue intensities along the coronal axis with a downsample factor of 3x. SR-DINOv3 shows strong MAE, PSNR, and SSIM, highlighting its superior performance in recovering otherwise lost anatomical features due to degradations. **RPLHR** is a more challenging dataset with a downsample factor of 5x and fine anatomical lung structures. Here we see the strength of SR-DINOv3 in reconstructing fine lung nodules with more accurate pixel intensity (HU), compared to other methods.

Contact

Deep Biomedical Imaging Lab:

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

Conclusion

- **SR-DINOv3** achieves the best balance of improved through-plane detail and quantitative HU fidelity.
- **DINOv3** trained on natural images provides strong baseline performance but can be improved with domain-aligned CT finetuning.
- **CycleGAN** reproduces high quality anatomical structure, but struggles to preserve accurate pixel intensities, resulting in noisy images and low SSIM.
- **Linear Interpolation** produces lower MAE, but this is because pixel intensities are linearly interpolated to be the exact value of nearby pixels, resulting in good evaluation metrics, but poor anatomical fidelity.

Acknowledgements

This research is supported in part by the National Institutes of Health under Award Numbers R56EB033332, R01DE033512, and R01CA272991.