Raptor: Scalable Train-free Embeddings for 3D Medical Volumes Leveraging Pretrained 2D Foundation Models

ICML 2025

Ulzee An* [1], Moonseong Jeong* [1], Simon A. Lee[2], Aditya Gorla[2,3], Yuzhe Yang[1,2], Sriram Sankararaman[1,2,4]

[1]Computer Science Department, [2]Department of Computational Medicine, [3]Bioinformatics Interdepartmental Program, [4]Department of Human Genetics, University of California, Los Angeles

UCLA Samueli School of Engineering
UCLA David Geffen School of Medicine

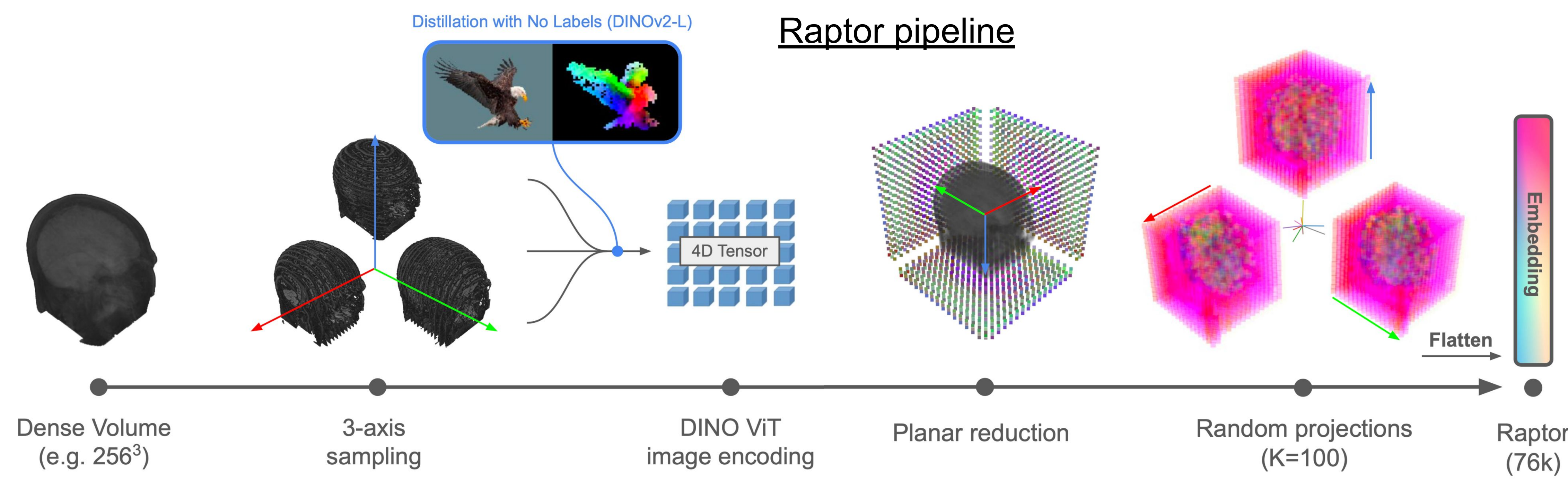
*Equal contribution

TL;DR

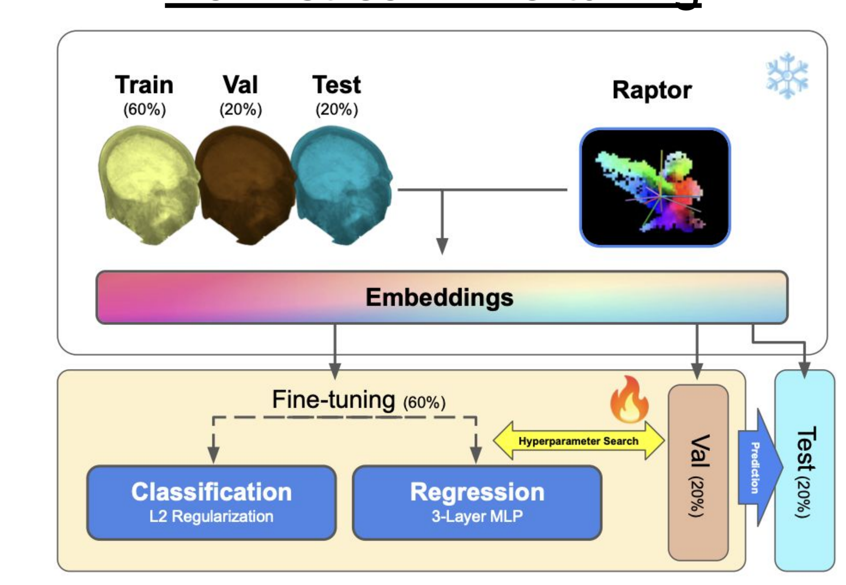
- Working with **3D medical volumes** is **hard**…
  - Limited 3D medical datasets
  - Heavy compute required
- … but with **Raptor** it's **easy**.
  - *Train-free* embeddings that obtain SOTA
  - >10x compression of 3D medical volumes
  - Runs on *RTX 2080Ti*

Raptor: Random Planar Tensor Reduction

Raptor pipeline



Downstream fine-tuning



Motivation

3D medical imaging models are **computationally expensive** and require **large labeled datasets**, which are often unavailable in medical domains.

Training state-of-the-art 3D models demands **high memory, compute, and time**, creating barriers for many researchers.

Image foundation models (2D) are powerful and widely available but **underutilized in 3D domains**.

We introduce **Raptor** — a **train-free method** that generates **compact, semantically rich embeddings** for 3D medical volumes by leveraging frozen pretrained **image foundation models (2D)** and **random projections**. The embeddings enable downstream analysis without needing to fit an expensive 3D model on large datasets.

Methods	Medical Pretraining Data	Domain
SLIViT [1]	14M images (ImageNet) +108K OCT images	Optical
SuPreM [2]	5K CT volumes	General
Merlin [3]	15K volumes	Chest
MISFM [4]	110K CT volumes	General
VoCo [5]	160K CT volumes	General
Raptor (Ours)	None (uses 2D foundation model)	General*

*we note that Raptor is agnostic to any downstream domain, medical or not

Experiments & Results

3D Medical MNIST [6] Classification

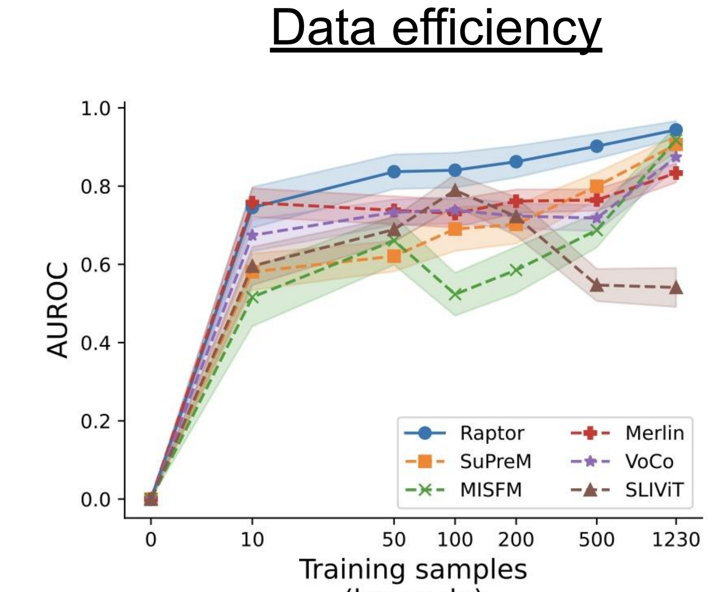
Methods	Organ AUC	Organ ACC	Nodule AUC	Nodule ACC	Fracture AUC	Fracture ACC	Adrenal AUC	Adrenal ACC	Vessel AUC	Vessel ACC	Synapse AUC	Synapse ACC
ResNet	0.995	0.918	0.886	0.860	**0.759**	0.487	0.869	0.835	0.932	0.915	0.889	0.853
MAE	0.982	0.800	0.820	0.828	0.600	0.481	0.710	0.752	0.607	0.880	0.560	0.737
MISFM	0.989	0.833	0.886	0.855	0.689	0.537	0.868	0.665	0.932	0.894	0.918	0.307
SuPreM	**0.999**	**0.968**	0.891	0.848	0.645	0.492	0.906	0.869	0.964	**0.929**	0.907	0.879
SLIViT	0.997	0.946	0.920	0.868	0.656	0.475	0.846	0.789	0.710	0.880	0.541	0.270
VoCo	0.992	0.870	0.797	0.836	0.699	0.535	0.913	**0.872**	0.799	0.880	0.844	0.830
Merlin	0.976	0.766	0.809	0.861	0.691	**0.549**	0.836	0.801	0.870	0.879	0.833	0.825
Raptor-B	0.998	0.958	0.904	0.858	0.647	0.501	**0.930**	0.858	0.945	0.919	0.922	0.894
Raptor	**0.999**	0.961	**0.929**	**0.870**	0.677	0.502	0.926	0.845	**0.966**	0.922	**0.943**	**0.911**

(*Raptor-B* is a 10x smaller variant of Raptor!)

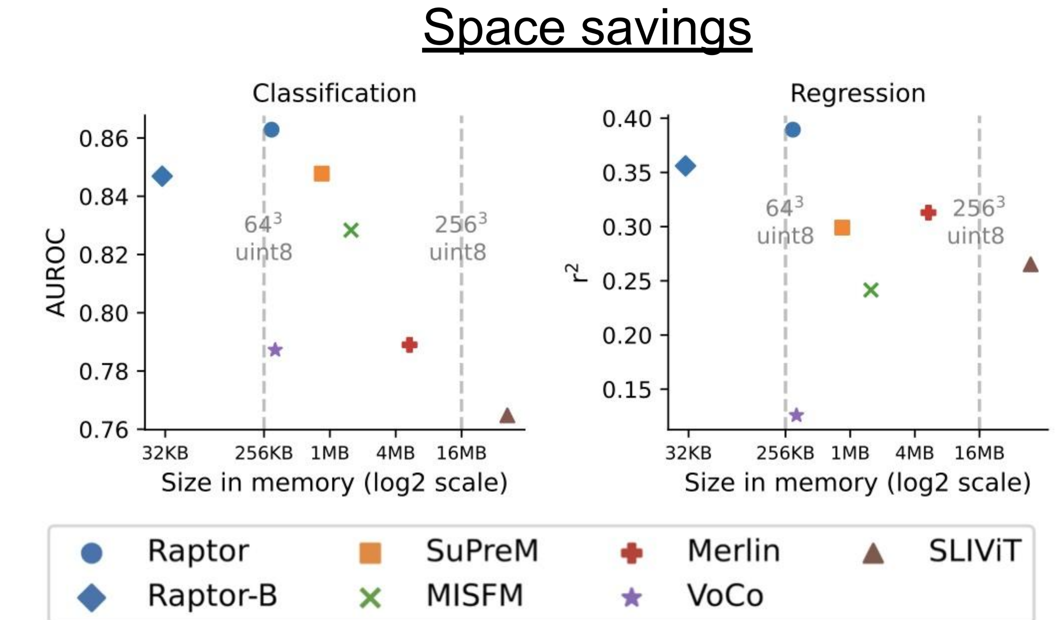
UK Biobank Brain Volumes Regression

Methods (r^2)	WhiteM	GrayM	Cereb	Amyg	Hippo	Cortex	Gyrus	Pall	Caud	Thal	Avg.
ResNet	0.417	0.562	0.193	0.072	0.108	0.125	0.099	0.055	0.162	0.134	0.193
MAE	0.036	0.045	0.072	0.036	0.040	0.043	0.032	0.012	0.037	0.036	0.039
MISFM	0.418	0.624	0.276	0.089	0.145	0.236	0.209	0.087	0.166	0.164	0.242
SuPreM	0.646	0.696	0.330	0.109	0.163	0.275	0.256	0.067	0.255	0.195	0.299
SLIViT	0.474	0.694	0.258	0.134	0.190	0.268	0.213	0.053	0.192	0.174	0.265
VoCo	0.225	0.375	0.189	0.071	0.113	0.059	0.048	0.043	0.060	0.075	0.126
Merlin	0.622	0.734	0.335	0.127	0.180	0.313	0.269	0.093	0.247	0.210	0.313
Raptor-B	0.614	0.742	0.398	**0.185**	0.247	0.355	0.314	0.116	0.331	0.258	0.356
Raptor	**0.681**	**0.777**	**0.437**	0.170	**0.262**	**0.404**	**0.340**	**0.142**	**0.381**	**0.300**	**0.389**

Data efficiency



Space savings



Ablation studies

Method	K	Seed 1	Seed 2	Seed 3	Std.
Raptor	1	0.818	0.817	0.793	0.0116
	5	0.890	0.860	0.864	0.0133
	10	0.866	0.896	0.876	0.0127
	100	0.901	0.899	0.900	0.0008
	150	0.898	0.897	0.897	0.0004

Do the **number of random projections** affect downstream accuracy?

Does the **number of views** represented in the embedding affect downstream accuracy?

Method	Viewpoint	AUC	ACC
Raptor	(A)xial	0.887	0.780
	(C)oronal	0.862	0.814
	(S)agittal	0.881	0.806
	A,C	0.893	0.826
	C,S	0.900	0.812
	A,S	0.884	0.822
	A,C,S	0.901	0.838

In conclusion, *Raptor* is…

★ **Train-free**: No need to train on 3D volumes — uses frozen 2D foundation models.
★ **Scalable**: Efficient for large, high-resolution volumes with sub-cubic complexity.
★ **Data-efficient**: Performs well even in low-data regimes common in medical settings.
★ **Compact**: Embeddings are up to **99%** smaller than raw voxel representations.
★ **Model-agnostic**: Compatible with any future 2D image foundation model.
★ **Strong performance**: Outperforms or matches SOTA pretrained 3D models across 10 diverse tasks.

Resources & Contact

Website **Github**

Email: {ulzee,bronsonj}@cs.ucla.edu

References

[1] Avram, Oren, et al. "SLIViT: a general AI framework for clinical-feature diagnosis from limited 3D biomedical-imaging data." Research Square (2023): rs-3.
[2] Li, Wenxuan, et al. "Abdomenatlas: A large-scale, detailed-annotated, & multi-center dataset for efficient transfer learning and open algorithmic benchmarking." Medical Image Analysis 97 (2024): 103285.
[3] Blankemeier, Louis, et al. "Merlin: A vision language foundation model for 3d computed tomography." Research Square (2024): rs-3.
[4] Wang, Guotai, et al. "Mis-fm: 3d medical image segmentation using foundation models pretrained on a large-scale unannotated dataset." arXiv preprint arXiv:2306.16925 (2023).
[5] Wu, Linshan, Jiaxin Zhuang, and Hao Chen. "Large-scale 3d medical image pre-training with geometric context priors." arXiv preprint arXiv:2410.09890 (2024).
[6] Yang, Jiancheng, et al. "Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification." *Scientific Data* 10.1 (2023): 41.