Lecture 4:
Vision Representation Learners in Biomedicine

Announcements

- A1 is now released on the class website, due Oct 16 at 11:59pm.
- Project instructions to come in the next few days.
- A1 release was also announced on Ed, please make sure you follow Ed for class announcements.

Assignment 1

- 1500-2000 word analysis paper using NeurIPS template
- Goal is to synthesize and think critically about the topics covered in class, through reading and analyzing recent papers
- Choose from one of two analysis categories:
  - “Extending representation learning to more complex modality setups”
  - “Diving into representation learning in a biomedical application”
- Whichever category you choose this time, you must choose the other for A2
- Detailed prompt instructions in the assignment guidelines
- We have provided possible paper suggestions. You are also very welcome to choose your own, but these must be checked with the course staff for suitability, through a private post on Ed.
  - You will hear back within 48 hours.

Today’s agenda

- Landscape of vision representation learners in biomedicine
- Specific examples

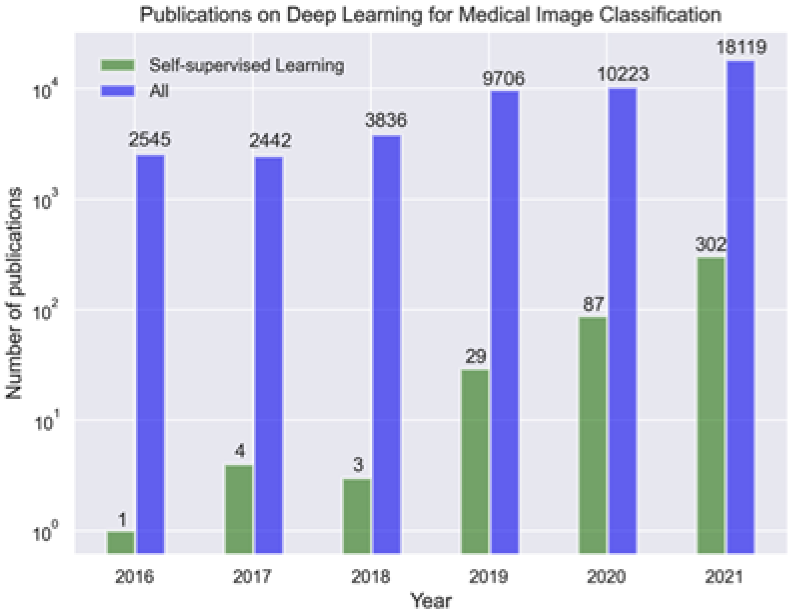
Visual representation learning for biomedical applications

- Impractical to curate large-scale labeled data for every different biomedical image analysis task or data distribution (e.g., different hospitals)
- Important strategy: learn powerful general feature representations from large amounts of unlabeled data. Then fine-tune using a much smaller amount of data for any given task.

Huang et al. 2023.

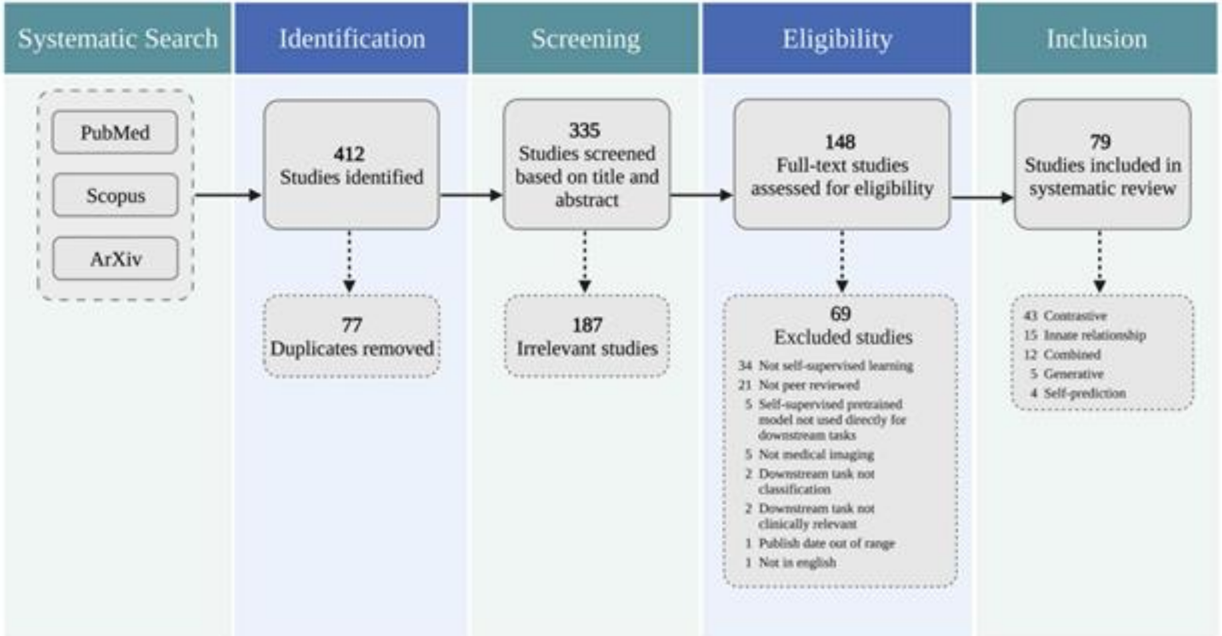
Self-supervised learning as a form of representation learning: a growing subset of deep learning for medical imaging literature

- Example of rising popularity of self-supervised learning between 2016-2021 (log-scale)
- Systematic literature search for “self-supervised learning” and “medical imaging” related terms, across PubMed, Scopus, and ArXiv
    - “Medical imaging” is defined broadly, including fundus photography, whole slide imaging, endoscopy, echocardiography, etc.



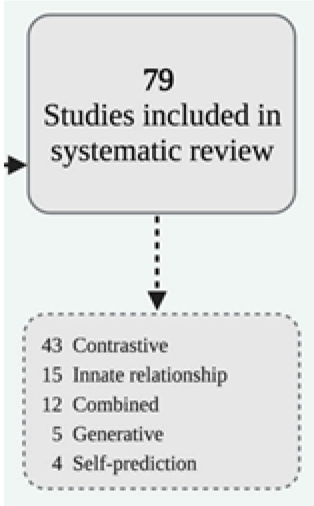
Huang et al. 2023.

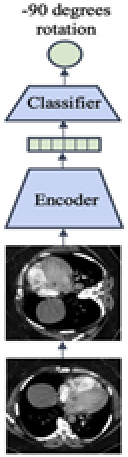
Systematic review of self-supervised papers that include downstream application to medical imaging classification



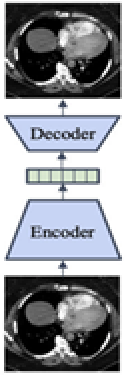
Huang et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. Npj Digital Medicine, 2023.

Remember: Different representation learning paradigms

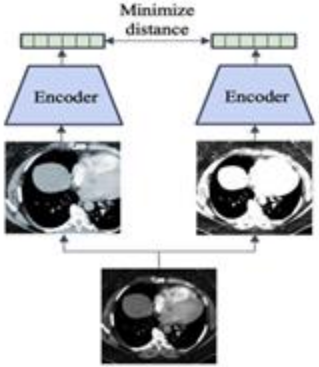




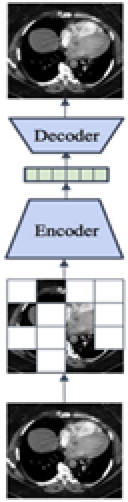
Innate relationship objective
E.g., predict rotation angle (or some other innate property) of an image



Generative objective
Compress and then reconstruct input image (e.g., variational autoencoders)



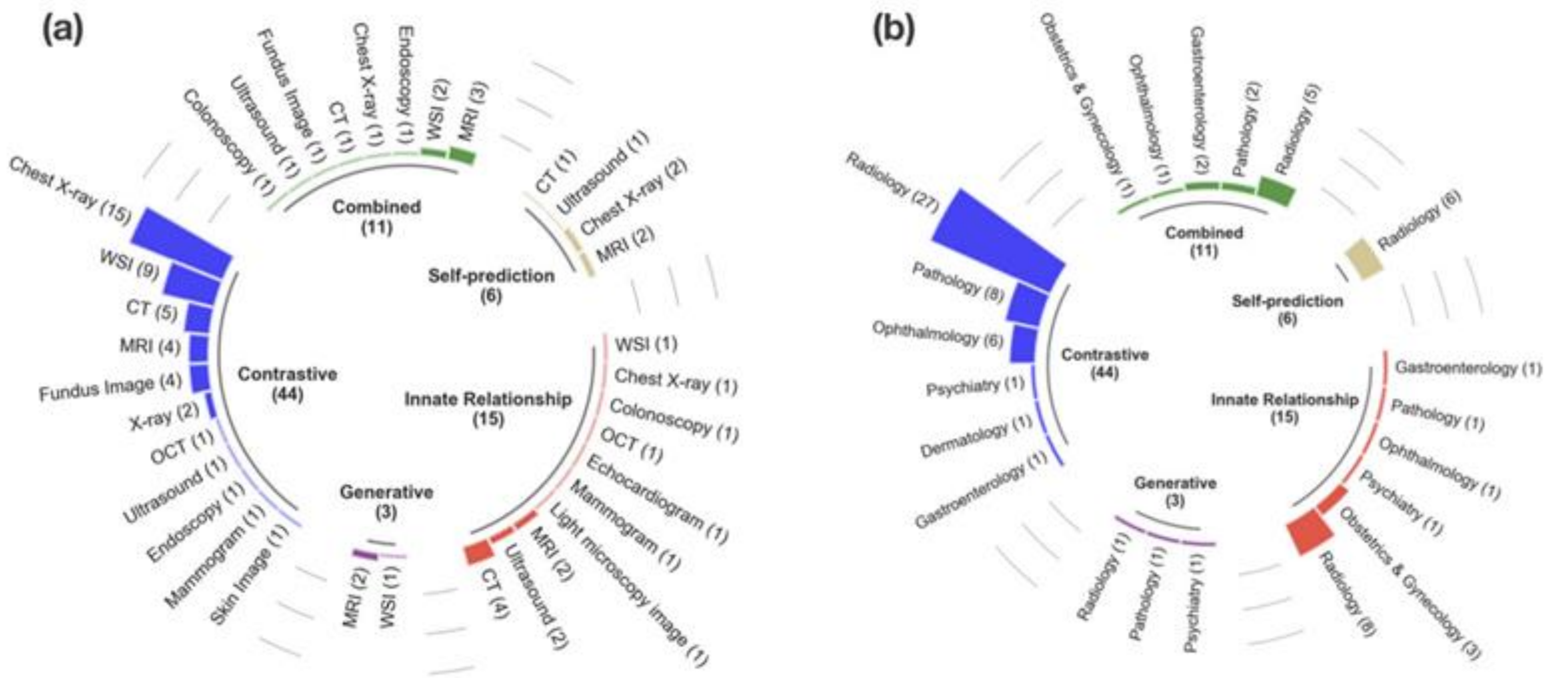
Contrastive objective
Different views of the same input should have more similar representation to each other than with a different input



Self-prediction objective
Mask parts of input data and predict these parts

Huang et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. Npj Digital Medicine, 2023.

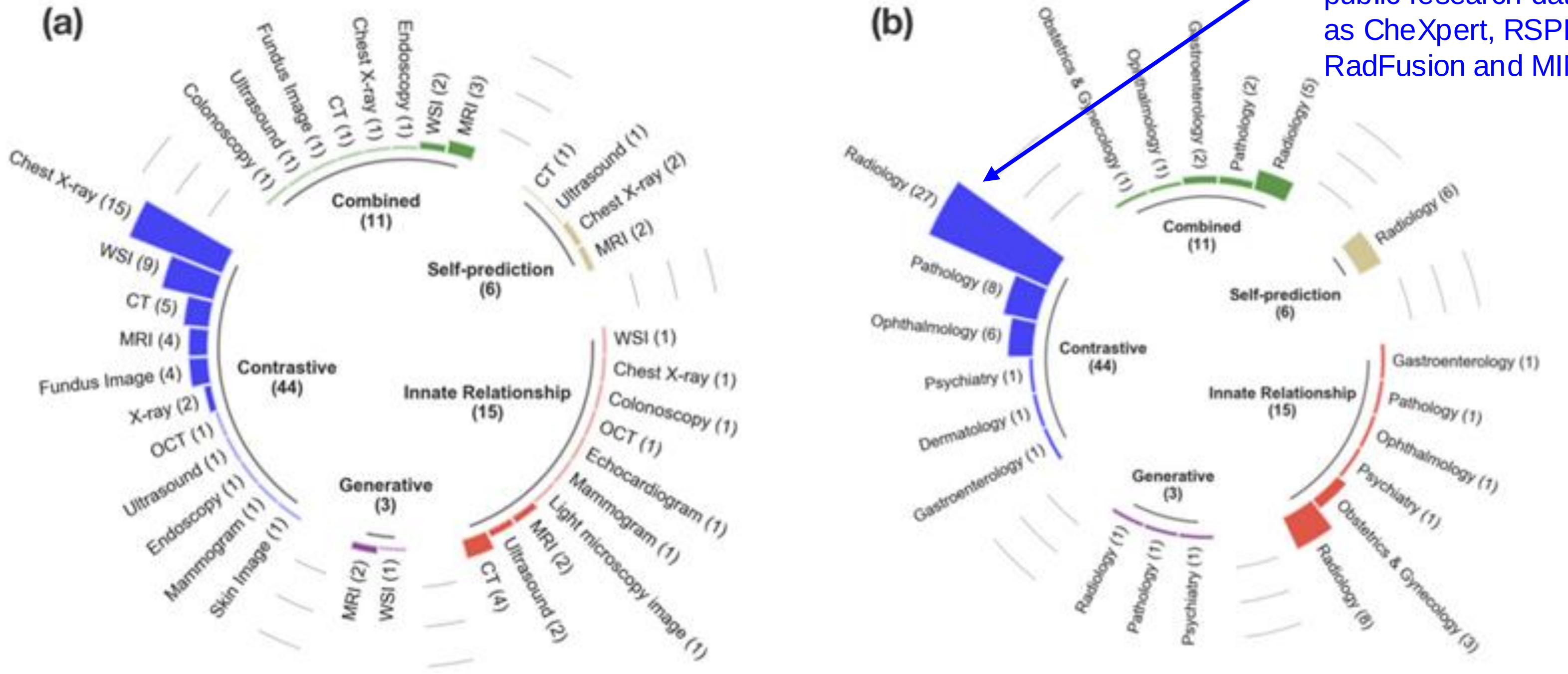
Distribution of medical imaging modalities and specialities



Huang et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. Npj Digital Medicine, 2023.

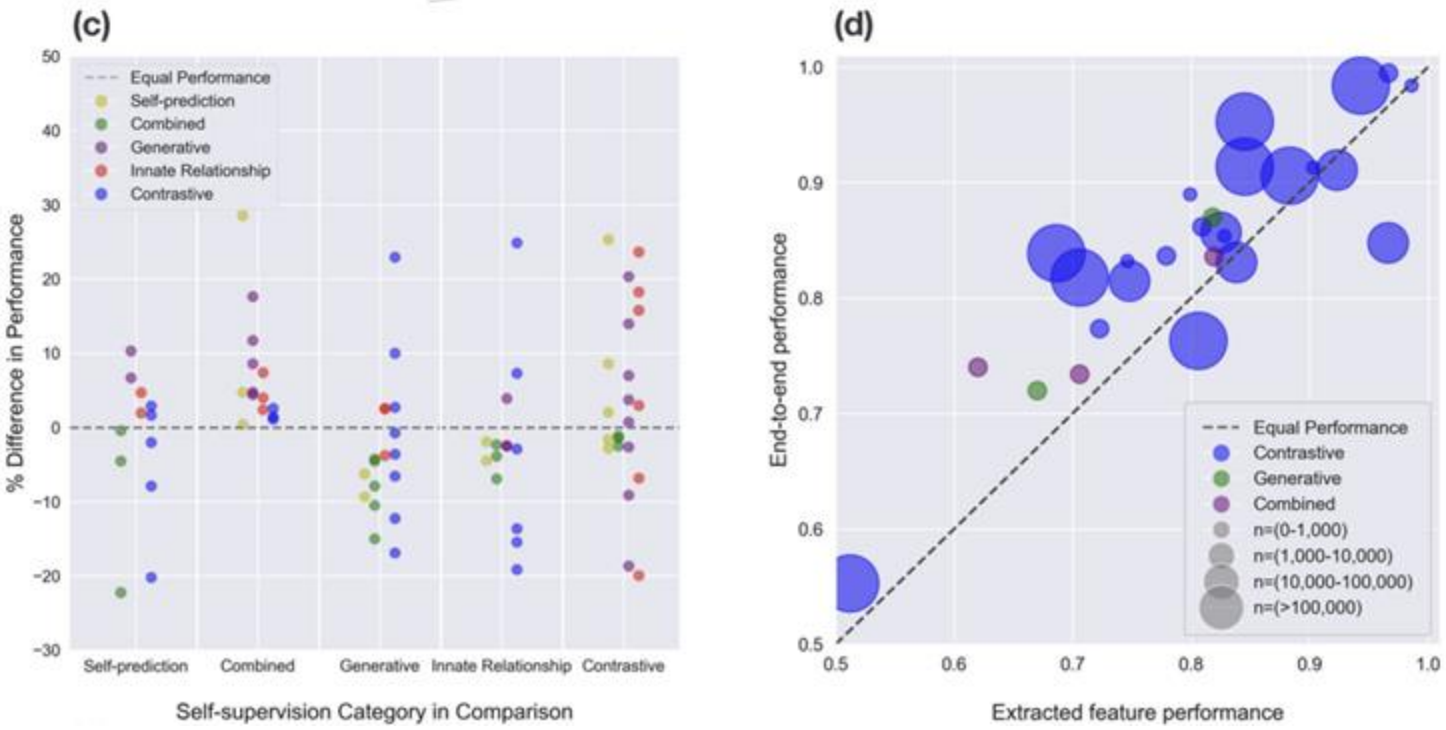
Distribution of medical imaging modalities and specialities

The radiology field has been active in releasing large scale, public research datasets such as CheXpert, RSPECT, RadFusion and MIMIC-CXR



Huang et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. Npj Digital Medicine, 2023.

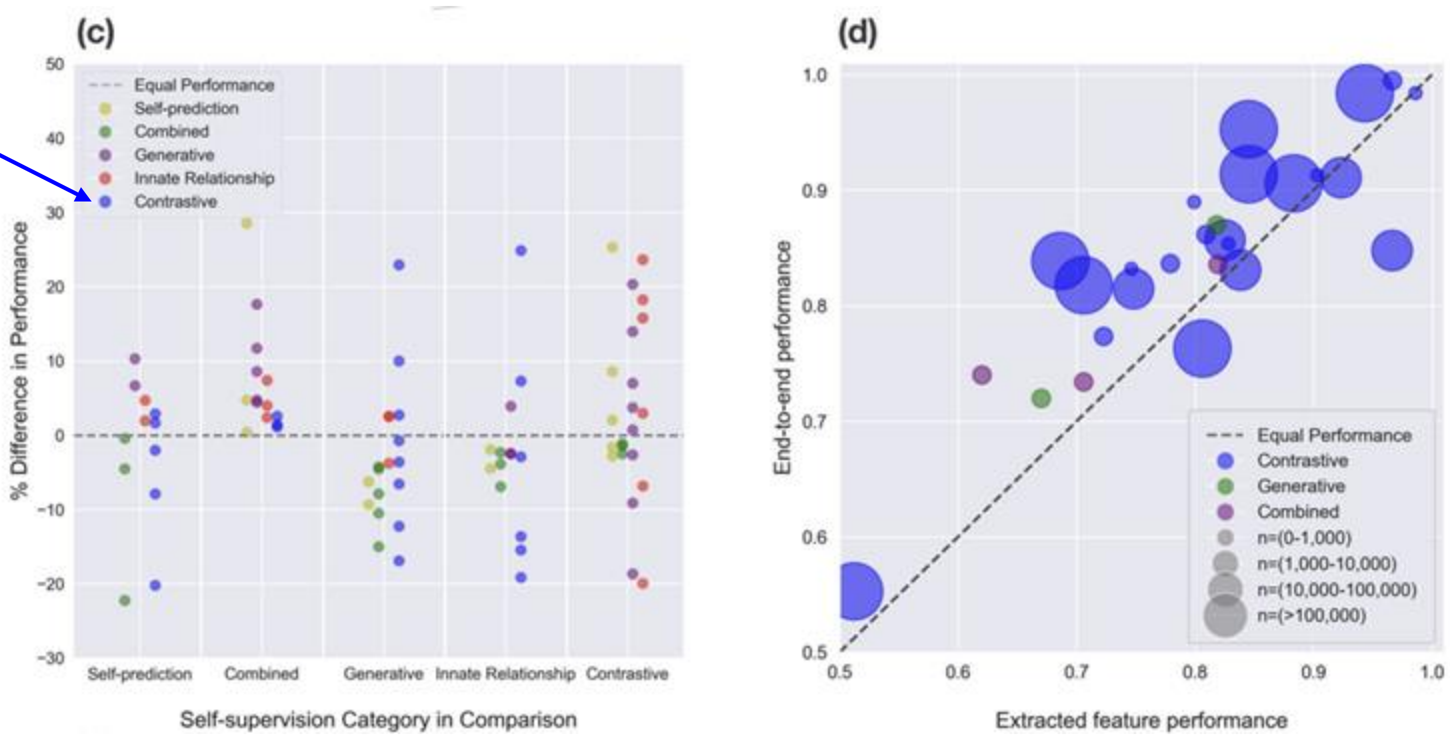
Self-supervised learning performance trends



Huang et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. Npj Digital Medicine, 2023.

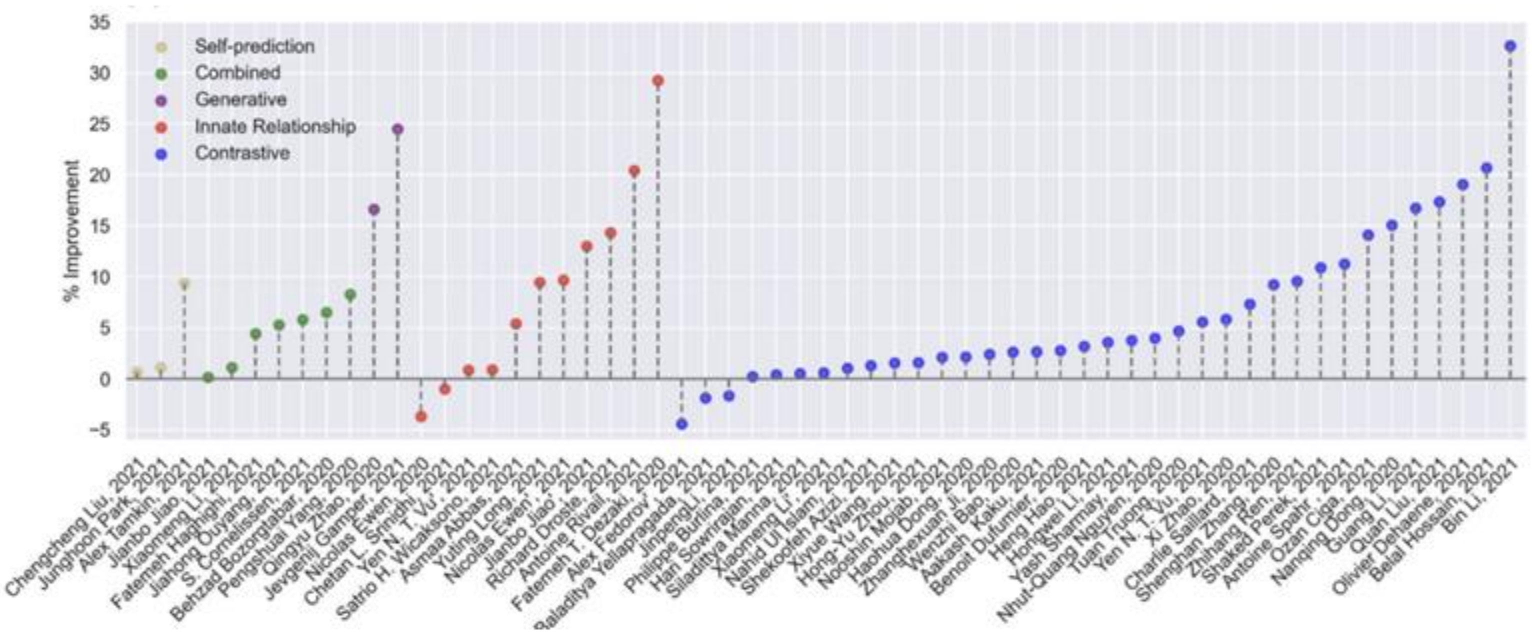
Self-supervised learning performance trends

Of contrastive approaches, SimCLR (13 papers) and MoCo (8 papers) were the most popular, out of 44.



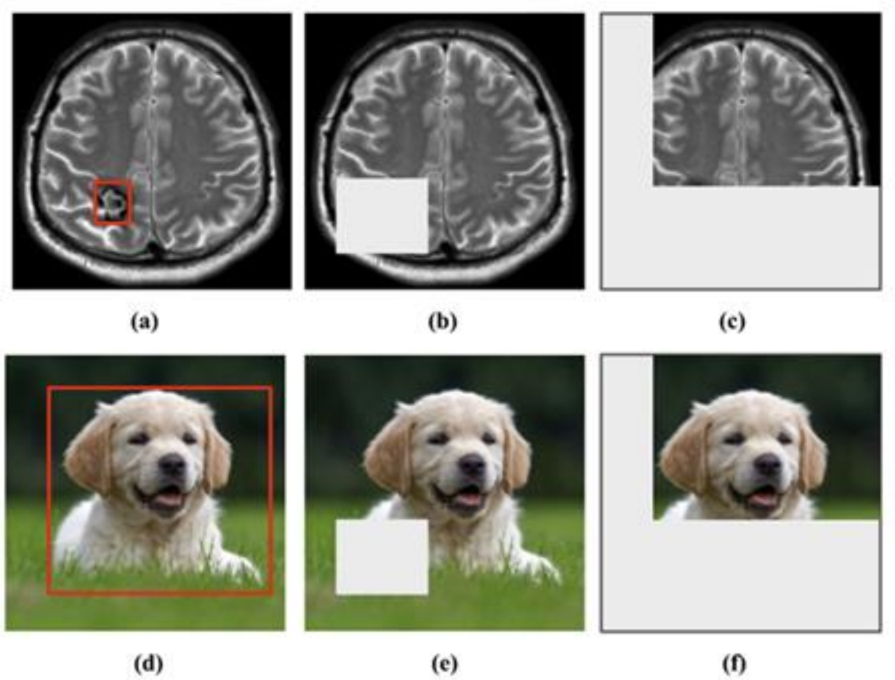
Huang et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. Npj Digital Medicine, 2023.

Reported performance improvement of fine-tuning after self-supervised learning vs. supervised learning



Huang et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. Npj Digital Medicine, 2023.

Aside: vanilla masking-based self-prediction methods may suffer from more noisy training signal for medical vs.natural images



Huang et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. Npj Digital Medicine, 2023.

A few more findings from the survey

- Many studies were found to be lacking in baseline and ablation experiments
  - Out of 79 studies, 60 compared with a supervised baseline, but only 33 with another self-supervised baseline
  - Of the 33 studies, 26 compared with a self-supervised category different from their best performing model. Of these, 16 compared with SimCLR, 11 with autoencoders, and 9 with MoCo.

Huang et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. Npj Digital Medicine, 2023.

A few more findings from the survey

- Many studies were found to be lacking in baseline and ablation experiments
  - Out of 79 studies, 60 compared with a supervised baseline, but only 33 with another self-supervised baseline
  - Of the 33 studies, 26 compared with a self-supervised category different from their best performing model. Of these, 16 compared with SimCLR, 11 with autoencoders, and 9 with MoCo.

Improving rigorous comparison of baselines and ablations is critical for research to be able to build on each other and to achieve rapid progress in the field

Huang et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. Npj Digital Medicine, 2023.

A few more findings from the survey

- Many studies were found to be lacking in baseline and ablation experiments
  - Out of 79 studies, 60 compared with a supervised baseline, but only 33 with another self-supervised baseline
  - Of the 33 studies, 26 compared with a self-supervised category different from their best performing model. Of these, 16 compared with SimCLR, 11 with autoencoders, and 9 with MoCo.
- Only 18 out of 79 studies reported results using natural image pre-trained weights (either supervised or self-supervised) to initialize their model for subsequent in-domain self-supervised pre-training

Huang et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. Npj Digital Medicine, 2023.

A few more findings from the survey

- Many studies were found to be lacking in baseline and ablation experiments
  - Out of 79 studies, 60 compared with a supervised baseline, but only 33 with another self-supervised baseline
  - Of the 33 studies, 26 compared with a self-supervised category different from their best performing model. Of these, 16 compared with SimCLR, 11 with autoencoders, and 9 with MoCo.
- Only 18 out of 79 studies reported results using natural image pre-trained weights (either supervised or self-supervised) to initialize their model for subsequent in-domain self-supervised pre-training

Worth stronger consideration in medical imaging.
This can often lead to the best results.

Huang et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. Npj Digital Medicine, 2023.

A few more findings from the survey

- Many studies were found to be lacking in baseline and ablation experiments
  - Out of 79 studies, 60 compared with a supervised baseline, but only 33 with another self-supervised baseline
  - Of the 33 studies, 26 compared with a self-supervised category different from their best performing model. Of these, 16 compared with SimCLR, 11 with autoencoders, and 9 with MoCo.
- Only 18 out of 79 studies reported results using natural image pre-trained weights (either supervised or self-supervised) to initialize their model for subsequent in-domain self-supervised pre-training
- Only 13 studies compared performance between classification using extracted features to end-to-end fine-tuning

Huang et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. Npj Digital Medicine, 2023.

A few more findings from the survey

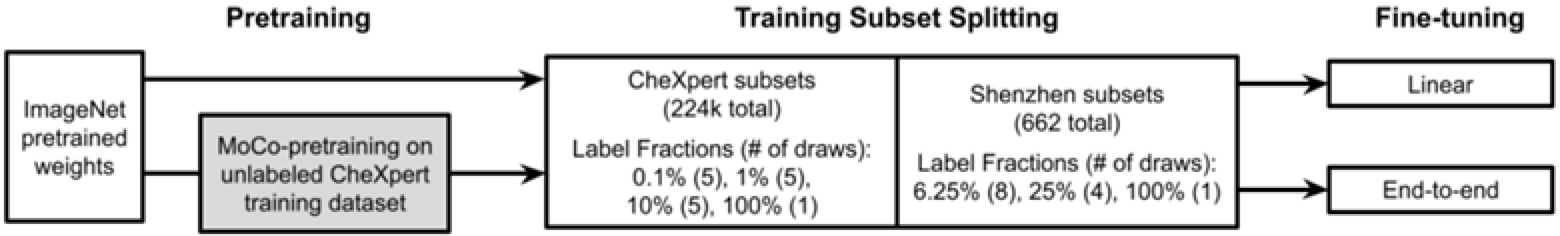
- Many studies were found to be lacking in baseline and ablation experiments
  - Out of 79 studies, 60 compared with a supervised baseline, but only 33 with another self-supervised baseline
  - Of the 33 studies, 26 compared with a self-supervised category different from their best performing model. Of these, 16 compared with SimCLR, 11 with autoencoders, and 9 with MoCo.
- Only 18 out of 79 studies reported results using natural image pre-trained weights (either supervised or self-supervised) to initialize their model for subsequent in-domain self-supervised pre-training
- Only 13 studies compared performance between classification using extracted features to end-to-end fine-tuning

End-to-end fine-tuning typically shows better results in these studies. However, stronger empirical study of different setups is needed (including tuning only later layers), and across different SSL methods.

Huang et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. Npj Digital Medicine, 2023.

MoCo-CXR: Contrastive self-supervised pre-training on chest X-rays

- Chest X-rays were some of the earliest large-scale datasets to be actively curated for AI model training
- Datasets contained up to several hundred thousand X-rays at the time of this work. (Nowadays, chest X-ray datasets have…)
- MoCo-CXR performed self-supervised training on the CheXpert dataset (224K scans from 65K patients), and evaluated on both CheXpert and the Shenzhen Hospital X-ray set (662 scans, of which 336 are abnormal for tuberculosis).

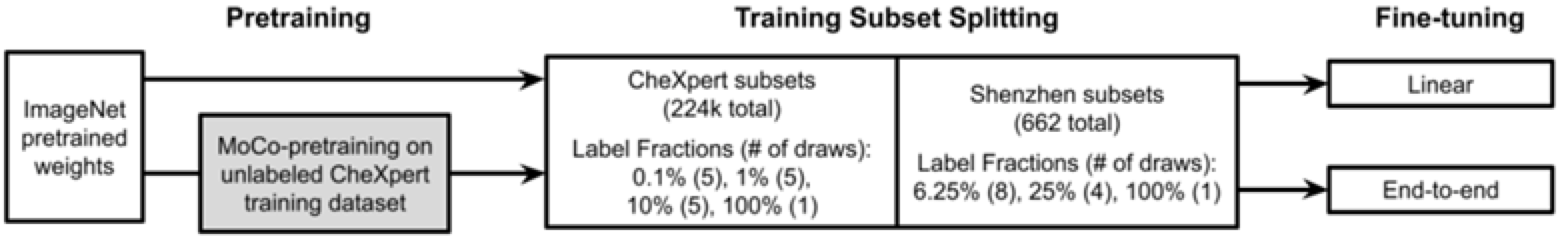


Sowrirajan et al. MoCo-CXR: MoCo Pretraining Improves Representation and Transferability of Chest X-ray Models. MIDL 2021.

MoCo-CXR: Contrastive self-supervised pre-training on chest X-rays

Selected MoCo due to compute benefits: able to train on a single NVIDIA GTX 1070 with a batch size of 16

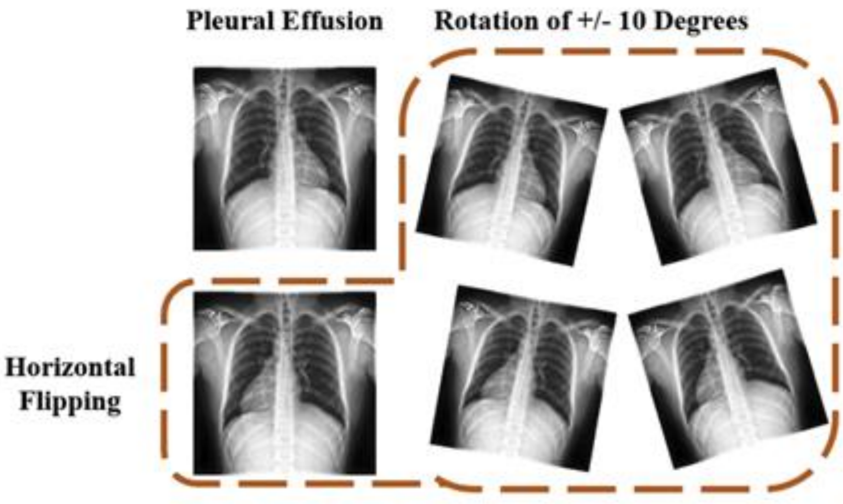
- Chest X-rays were some of the earliest large-scale datasets to be actively curated for AI model training
- Datasets contained up to several hundred thousand X-rays at the time of this work. (Nowadays, chest X-ray datasets have…)
- MoCo-CXR performed self-supervised training on the CheXpert dataset (224K scans from 65K patients), and evaluated on both CheXpert and the Shenzhen Hospital X-ray set (662 scans, of which 336 are abnormal for tuberculosis).



Sowrirajan et al. MoCo-CXR: MoCo Pretraining Improves Representation and Transferability of Chest X-ray Models. MIDL 2021.

Adapting data augmentation strategy for chest X-rays

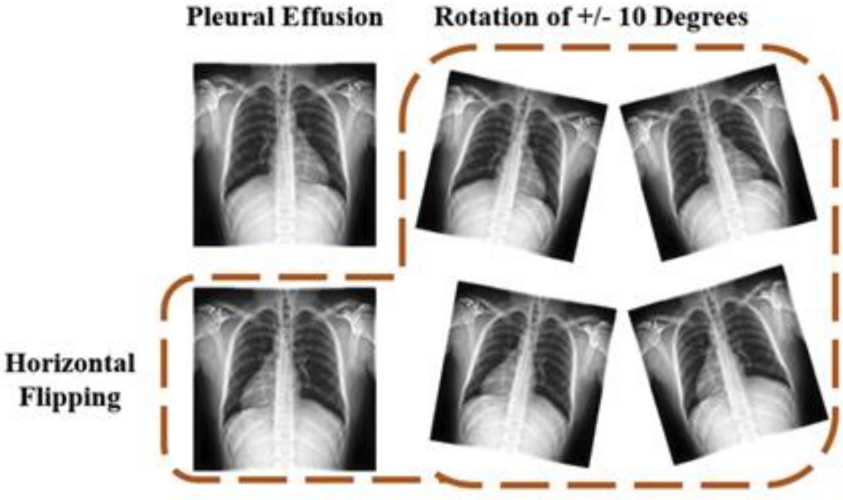
- Standard augmentation strategies for self-supervised learning in natural images may not be appropriate for chest X-rays
  - Cropping or Gaussian blurring could change the disease label or make it impossible to distinguish between diseases
  - Color jittering and random grayscale are not meaningful for X-rays that are already grayscale
- Instead, use random rotation by 10 degrees, and horizontal flipping here



Sowrirajan et al. MoCo-CXR: MoCo Pretraining Improves Representation and Transferability of Chest X-ray Models. MIDL 2021.

Adapting data augmentation strategy for chest X-rays

- Standard augmentation strategies for self-supervised learning in natural images may not be appropriate for chest X-rays
  - Cropping or Gaussian blurring could change the disease label or make it impossible to distinguish between diseases
  - Color jittering and random grayscale are not meaningful for X-rays that are already grayscale
- Instead, use random rotation by 10 degrees, and horizontal flipping here

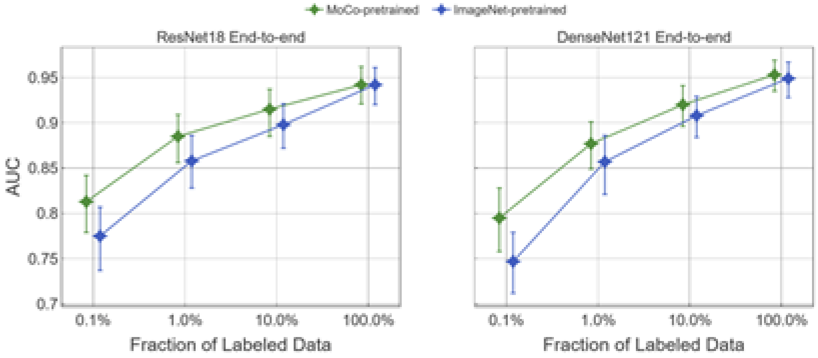


Work acknowledges opportunity for future improvement of augmentation design

Sowrirajan et al. MoCo-CXR: MoCo Pretraining Improves Representation and Transferability of Chest X-ray Models. MIDL 2021.

Comparing MoCo-pretrained vs. ImageNet-pretrained models

Predicting pleural effusion in CheXpert

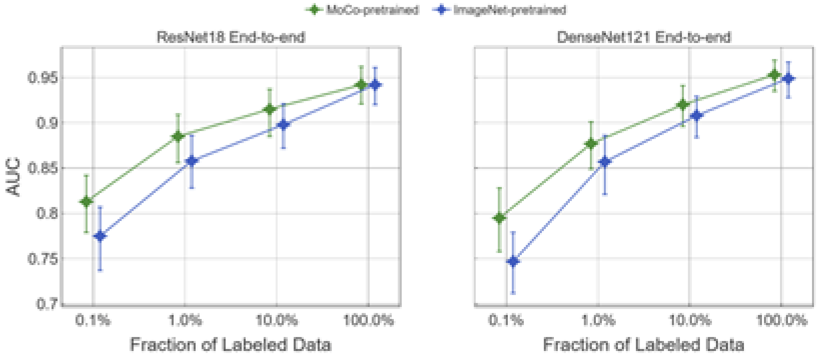


Observed similar trends for other pathology prediction tasks in CheXpert (cardiomegaly, edema, etc.)

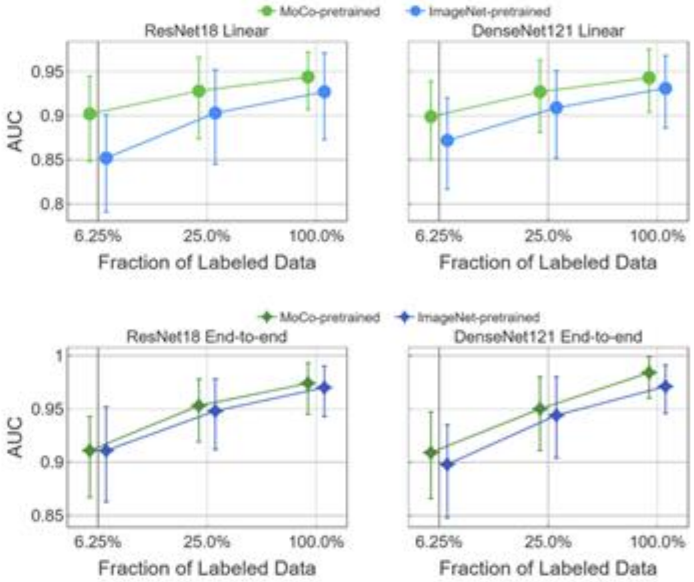
Sowrirajan et al. MoCo-CXR: MoCo Pretraining Improves Representation and Transferability of Chest X-ray Models. MIDL 2021.

Comparing MoCo-pretrained vs. ImageNet-pretrained models

Predicting pleural effusion in CheXpert



Predicting tuberculosis in the Shenzhen dataset



Observed similar trends for other pathology prediction tasks in CheXpert (cardiomegaly, edema, etc.)

Sowrirajan et al. MoCo-CXR: MoCo Pretraining Improves Representation and Transferability of Chest X-ray Models. MIDL 2021.

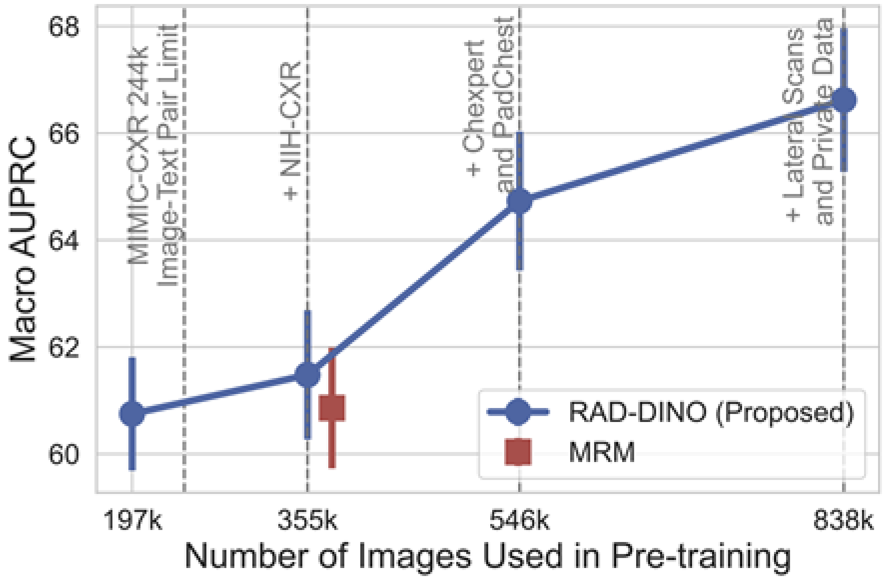
RAD-DINO: Improving scale and model for chest X-ray pre-training

- Used the DINOv2 self-supervised method, with ViT architecture. Started from public DINOv2 ViT-B/14 model and continued training with CXR images.
- Trained on an extended CXR dataset, Multi-CXR, aggregating multiple datasets including CheXpert

Dataset	View	Patient cohort	Number of subjects	Number of images
BRAX [122]	frontal, lateral	all available in institutional PACS	19,351	41,620
CheXpert [69]	frontal, lateral	inpatient and outpatient	65,240	223,648
MIMIC-CXR [17]	frontal	ICU	188,546	210,491
NIH-CXR [111]	frontal	not specified	32,717	112,120
PadChest [57]	frontal, lateral	all available	67,000	160,817
Private	frontal, lateral	outpatient	66,323	90,000
Total			439,177	838,336

Perez-Garcia et al. RAD-DINO: Exploring scalable medical image encoders beyond text supervision. 2024.

Linear probing performance on VinDr-CXR vs. number of images used in RAD-DINO pre-training



Perez-Garcia et al. RAD-DINO: Exploring scalable medical image encoders beyond text supervision. 2024.

Comparison with methods using vision-language representation learning

Model	VinDr-CXR [46] (AUPRC) LO	CM	PL-T	AE	PF	TB	Agg
CLIP@224 [2]	22.95 ± 1.17	52.91 ± 1.53	22.31 ± 2.33	56.45 ± 2.02	25.12 ± 1.35	18.41 ± 1.43	33.03
CLIP@336 [2]	23.88 ± 1.05	53.87 ± 1.40	23.71 ± 3.61	58.16 ± 1.62	27.67 ± 0.69	19.93 ± 1.99	34.54
BioViL-T [39]	34.07 ± 3.20	41.10 ± 1.94	26.50 ± 4.32	37.01 ± 1.19	32.35 ± 2.38	34.08 ± 3.65	34.19
BiomedCLIP [40]	38.33 ± 1.93	66.94 ± 0.95	29.62 ± 3.57	62.66 ± 1.90	36.78 ± 3.85	38.21 ± 4.85	45.42
MRM [7]	43.38 ± 2.36	82.04 ± 1.43	45.05 ± 4.94	75.14 ± 1.30	55.12 ± 3.44	64.42 ± 4.15	60.86
RAD-DINO	47.55 ± 2.96	82.51 ± 1.66	46.13 ± 4.98	83.53 ± 0.73	69.67 ± 2.29	70.36 ± 3.51	66.63
Δ	+4.17	+0.47	+1.08	+8.39	+14.55	+5.94	+5.77

LO: Lung Opacity, CM: Cardiomegaly, PL-T: Pleural Thickening, AE: Aortic Enlargement,
PF: Pulmonary Fibrosis, TB: Tuberculosis, Agg: Macro Average

Perez-Garcia et al. RAD-DINO: Exploring scalable medical image encoders beyond text supervision. 2024.

Comparison with methods using vision-language representation learning

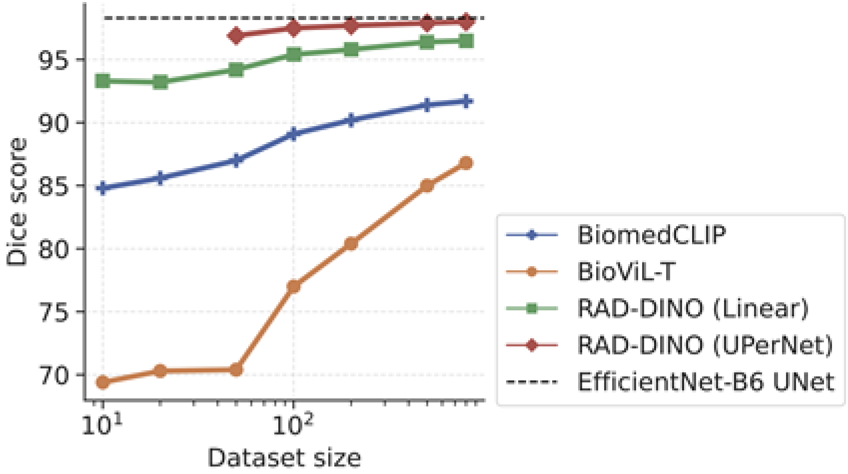
Although multimodal representation learning is very popular (e.g. with paired image and text data) and can have benefits such as zero-shot recognition, unimodal representation learning using only image data can outperform these methods for fine-tuning representations on downstream tasks.

- Unimodal methods can leverage larger amounts of data
- Text supervision may not be detailed enough to learn representations that are effective for precise image analysis (e.g. segmenting small abnormalities)
- Text alignment can lead to undesired invariances to anatomical variations, e.g. the phrase “No cardiopulmonary process” for healthy individuals that have variation in appearance

Perez-Garcia et al. RAD-DINO: Exploring scalable medical image encoders beyond text supervision. 2024.

Comparison with methods using vision-language representation learning

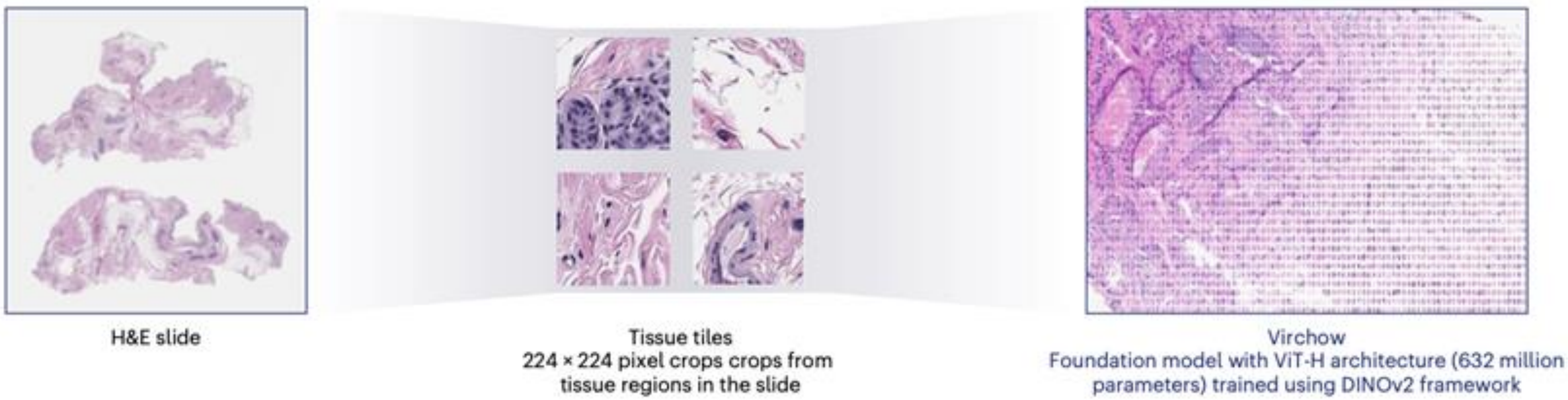
- Comparing transfer to semantic segmentation of right vs left lung using dataset derived from MIMIC-CXR



Perez-Garcia et al. RAD-DINO: Exploring scalable medical image encoders beyond text supervision. 2024.

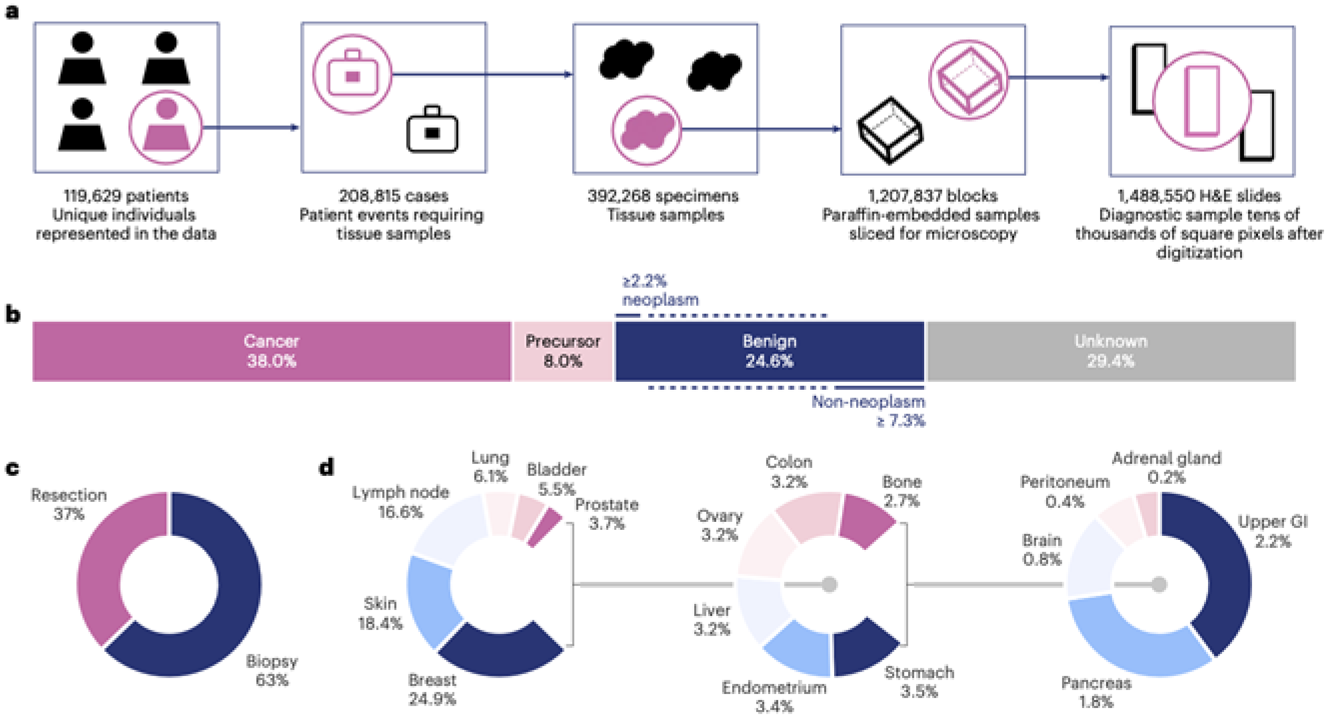
Virchow: Large-scale self-supervised learning on digital pathology images

- Trained a DINO v2 based model on 1.5 million H&E stained whole slide images (WSIs) from Memorial Sloan Kettering Cancer Center and external consults, corresponding to ~100,000 patients



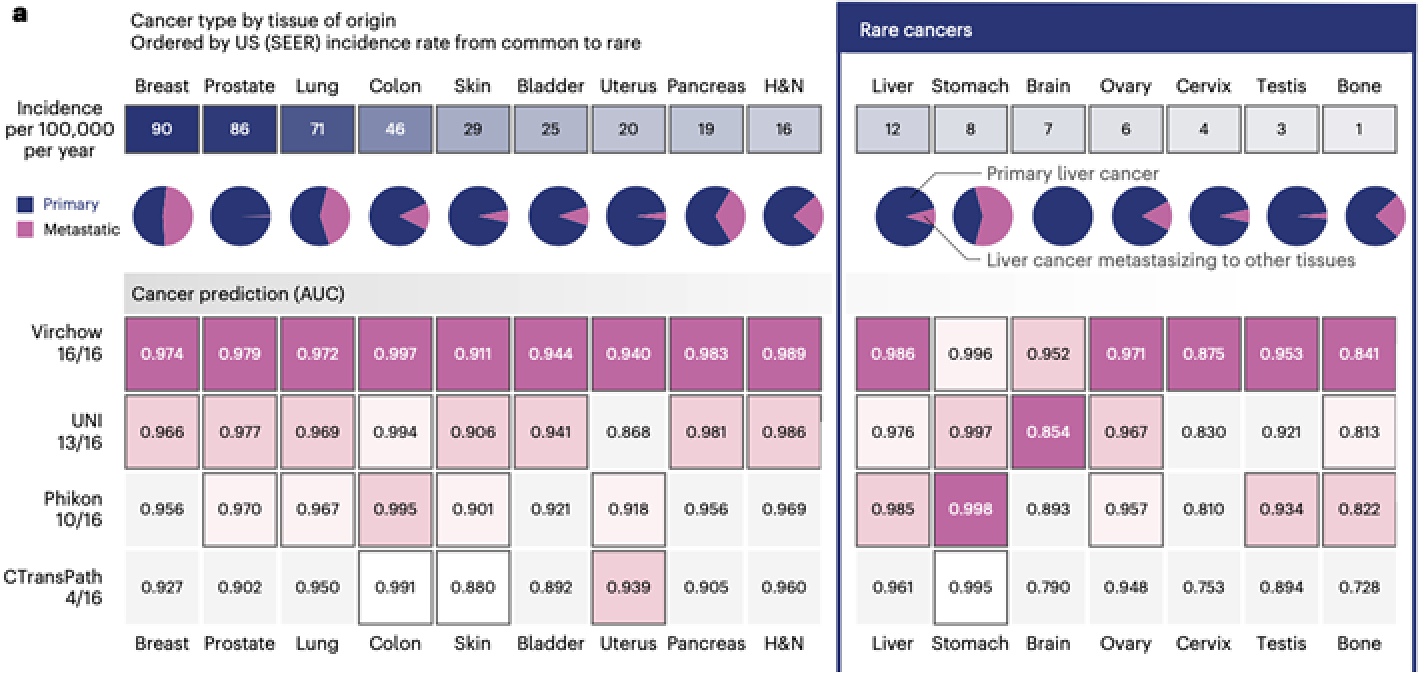
Vorontsov et al. A foundation model for clinical-grade computational pathology and rare cancers detection. Nature Medicine, 2024.

Characteristics of self-supervised training data



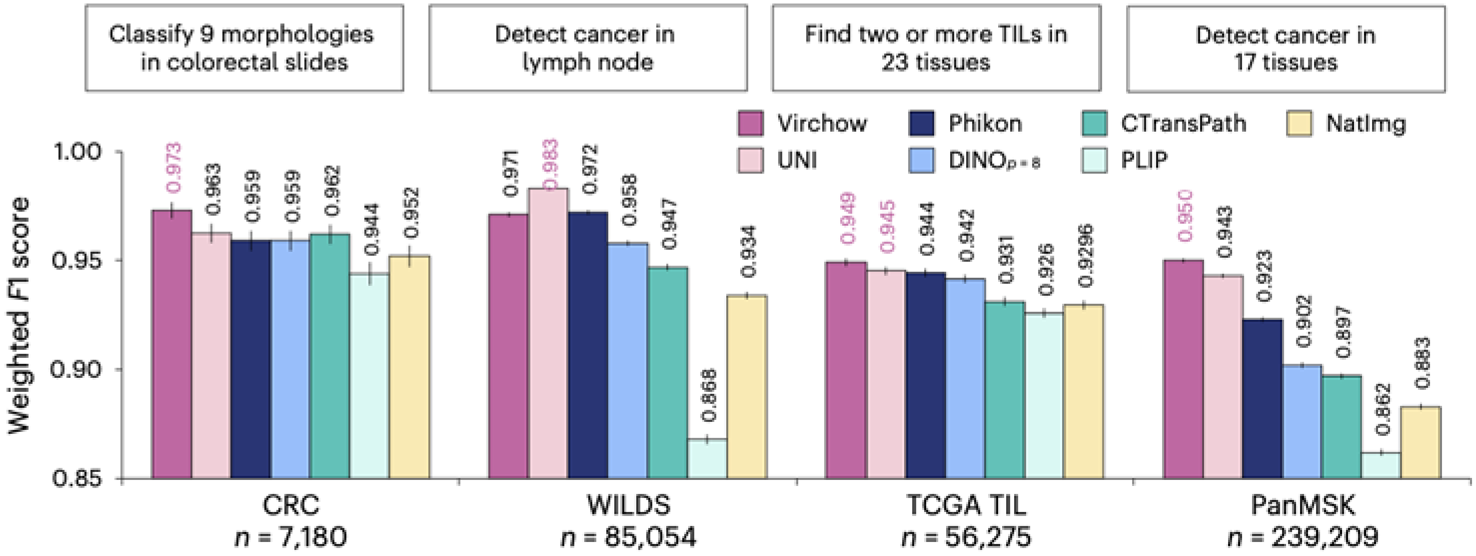
Vorontsov et al. A foundation model for clinical-grade computational pathology and rare cancers detection. Nature Medicine, 2024.

“Pan-cancer detection” across nine common and seven rare cancers



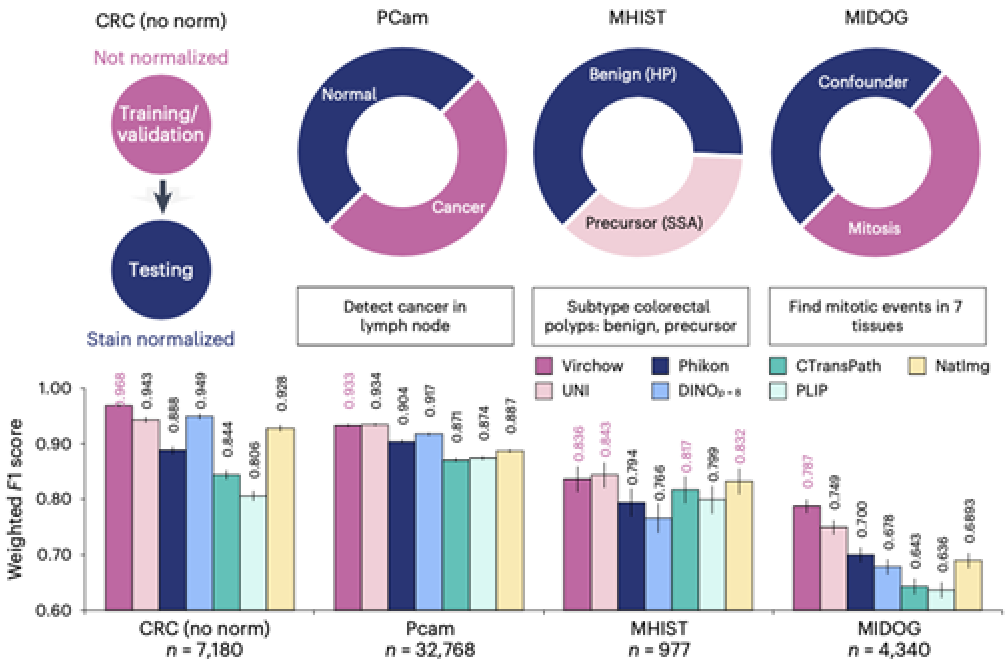
Vorontsov et al. A foundation model for clinical-grade computational pathology and rare cancers detection. Nature Medicine, 2024.

Comparison with baselines including vision-language models on varied tasks



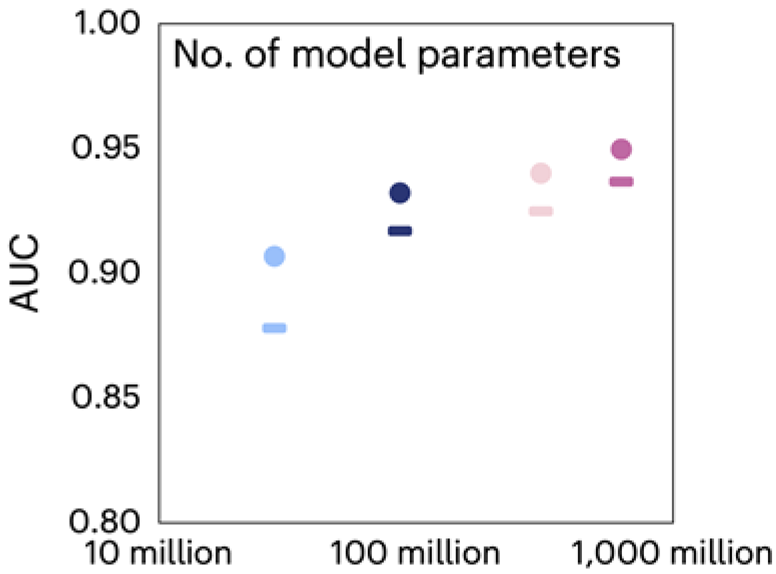
Vorontsov et al. A foundation model for clinical-grade computational pathology and rare cancers detection. Nature Medicine, 2024.

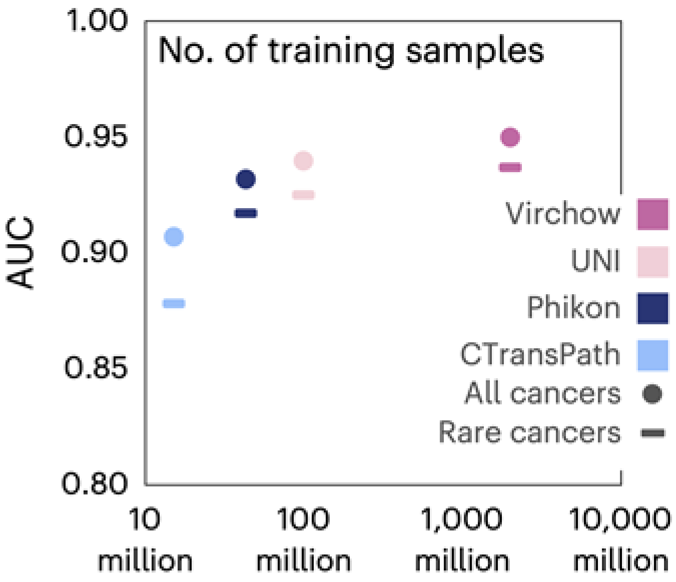
Comparison with baselines including vision-language models on varied tasks



Vorontsov et al. A foundation model for clinical-grade computational pathology and rare cancers detection. Nature Medicine, 2024.

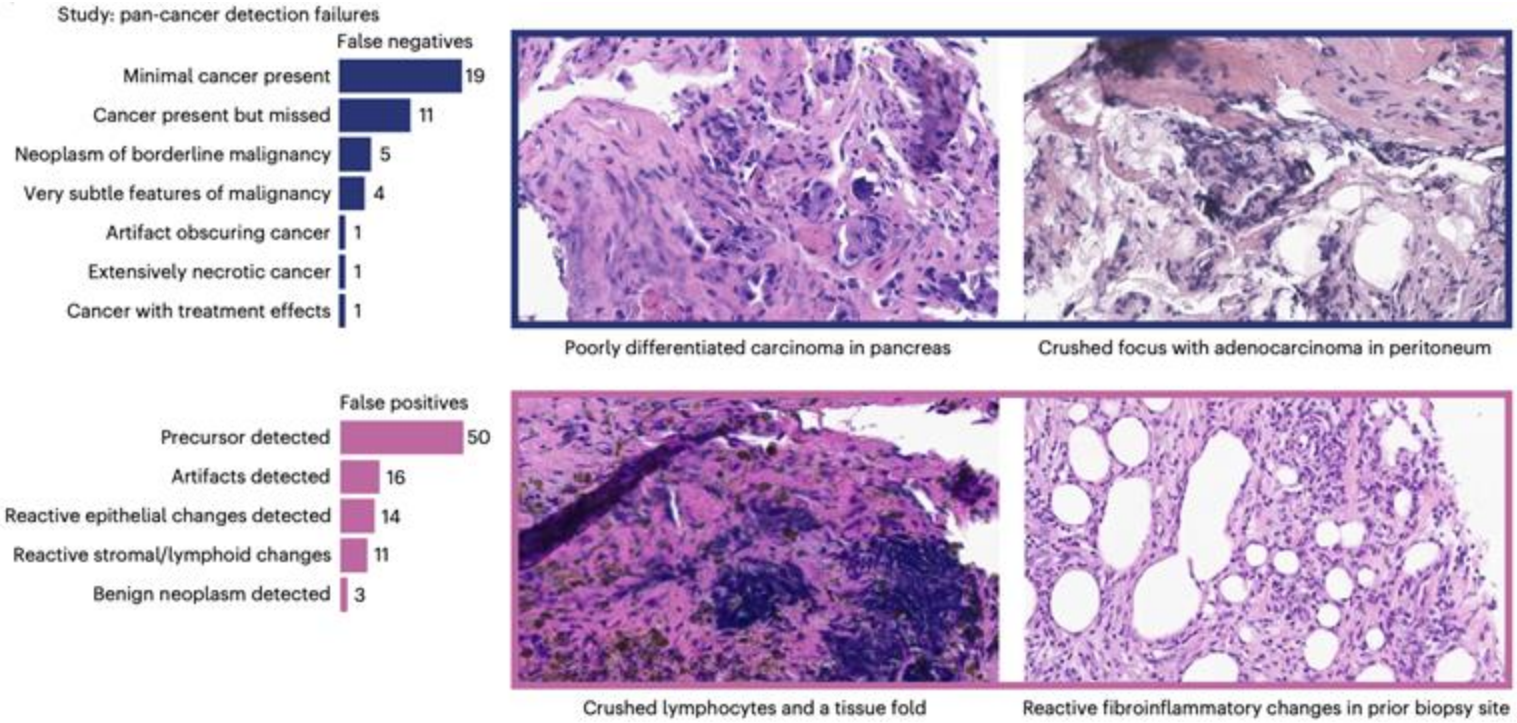
Foundation model scaling behavior





Vorontsov et al. A foundation model for clinical-grade computational pathology and rare cancers detection. Nature Medicine, 2024.

Analysis of model failures



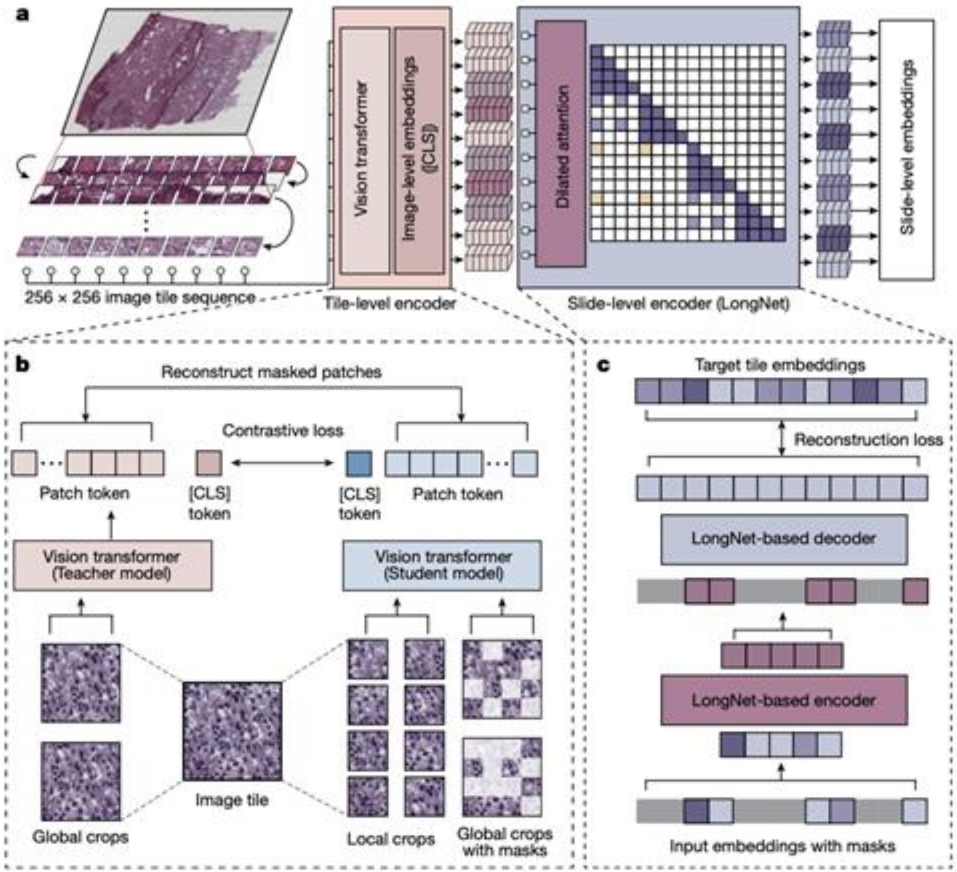
Vorontsov et al. A foundation model for clinical-grade computational pathology and rare cancers detection. Nature Medicine, 2024.

Prov-GigaPath: Another large pathology foundation model

- Trained on 171,189 whole slide images from 30,000 patients, corresponding to 1.3 billion slide tiles, from the Providence health system comprising 28 cancer centres. Slides cover 31 major tissue types.
- Work also introduces a digital pathology benchmark spanning 26 prediction tasks such as mutation prediction and cancer subtyping, using data from Providence ad TCGA. Achieves state-of-the-art on 25 out of 26 tasks.
- Adopts the LongNet model to digital pathology to scale to slide-level learning with thousands of image tiles

Xu et al. A whole-slide foundation model for digital pathology from real-world data. Nature, 2024.

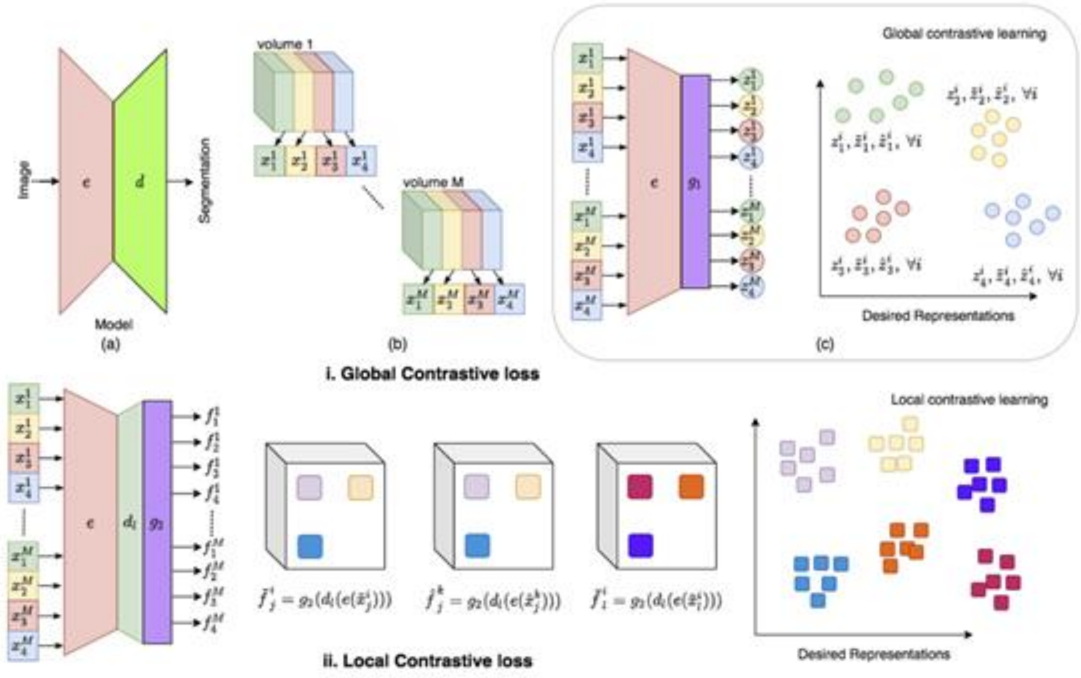
GigaPath model



Xu et al. A whole-slide foundation model for digital pathology from real-world data. Nature, 2024.

Self-supervised learning for volumetric data (e.g. CT and MRI)

- Challenge: much less data (even unlabeled) is typically available for 3D
- One option: continue to learn 2D representations, but incorporate 3D by encouraging additional relationships among slices of a volume
- Example of doing this within a SimCLR framework

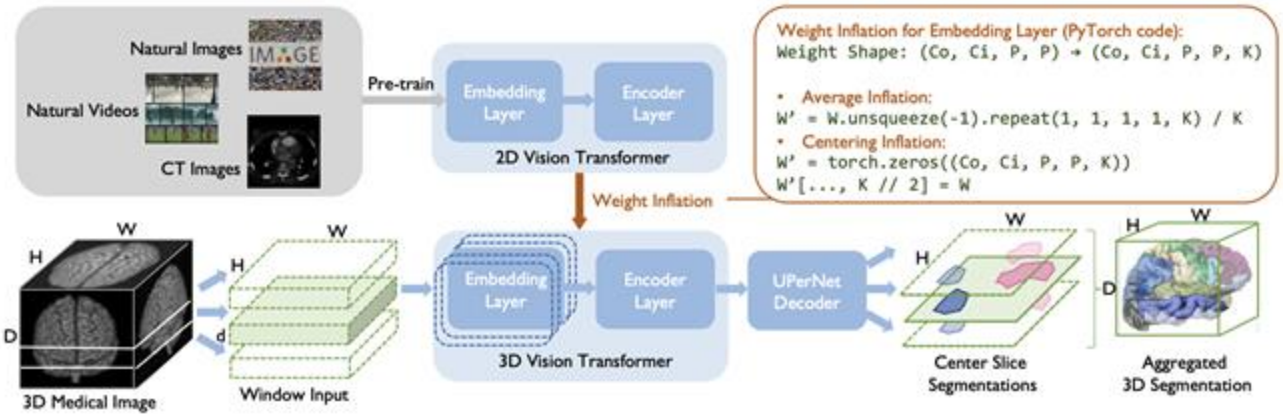


Chaitanya et al. Contrastive learning of global and local features for medical image segmentation with limited annotations. NeurIPS 2020.

Another approach: leverage pre-trained natural image models

Options include 2D natural image models (inflate weights to 3D), or video models (time is naturally a 3rd dimension). Can often improve over pre-training only on limited 3D medical data.

Often, the volumes that are used for model training and inference comprise only a subset of neighboring depth slices (as opposed to an entire 3D MRI or CT) to reduce computational needs. This may be sufficient for segmentation, or can be later aggregated if needed (e.g. for classification).



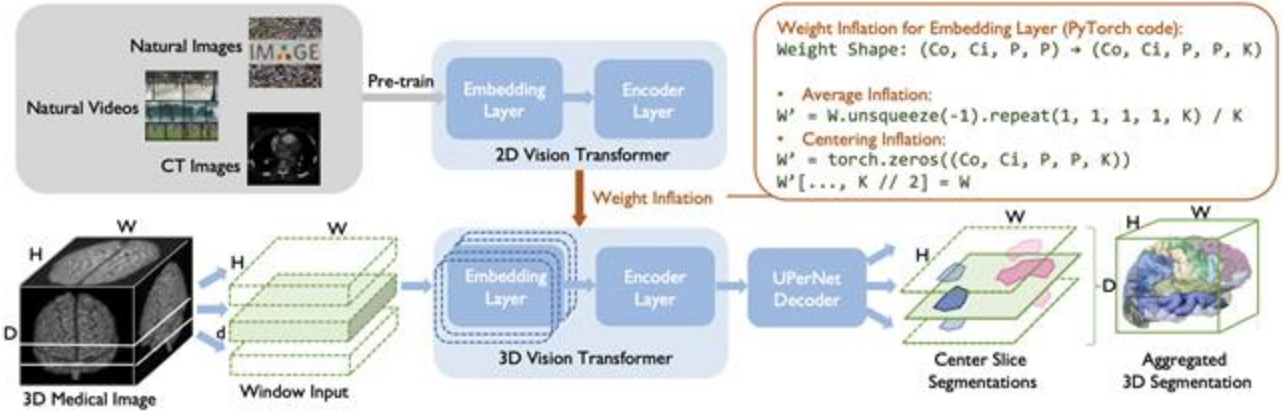
Zhang et al. Adapting Pre-trained Vision Transformers from 2D to 3D through Weight Inflation Improves Medical Image Segmentation. ML4H 2022.

Another approach: leverage pre-trained natural image models

Options include 2D natural image models (inflate weights to 3D), or video models (time is naturally a 3rd dimension). Can often improve over pre-training only on limited 3D medical data.

Often, the volumes that are used for model training and inference comprise only a subset of neighboring depth slices (as opposed to an entire 3D MRI or CT) to reduce computational needs. This may be sufficient for segmentation, or can be later aggregated if needed (e.g. for classification).

Note: in this particular work, the pre-training is supervised not self-supervised, but the transfer paradigms are still relevant



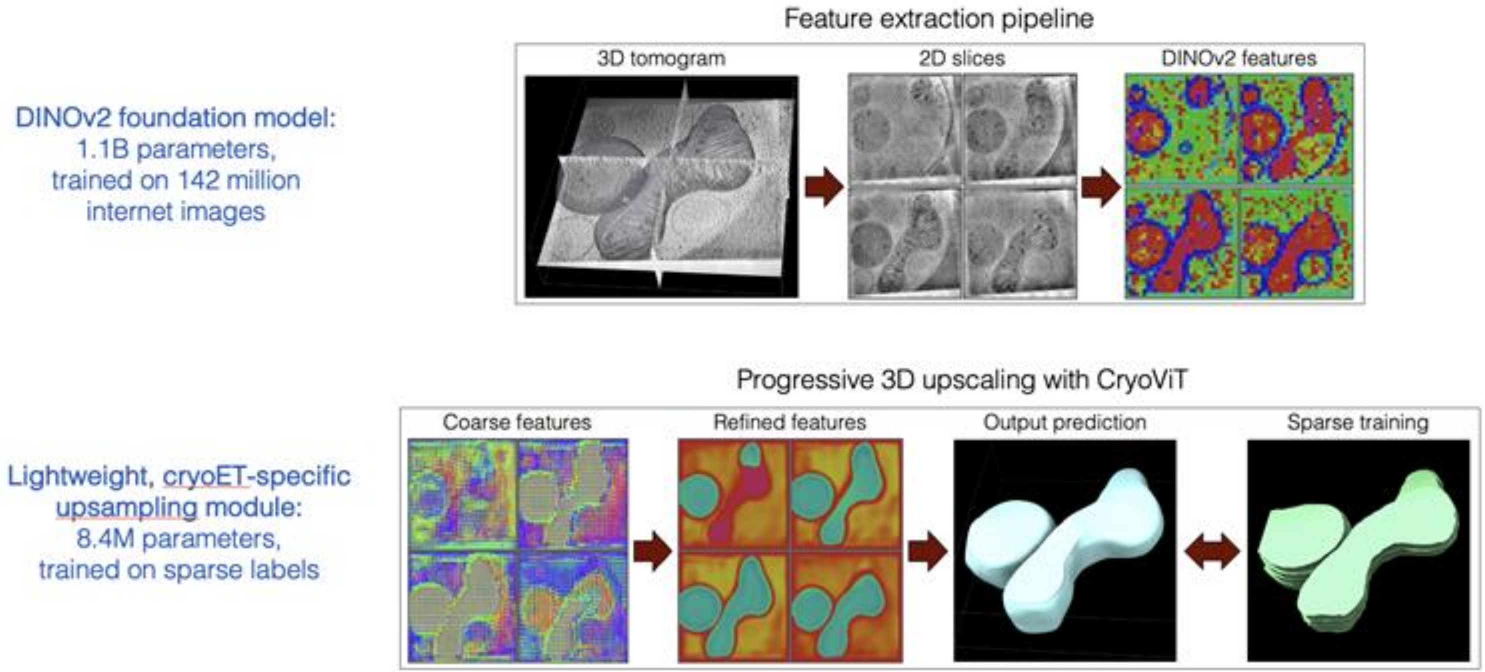
Zhang et al. Adapting Pre-trained Vision Transformers from 2D to 3D through Weight Inflation Improves Medical Image Segmentation. ML4H 2022.

CryoViT: Leveraging 2D foundation models for 3D segmentation in cryo-ET tomograms

- Cryo-electron tomography (cryoET): nanometer-resolution volumes of cells flash-frozen and imaged at cryogenic temperatures (-150 degrees Celsius or lower), allows visualizing subcellular structures in their native structural state
- Dataset size: 256 tomograms of Huntington’s Disease iPSC-derived neurons. Large relative to cryoET domain, small relative to foundation models
- Leveraged 2D, natural image-pretrained DINO v2 features, then performed semi-supervised learning using 3D cryoET training data (4% of 2D slices were labeled)

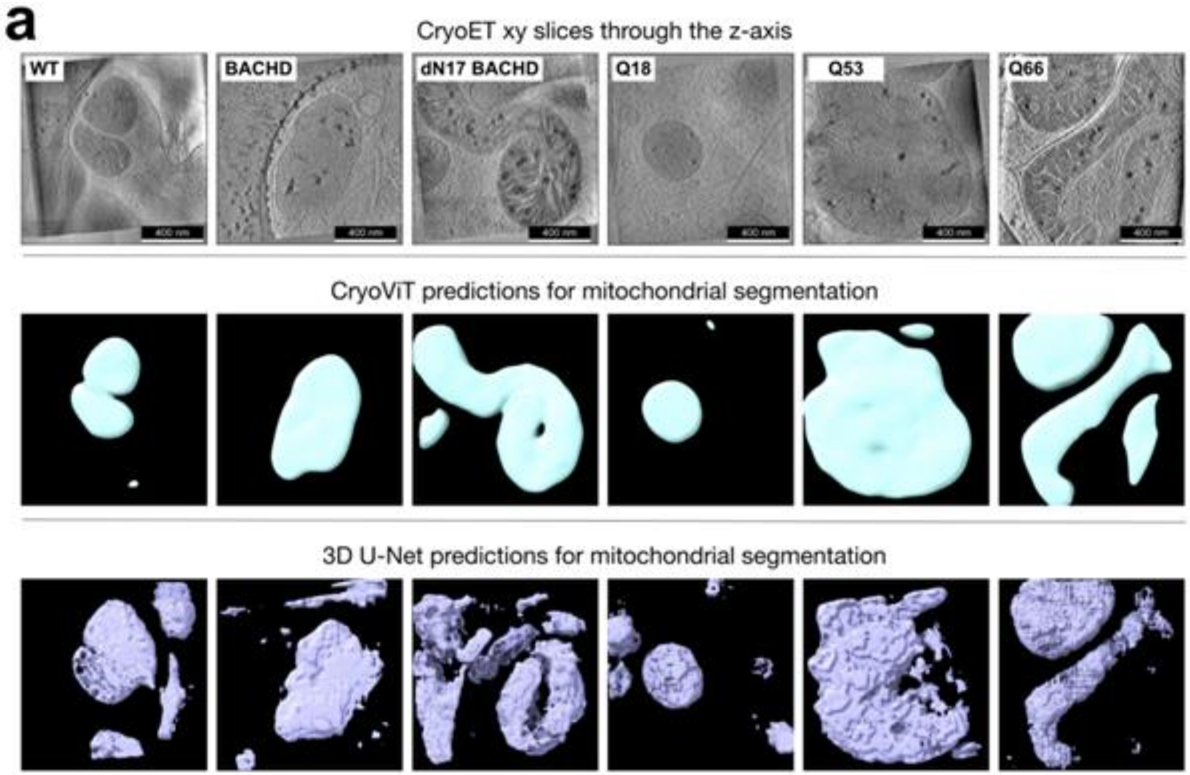
Gupte et al. CryoViT: Efficient Segmentation of Cryogenic Electron Tomograms with Vision Foundation Models. arXiv, 2024.

Cryo-VisionTransformer (Cryo-ViT)



Gupte et al. CryoViT: Efficient Segmentation of Cryogenic Electron Tomograms with Vision Foundation Models. arXiv, 2024.

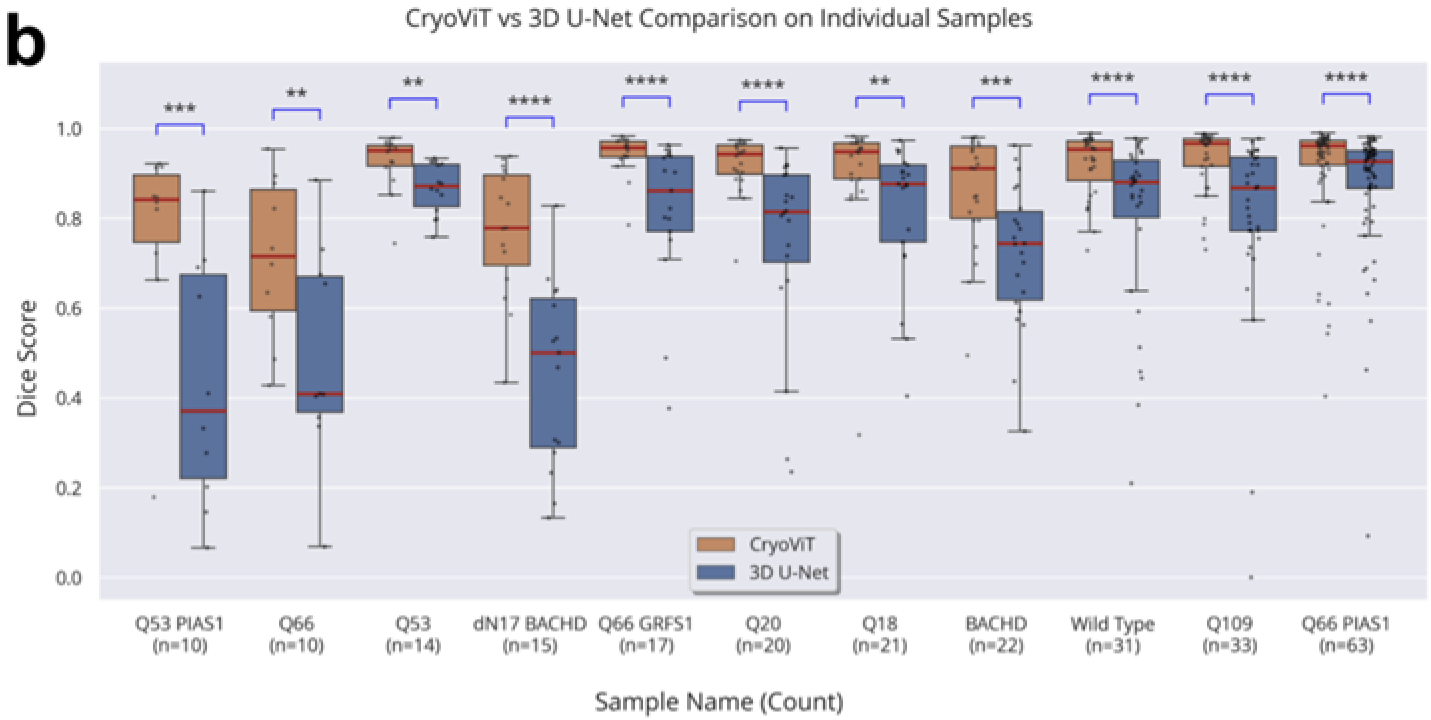
3D segmentations of CryoViT vs 3D U-Net



Gupte et al. CryoViT: Efficient Segmentation of Cryogenic Electron Tomograms with Vision Foundation Models. arXiv, 2024.

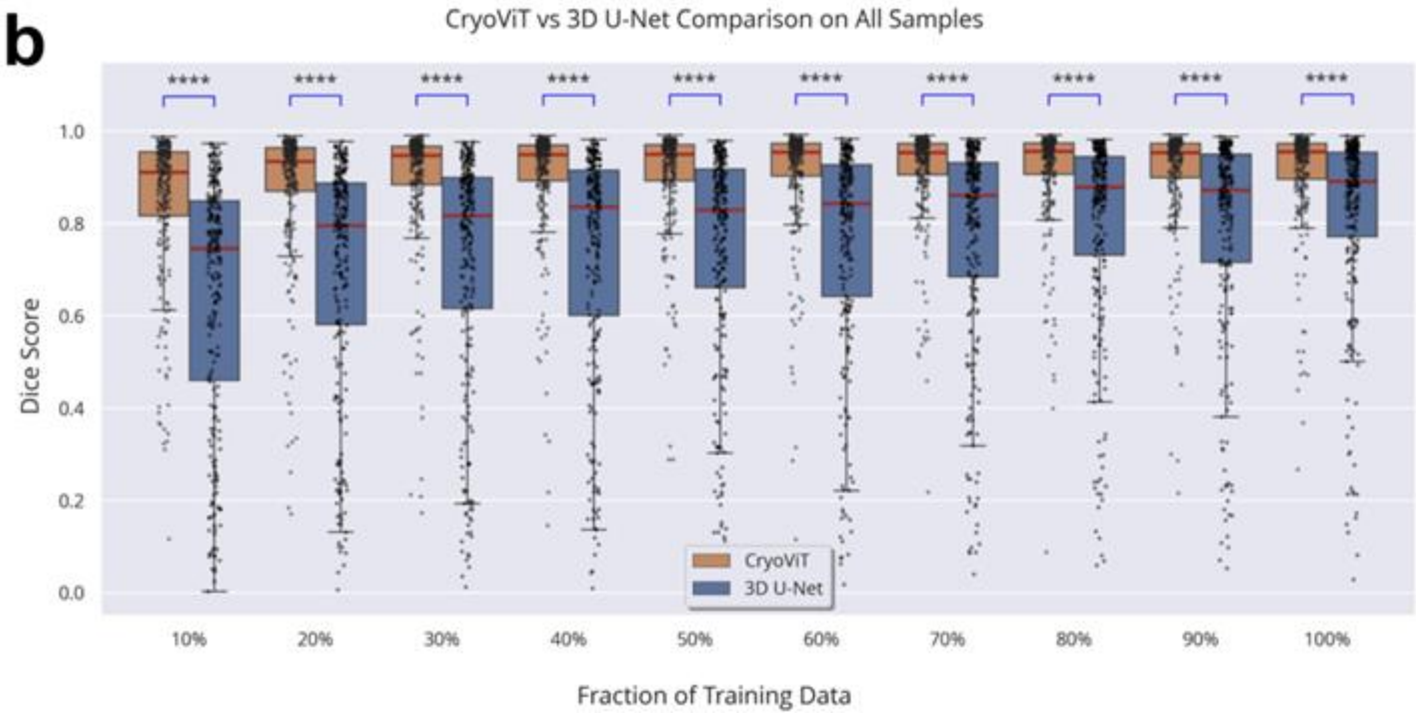
Dice Score across different sample types

b



Gupte et al. CryoViT: Efficient Segmentation of Cryogenic Electron Tomograms with Vision Foundation Models. arXiv, 2024.

Performance vs. fraction of labeled training data used



Gupte et al. CryoViT: Efficient Segmentation of Cryogenic Electron Tomograms with Vision Foundation Models. arXiv, 2024.

Endo-FM: Foundation model for endoscopy video analysis

- Endoscopy video foundation model trained on 33K video clips comprising 5 million video frames, aggregated through 9 publicly available dataset and a privately collected dataset from Renji Hospital in Shanghai
- Self-supervised training method based on DINO, adapted for video using spatial as well as temporal view transformations

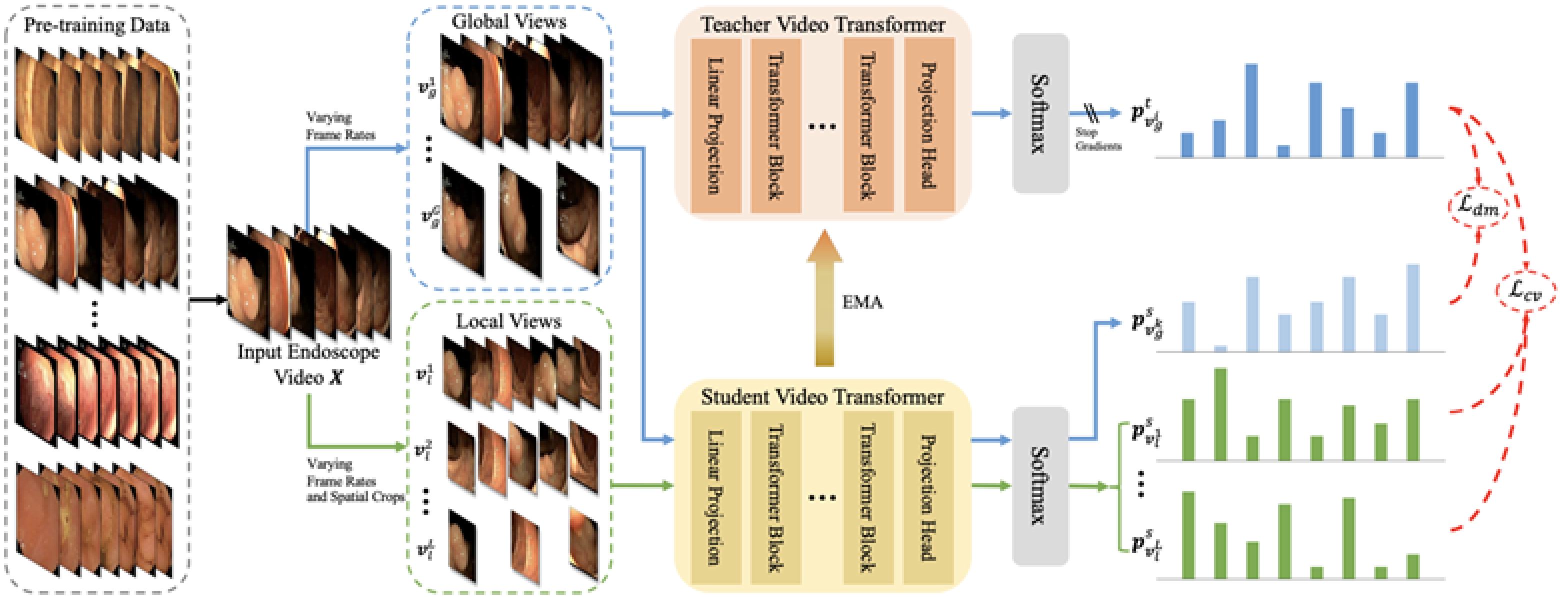
Wang et al. Foundation Model for Endoscopy Video Analysis via Large-scale Self-supervised Pre-train. MICCAI 2023.

Pre-training and downstream datasets and tasks

Phase	Dataset	Provider	Videos	Frames	Protocol	Disease
Pre-train	Colonoscopic [19]	CNRS	210	36534	colonoscope	adenoma, hyperplasia
	SUN [20] & SUN-SEG [13]	ANU	1018	159400	colonoscope	SSL, adenoma, hyperplasia, T1b
	LDPolypVideo [18]	USTC	237	40186	colonoscope	polyp
	Hyper-Kvasir [5]	Simula	5704	875940	gastroscope	barrett's oesophagus, polyp, cancer
	Kvasir-Capsule [31]	Simula	1000	158892	gastroscope	erosion, erythema, etc.
	CholecTriplet [24]	BIDMC	580	90444	laparoscope	cholecystectomy
	Ours	Baoshan Branch of Renji Hospital	16494 7653	2491952 1170753	colonoscope gastroscope	polyp, erosion, etc.
	Summary	6 providers	32896	5024101	3 protocols	10+ diseases
Downstream	PolypDiag [33]	Adelaide	253	485561	gastroscope	polyp, cancer
	CVC-12k [2]	UAB	29	612	colonoscope	polyp
	KUMC [15]	Kansas	53	19832	colonoscope	adenoma, hyperplasia
	Summary	3 providers	335	506005	2 protocols	4 diseases

Wang et al. Foundation Model for Endoscopy Video Analysis via Large-scale Self-supervised Pre-train. MICCAI 2023.

Endo-FM model



Wang et al. Foundation Model for Endoscopy Video Analysis via Large-scale Self-supervised Pre-train. MICCAI 2023.

Comparison with other SOTA methods on downstream tasks

Table 2. Comparison with other latest SOTA methods on 3 downstream tasks. We report F1 score (%) for PolypDiag, Dice (%) for CVC-12k, and F1 score (%) for KUMC.

Method	Venue	Pre-training Time (h)	PolypDiag (Classification)	CVC-12k (Segmentation)	KUMC (Detection)
Scratch (Rand. init.)		N/A	83.5±1.3	53.2±3.2	73.5±4.3
TimeSformer [3]	ICML'21	104.0	84.2±0.8	56.3±1.5	75.8±2.1
CORP [12]	ICCV'21	65.4	87.1±0.6	68.4±1.1	78.2±1.4
FAME [7]	CVPR'22	48.9	85.4±0.8	67.2±1.3	76.9±1.2
ProViCo [26]	CVPR'22	71.2	86.9±0.5	69.0±1.5	78.6±1.7
VCL [27]	ECCV'22	74.9	87.6±0.6	69.1±1.2	78.1±1.9
ST-Adapter [25]	NeurIPS'22	8.1	84.8±0.7	64.3±1.9	74.9±2.9
Endo-FM (Ours)		20.4	**90.7±0.4**	**73.9±1.2**	**84.1±1.3**

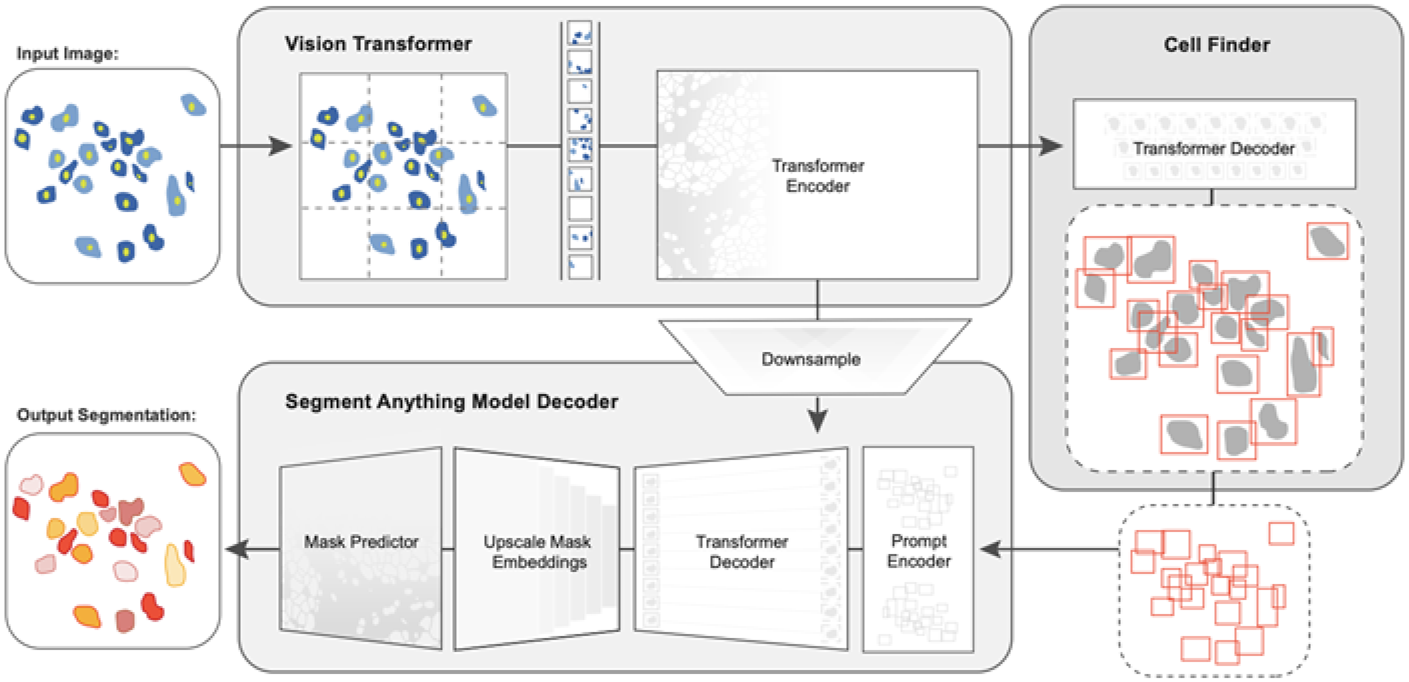
Wang et al. Foundation Model for Endoscopy Video Analysis via Large-scale Self-supervised Pre-train. MICCAI 2023.

CellSAM: An example of Segment Anything (SAM) adapted for cell segmentation

- Fine-tuned SAM on a collection of public cell image datasets and a private dataset to improve ability to segment cells
- Final CellSAM model combines a cell detection model (Cell Finder) with the fine-tuned SAM by using CellFinder detections as input prompts to SAM

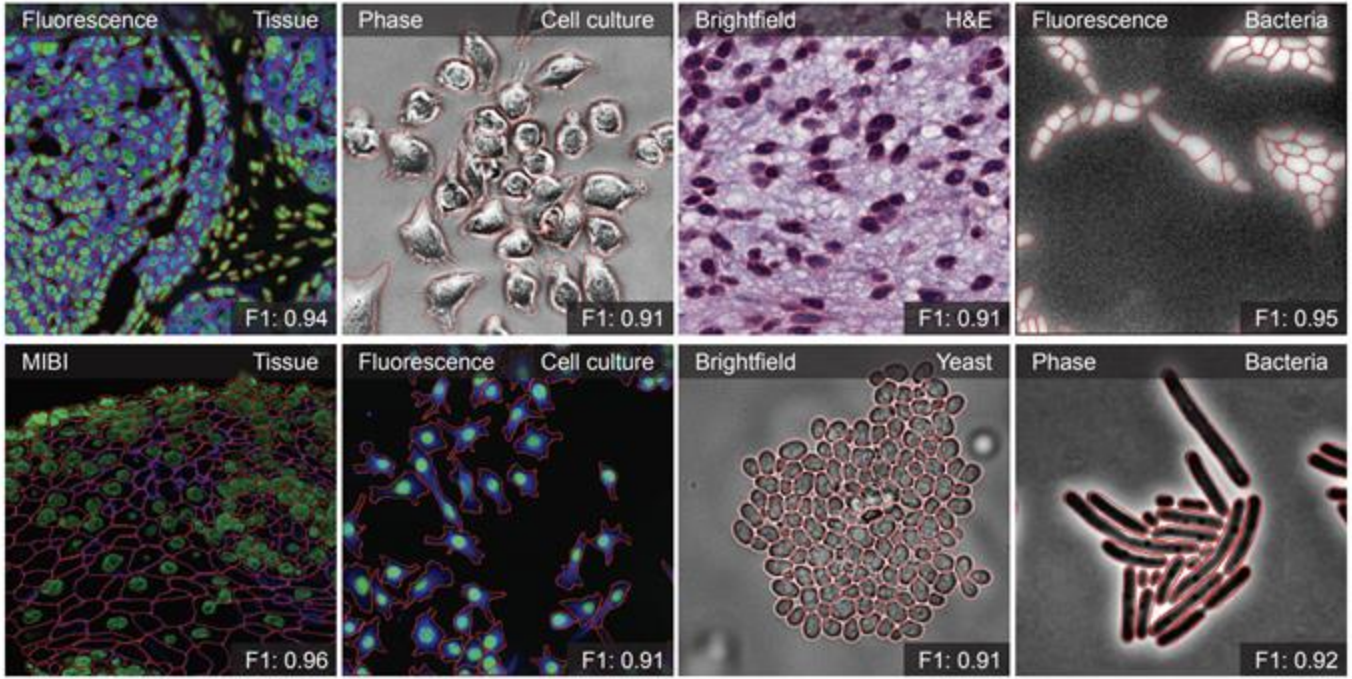
Israel et al. A Foundation Model for Cell Segmentation. arXiv 2023.

CellSAM approach



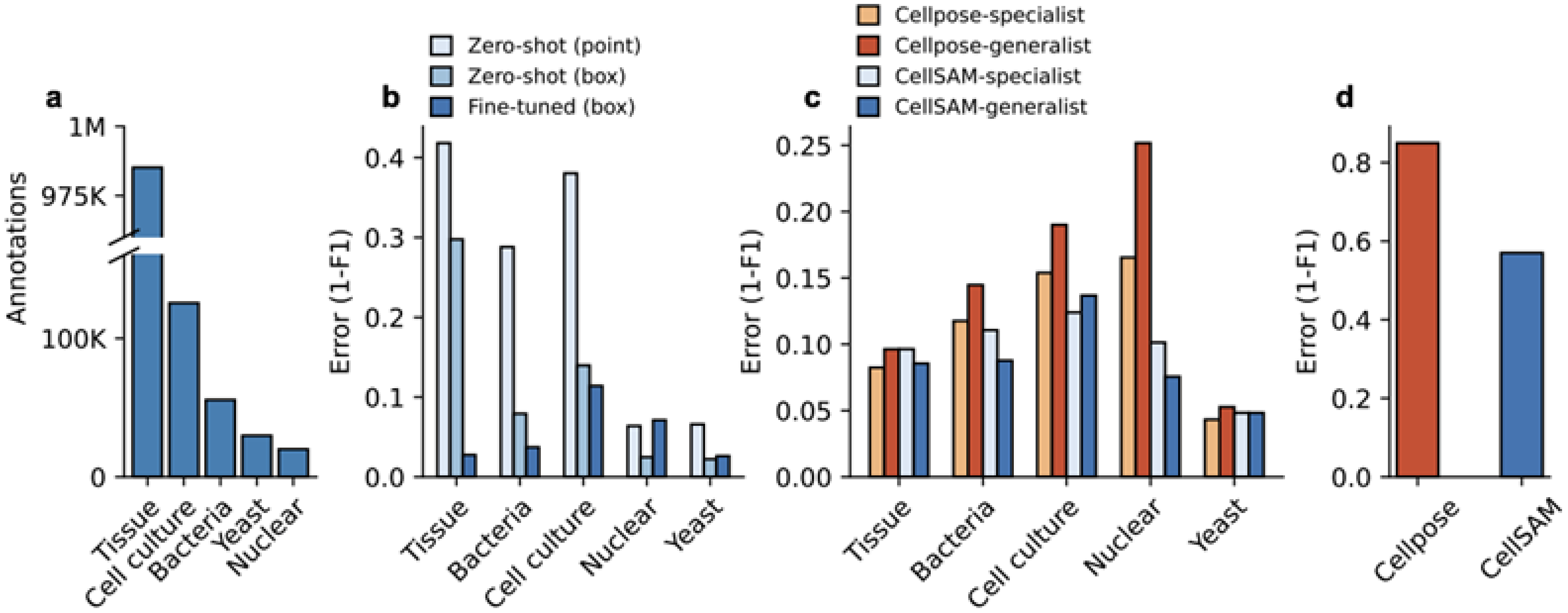
Israel et al. A Foundation Model for Cell Segmentation. arXiv 2023.

Example outputs for different data and imaging modalities



Israel et al. A Foundation Model for Cell Segmentation. arXiv 2023.

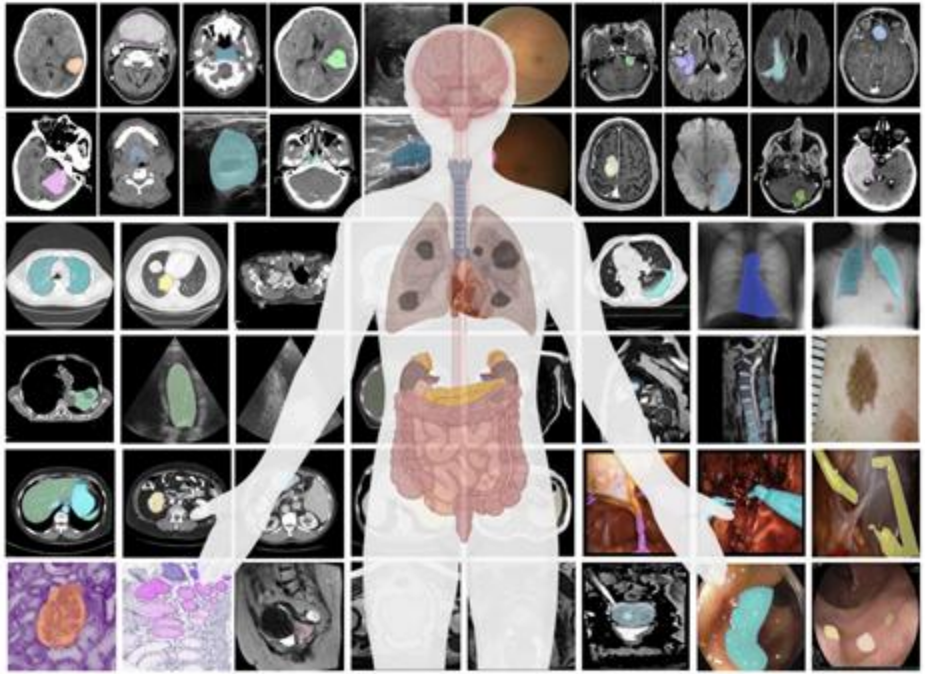
Comparison with baseline approaches



Israel et al. A Foundation Model for Cell Segmentation. arXiv 2023.

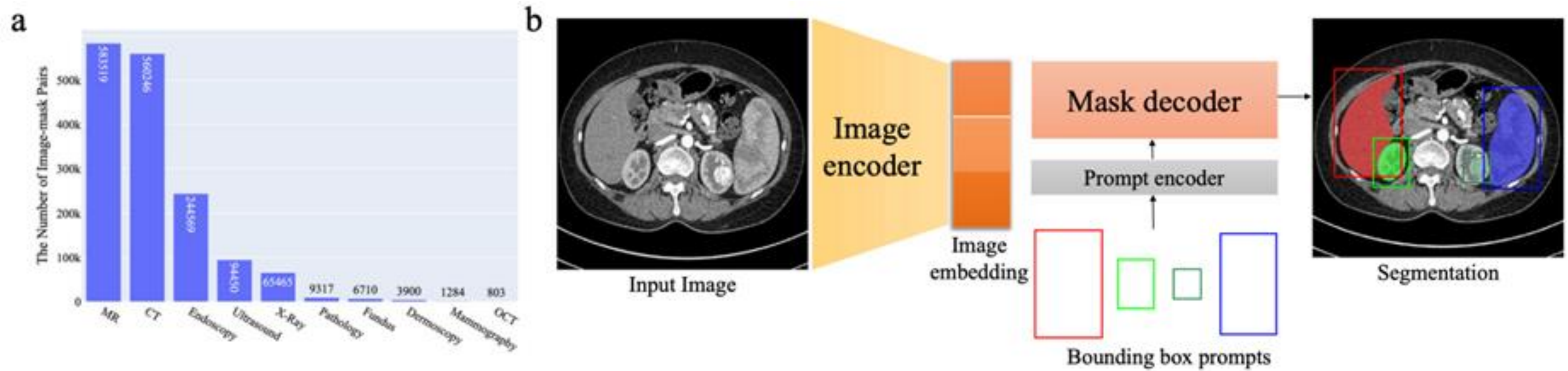
MedSAM: Segment anything in medical images

- Tackles universal medical image segmentation across many imaging modalities and tasks
- Trained on a large-scale medical image dataset with 1,570,263 image-mask pairs, covering 10 imaging modalities and over 30 cancer types
- Evaluated on 86 internal validation and 60 external validation tasks



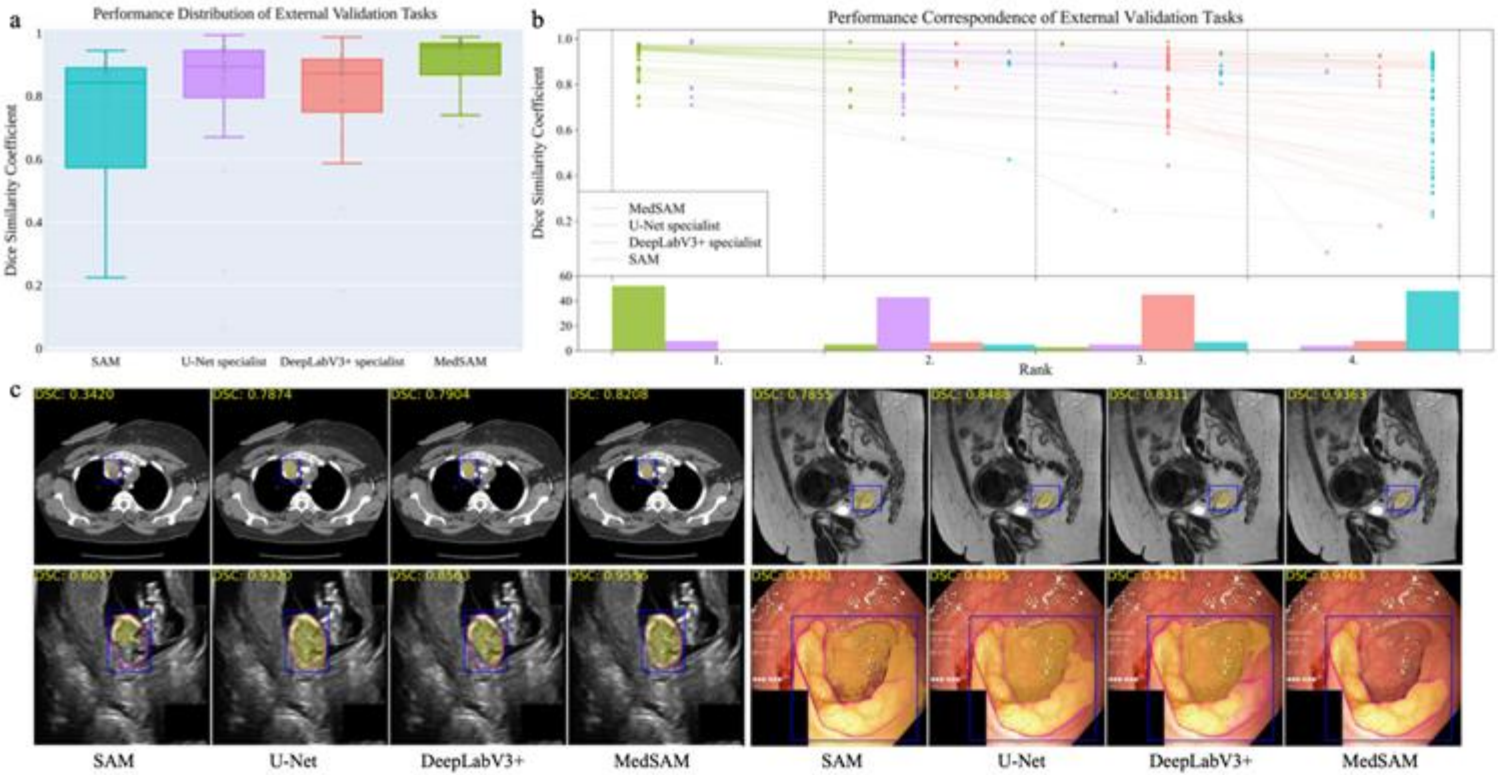
Ma et al. Segment anything in medical images. Nature Communications, 2024.

Modality distribution of training dataset



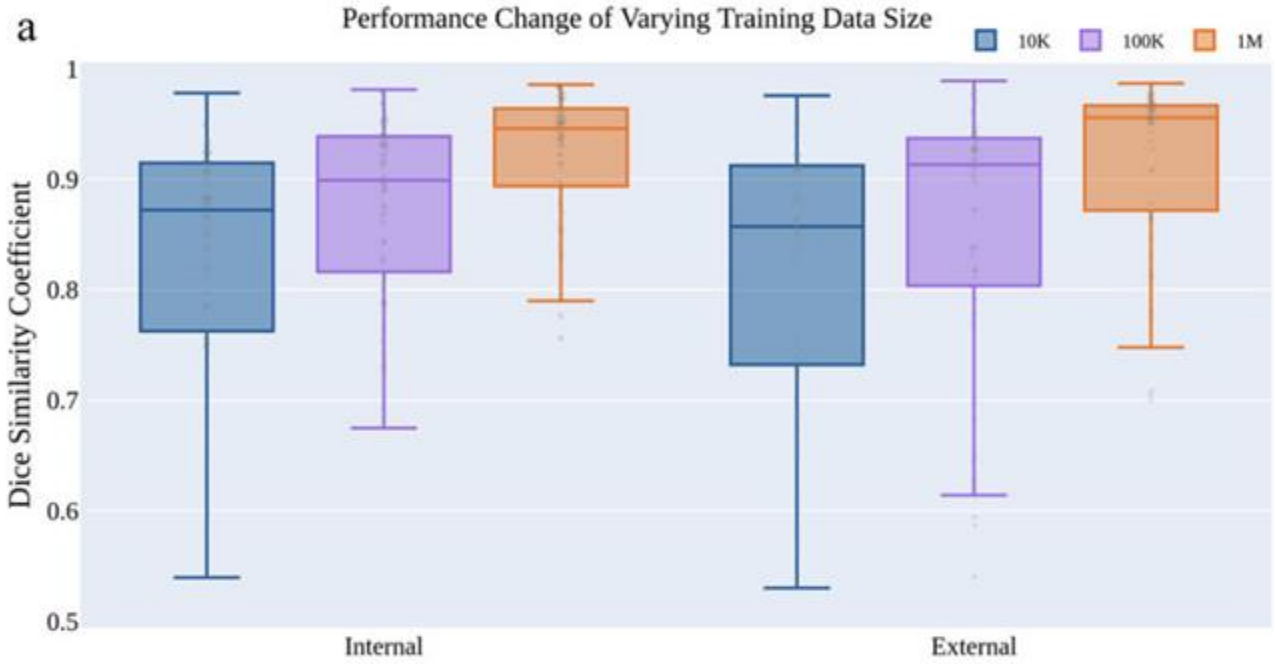
Ma et al. Segment anything in medical images. Nature Communications, 2024.

Performance comparison on external validation tasks



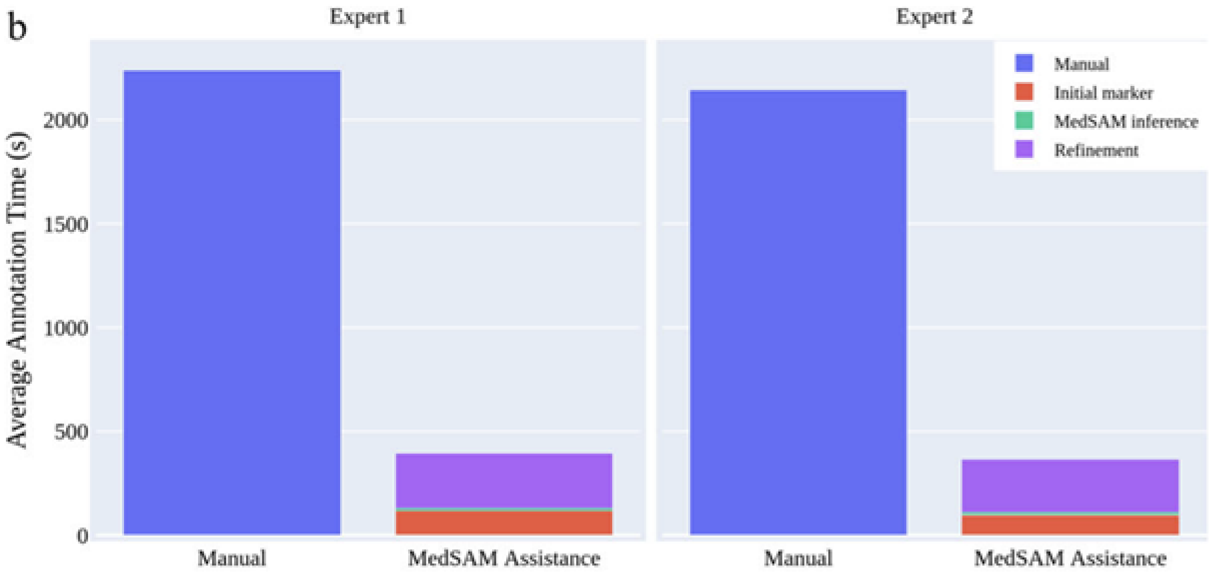
Ma et al. Segment anything in medical images. Nature Communications, 2024.

Effect of self-supervised training dataset size on downstream performance



Ma et al. Segment anything in medical images. Nature Communications, 2024.

Using MedSAM to reduce data annotation time



Ma et al. Segment anything in medical images. Nature Communications, 2024.

Next time

- Vision-language representation learners