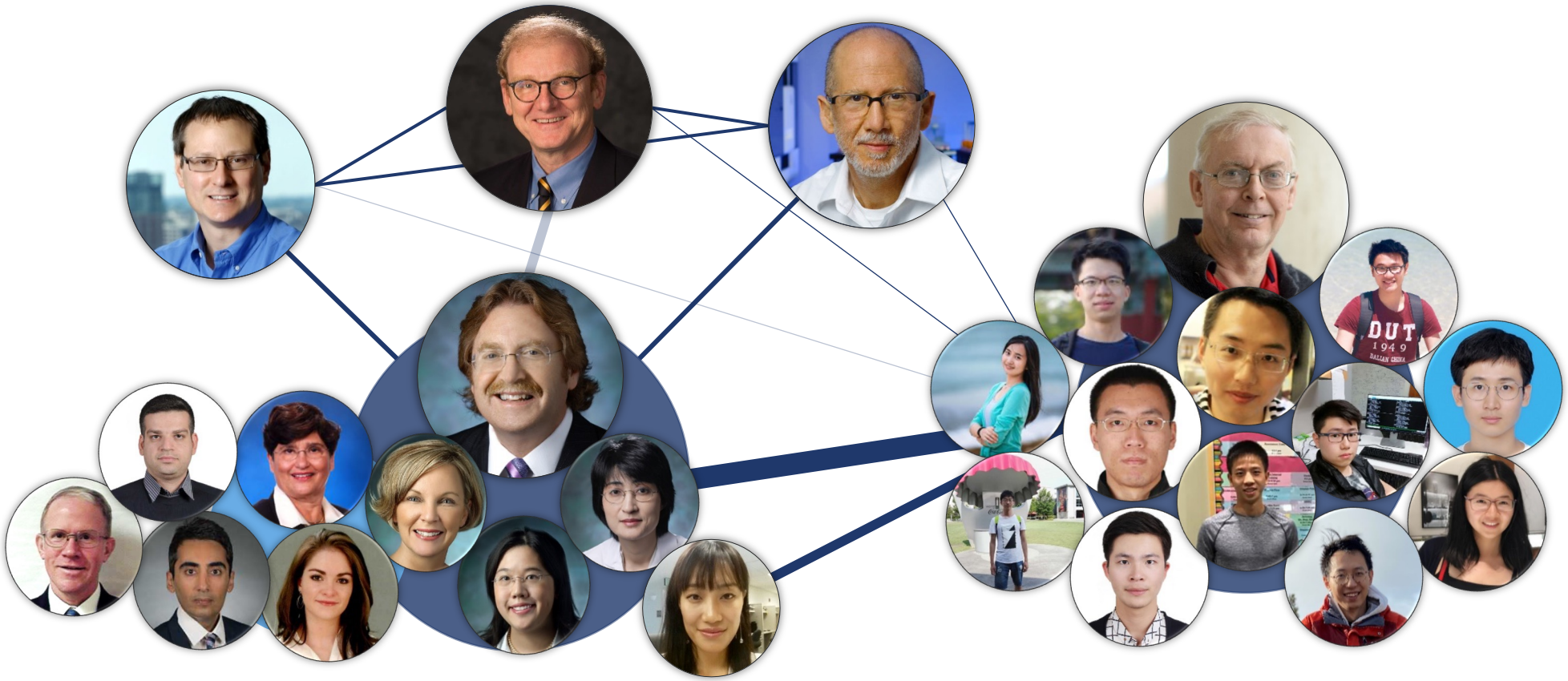
Medical Image Analysis: Scaling Annotations, Datasets, and Algorithms

Zongwei Zhou, PhD

Postdoc, Department of Computer Science
Johns Hopkins University, Baltimore, MD
P: 1-(480)738-2575 | E: zzhou82@jh.edu

The Hopkins Team



Medical School

Engineering School

Objective

Early Cancer Detection

Pancreatic Cancer (FELIX)

High-performance AI algorithms

+ Sensitivity = 97%, Specificity = 99%

+ Performance similar to radiologists

+ Generalizable to multiple hospitals

- 5,038 annotated CT volumes

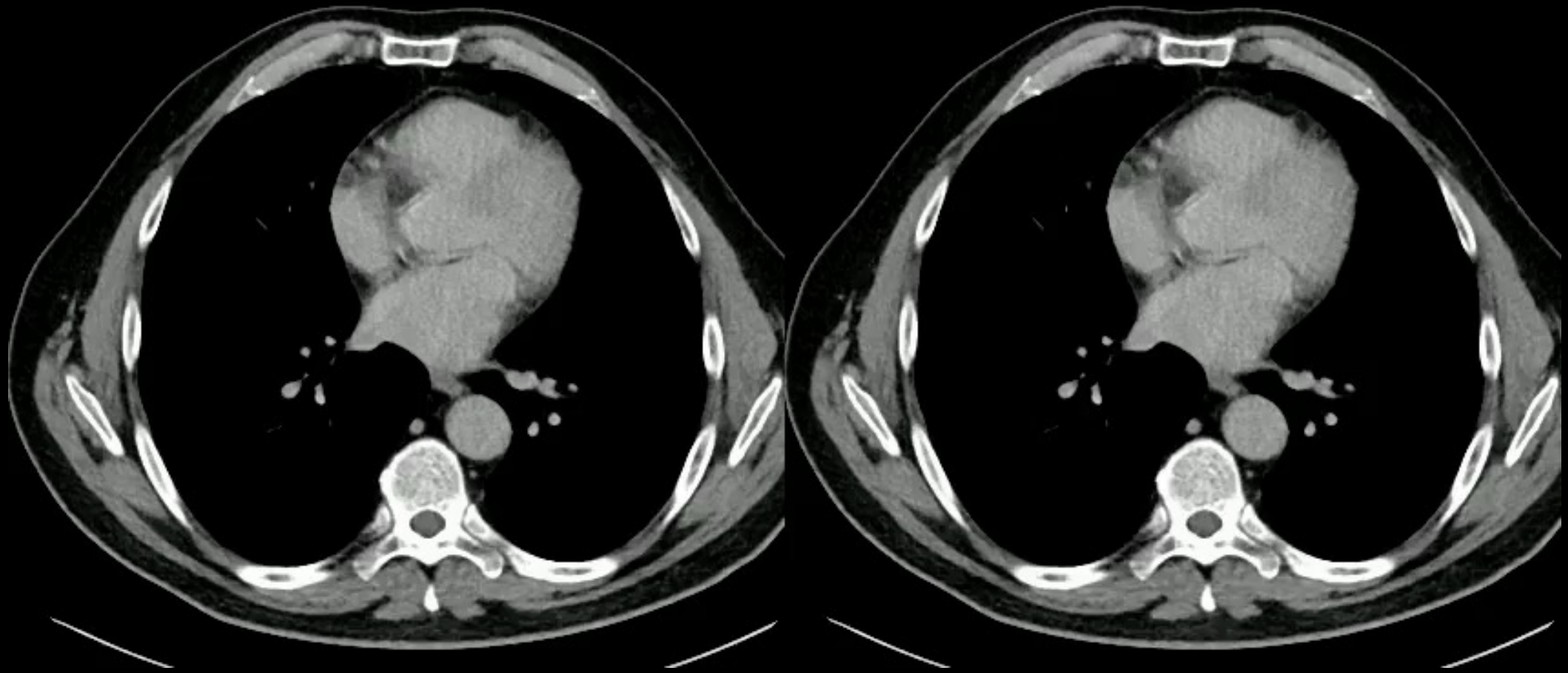
- 15 human-year to create

- Only for pancreatic tumors

Radiologists hate annotation, but computer scientists love it.

CT

AI predictions



Pancreas

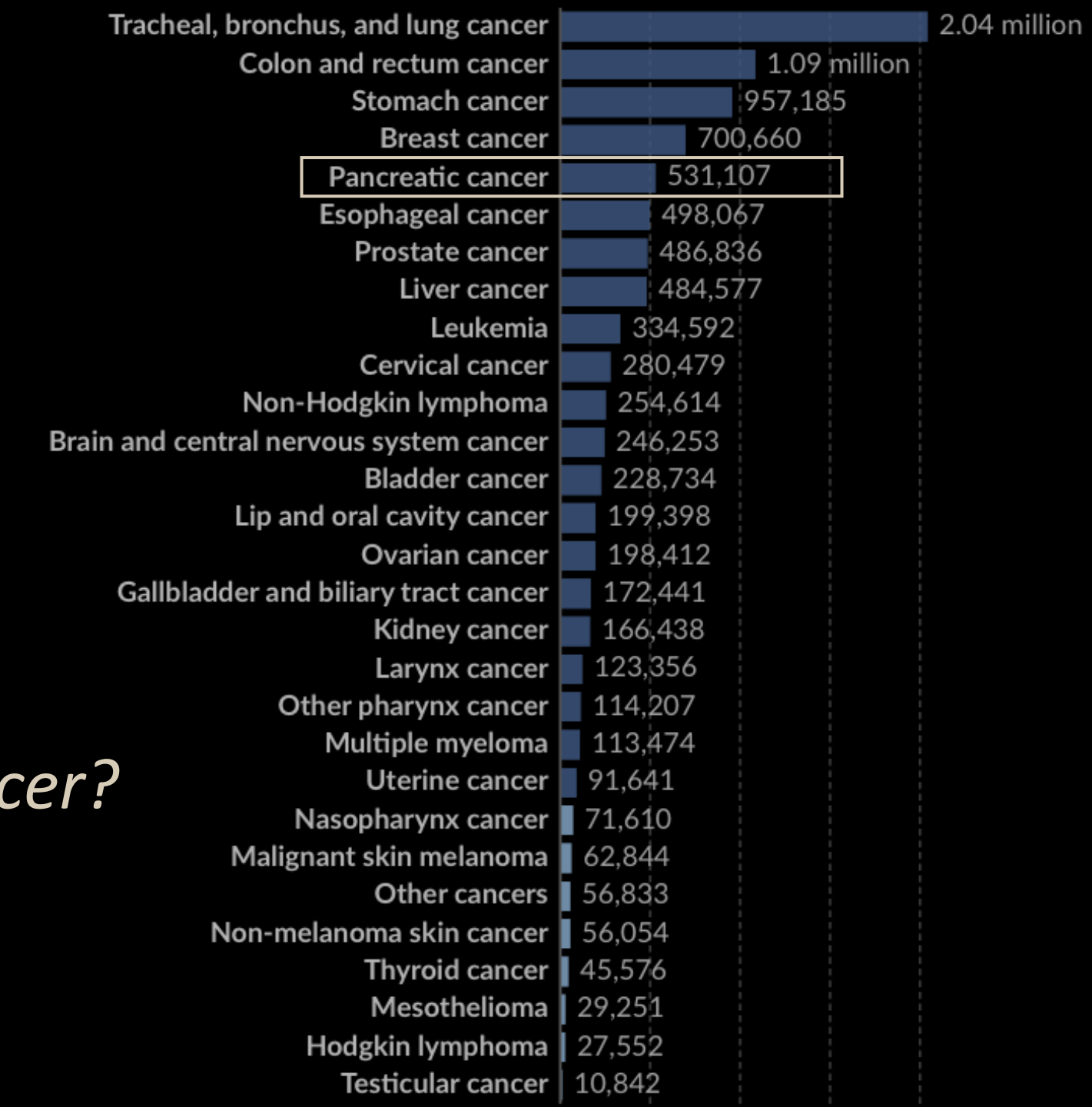
Pancreatic tumor

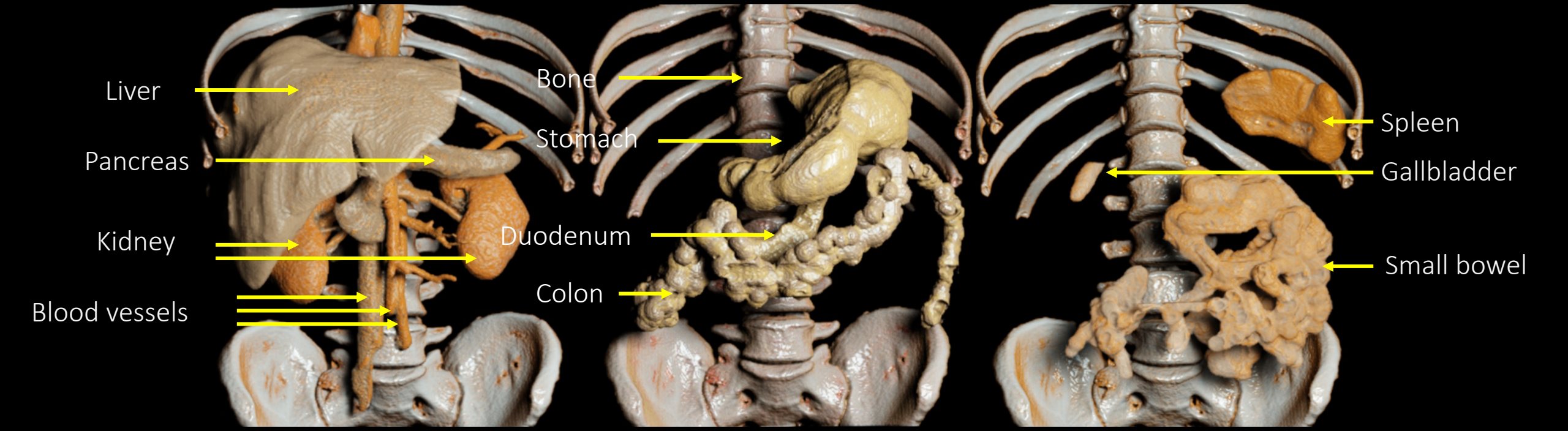
Pancreatic duct

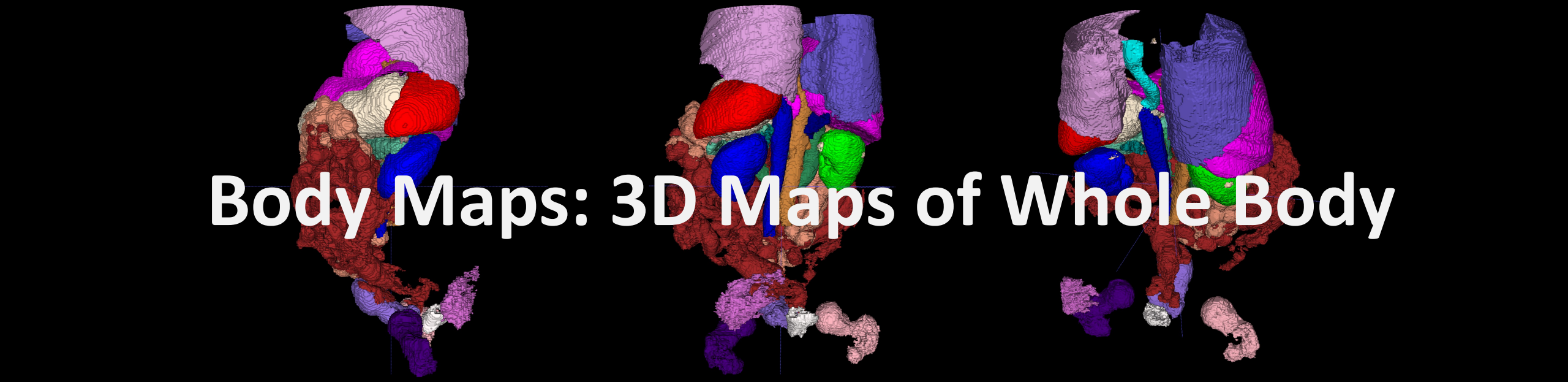
Objective

Early Cancer Detection

How do we deal with many other types of cancer?

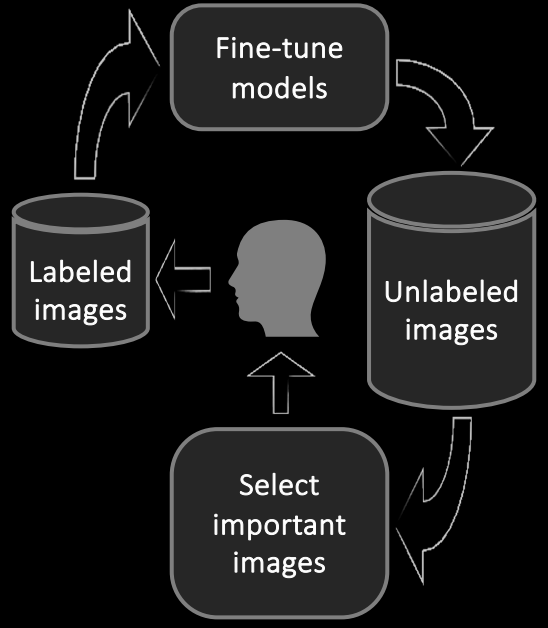






Body Maps: 3D Maps of Whole Body

Medical Image Analysis: Scaling Annotations, Datasets, and Algorithms

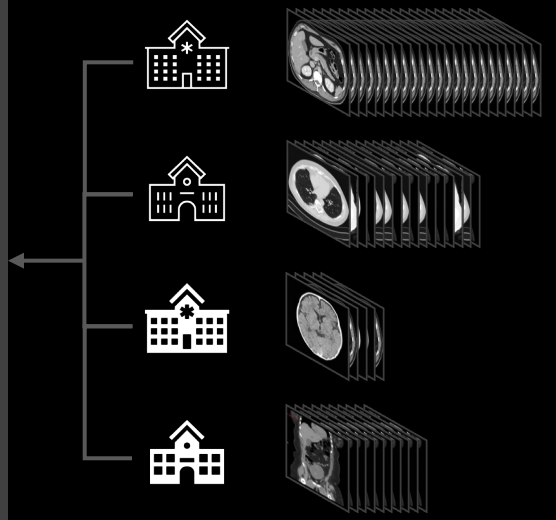


Interactive segmentation

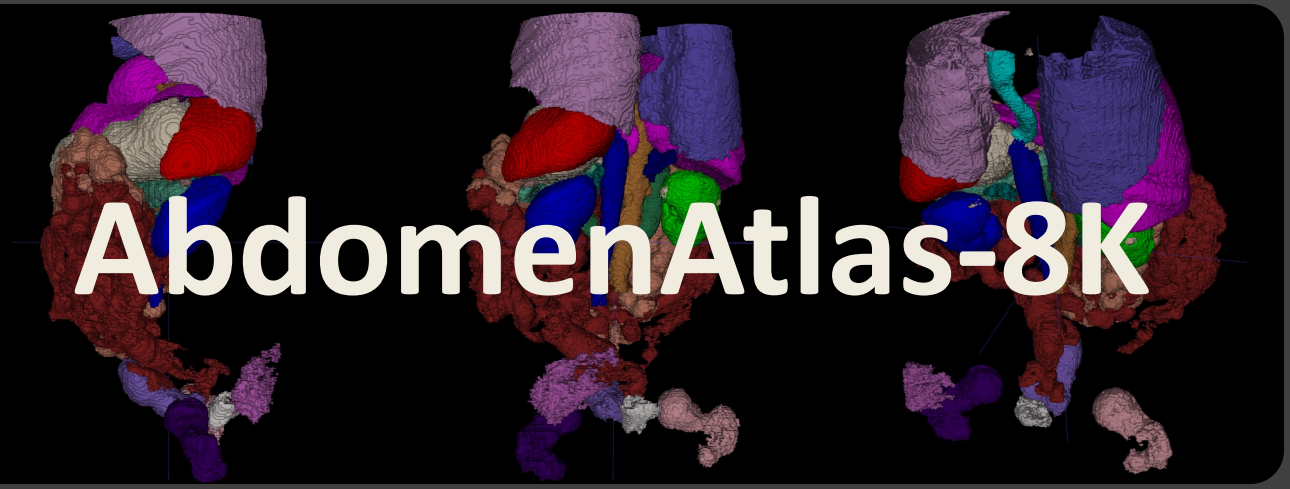
Annotated
25
organs

Annotated
7
cancers

Integrated
15
public datasets



Collected from
27 hospitals
worldwide



AbdomenAtlas-8K

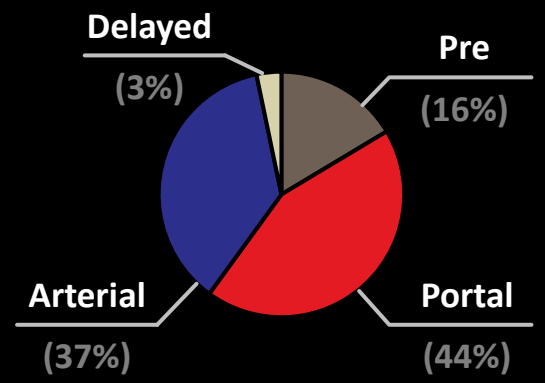
Up to
533x faster
than previous strategies

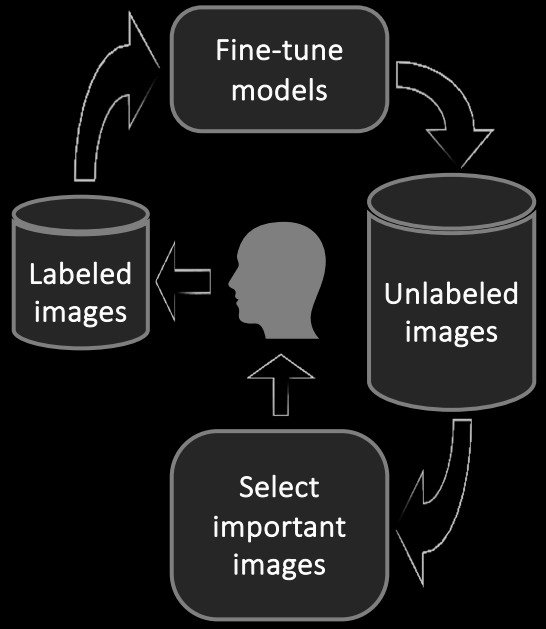
MONAI

Annotated
3.2M
images

Annotated
8,448
CT volumes

Created in
3 Weeks
by 1 annotator





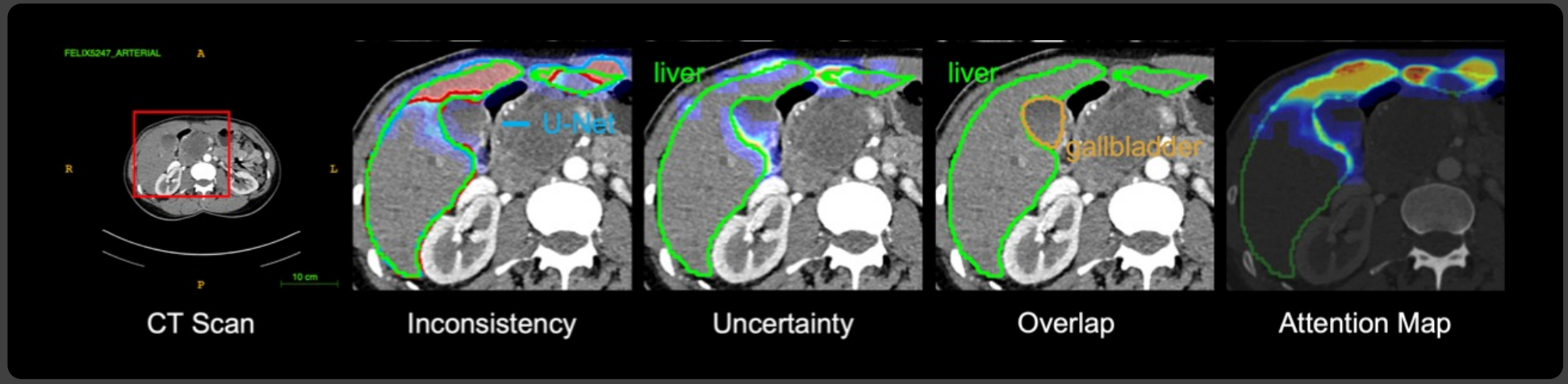
Interactive segmentation

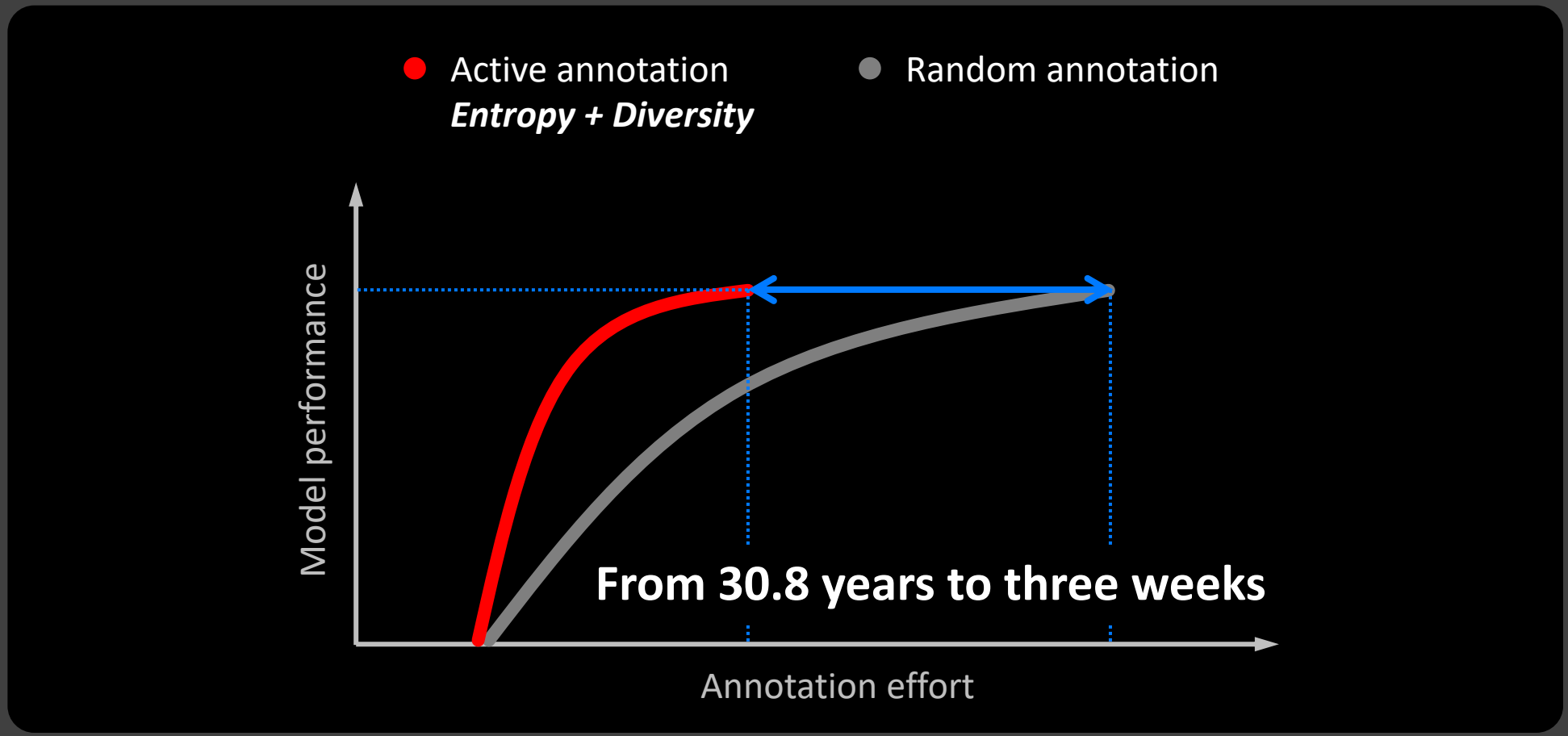
Up to

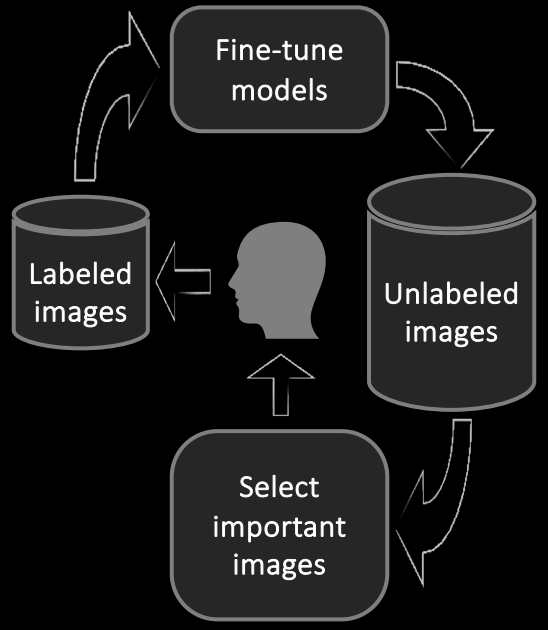
533x faster

than previous strategies

MONAI







Interactive segmentation

Up to

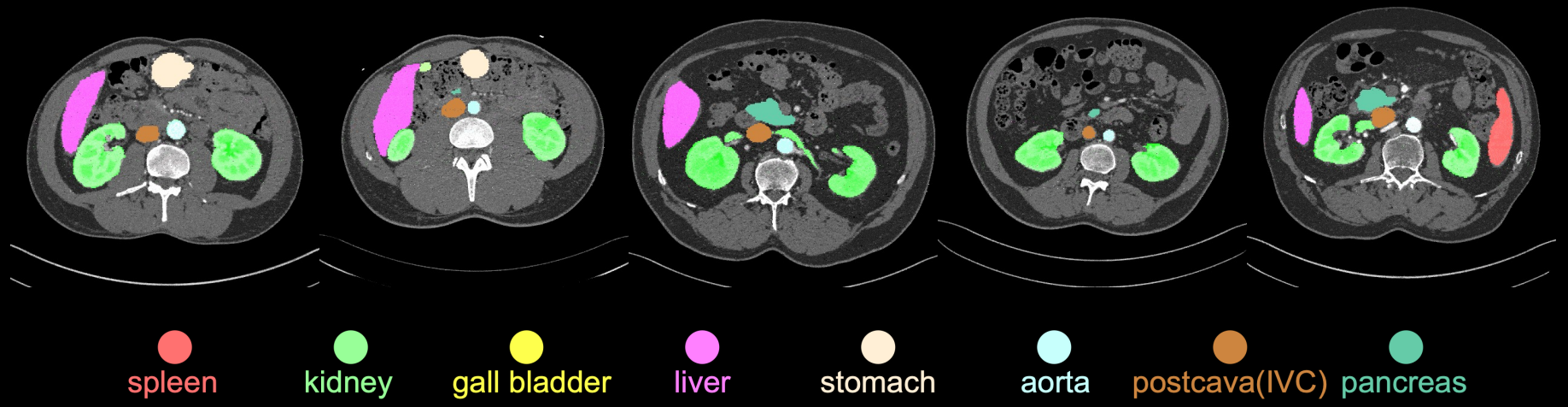
533x faster

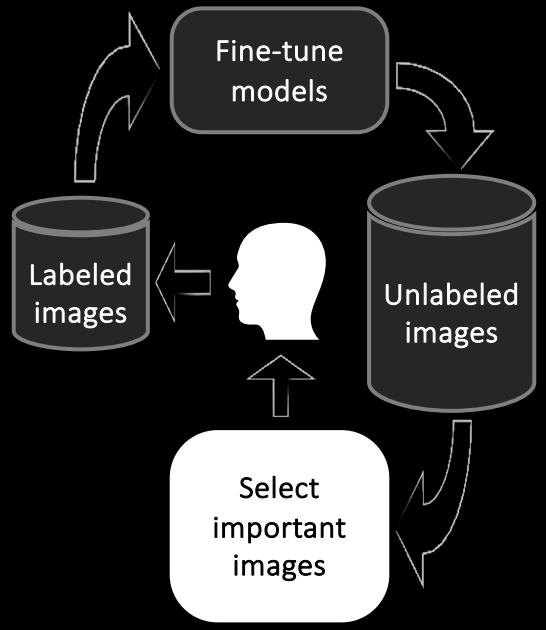
than previous strategies

MONAI

We will release **AbdomenAtlas-8K** of
8,448 CT volumes, totaling **3.2 million** CT slices

[Qu et al., NeurIPS 2023]





Interactive segmentation

Up to

533x faster

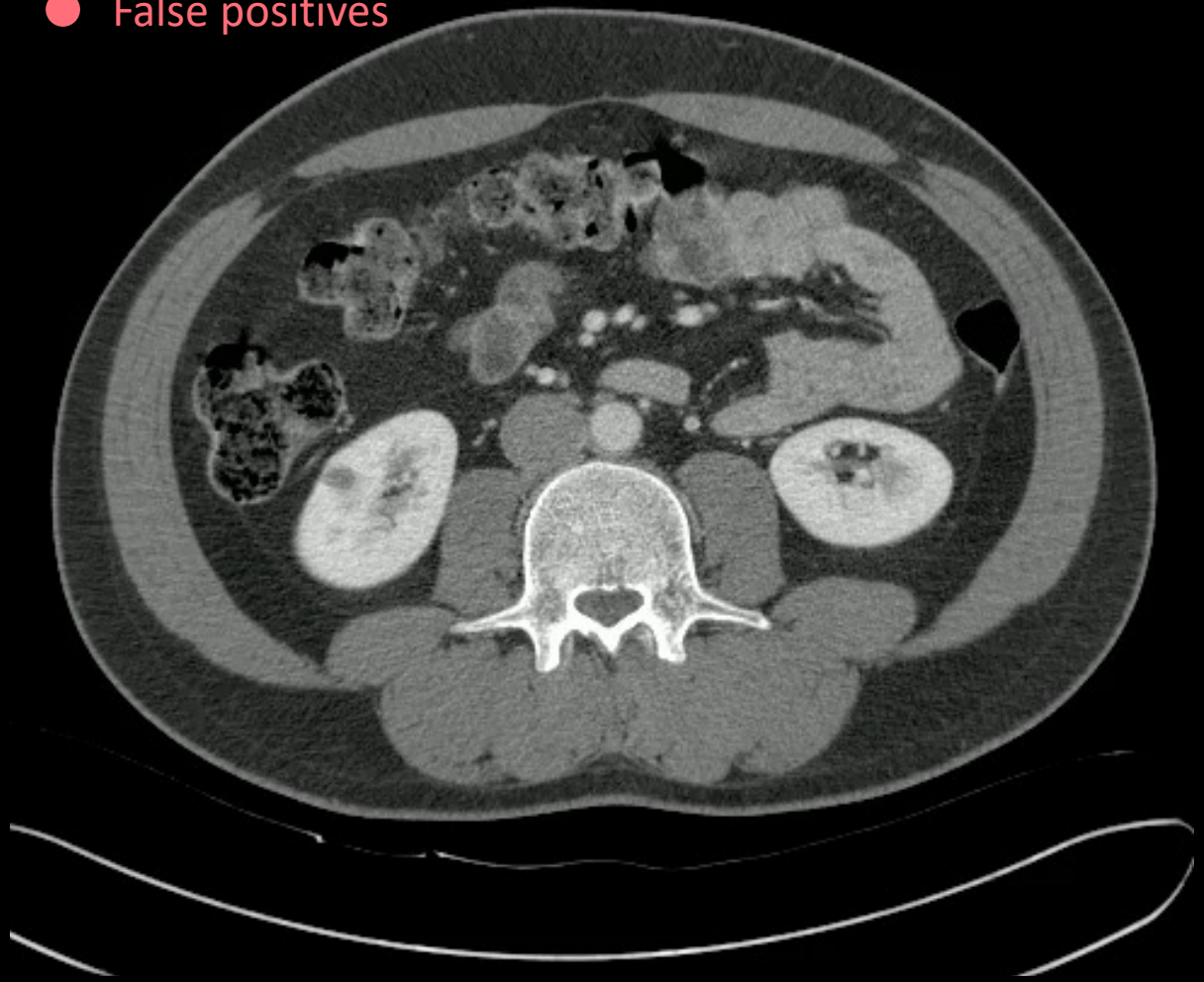
than previous strategies

MONAI

False positives ↓ true negatives ↑

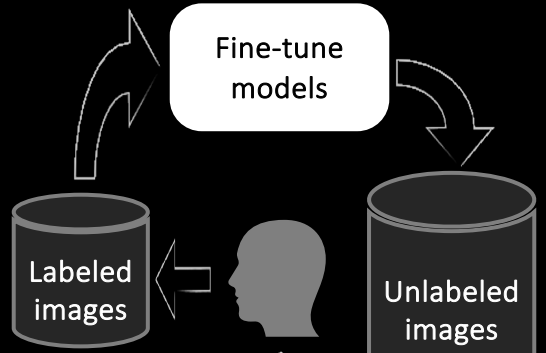
Pathology reports

● False positives



Findings:
The liver is normally positioned and normal in size and morphology with smooth border. Its internal structure and attenuation values are normal. No intra-hepatic biliary ductal dilatation.

Impression:
Normal liver



Interactive segmentation

Up to

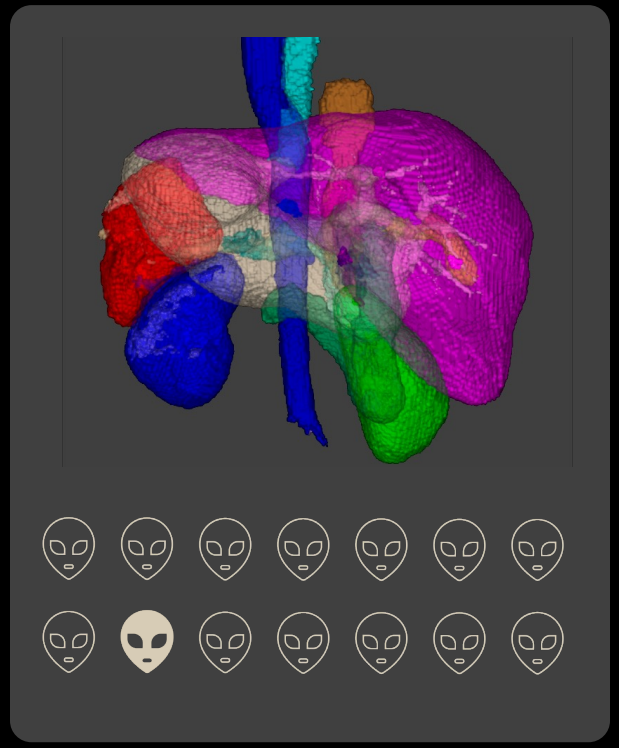
533x faster

than previous strategies

MONAI

False negatives ↓ true positives ↑

Manual annotations



Anomaly detection

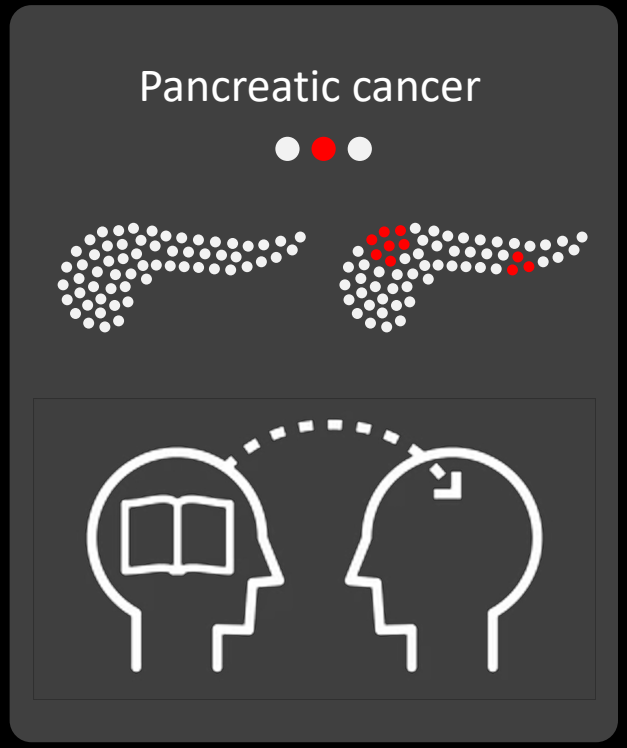
Pathology reports

[Xiang et al., CVPR 2023]

Weak annotation

Visual prompts

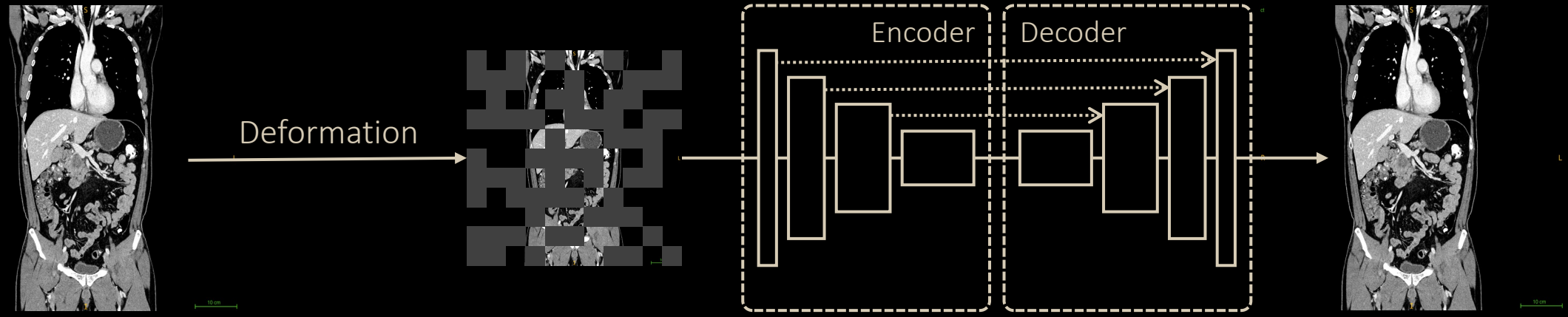
[Chou et al., MIR 2023]



Transfer learning

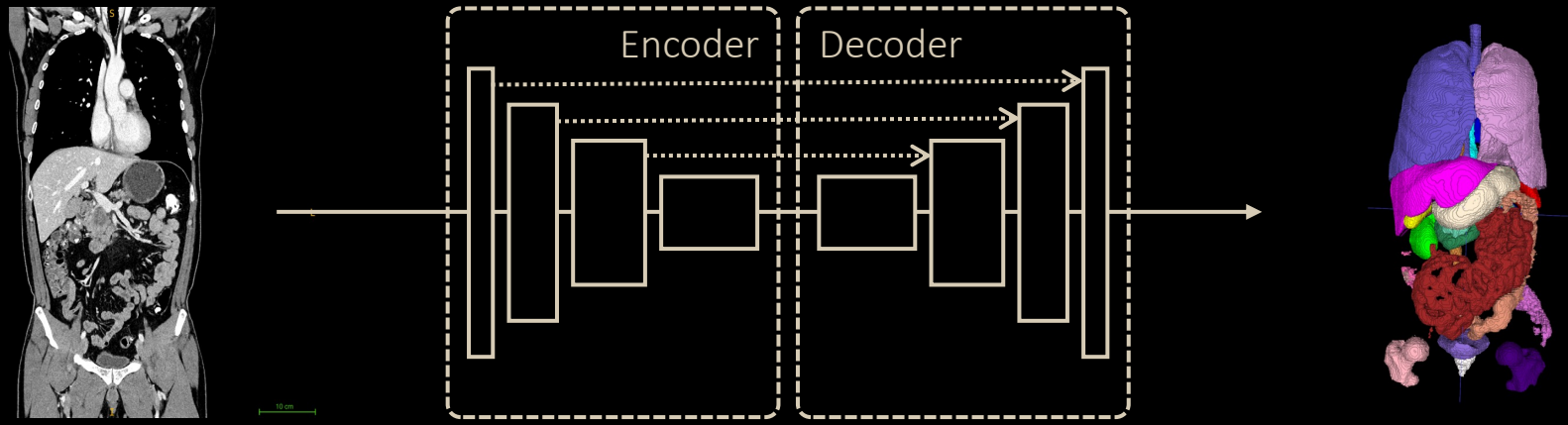
Per-voxel annotations

[Li et al., under review]



Self-supervised Pre-training

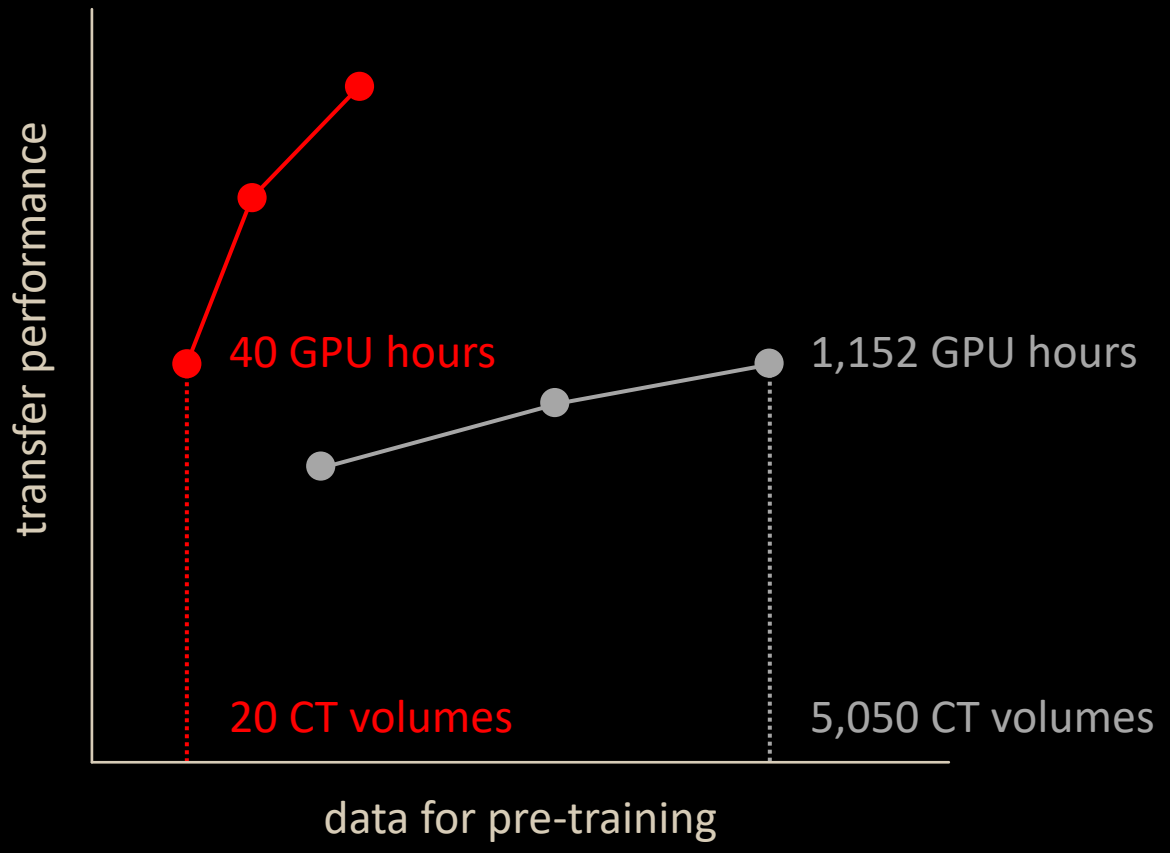
5,000 data + 1,152 GPU hours



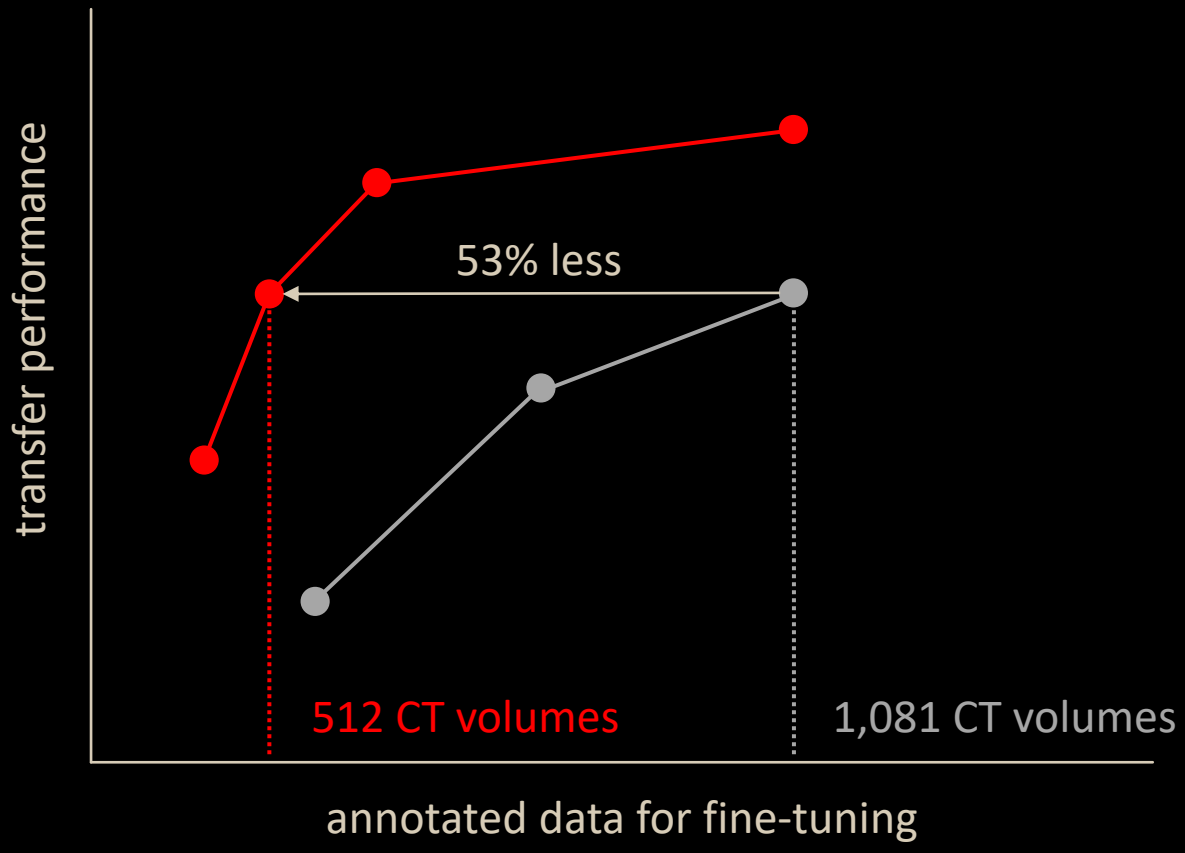
Supervised Pre-training

20 data + 20 annotation + 40 GPU hours

Supervised > Self-supervised
data & computation efficiency



Supervised > Self-supervised
annotation & learning efficiency

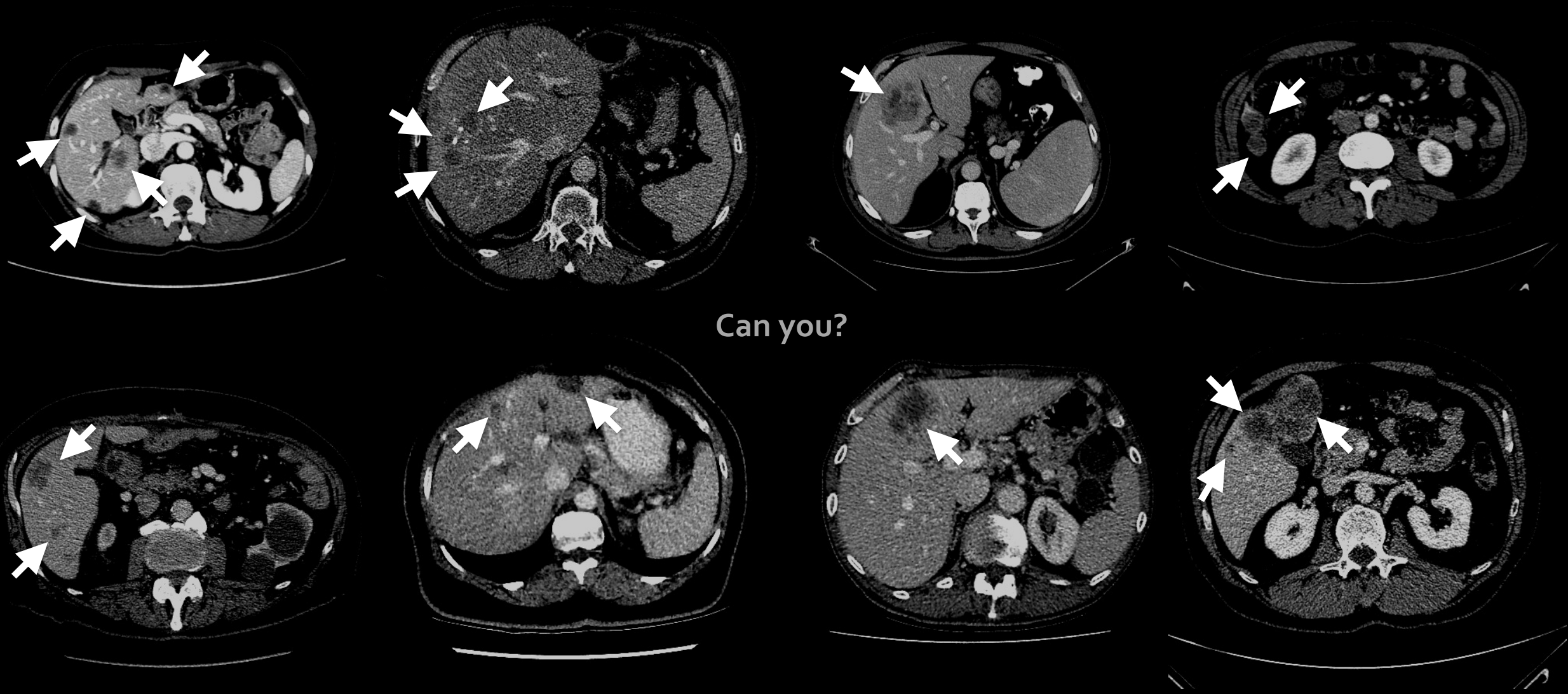


Medical Image Analysis: Scaling Annotations, Datasets, and Algorithms

- **We created AbdomenAtlas-8K for organs**
- **Now, scaling annotations for cancers is challenging**
  - Pathology reports
  - Manual annotations
  - *Collaborations (academia, industry, & hospital)*

Medical Image Analysis: Scaling Annotations, Datasets, and Algorithms

Medical professionals cannot tell which are real and which are synthetic tumors



Can you?

Training AI on synthetic tumors performs as well as training it on real tumors

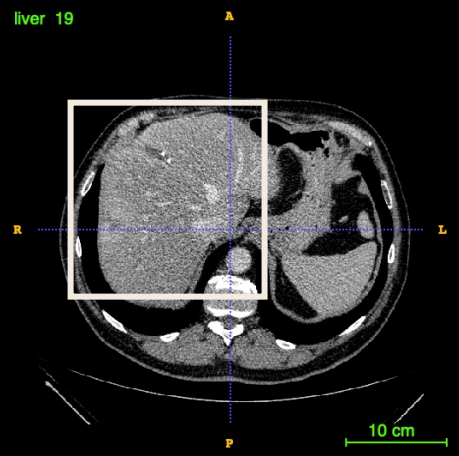
CT

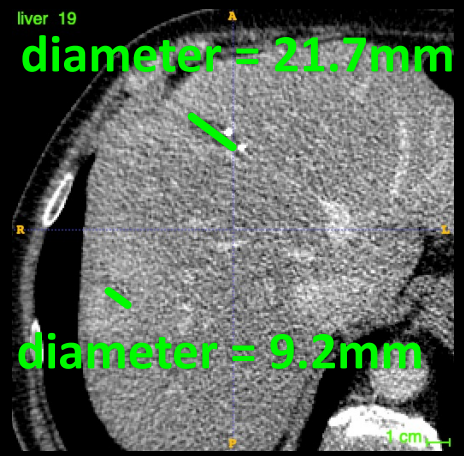
AI prediction
trained on real tumors
with per-voxel annotation
DSC = 58% [52% - 63%]

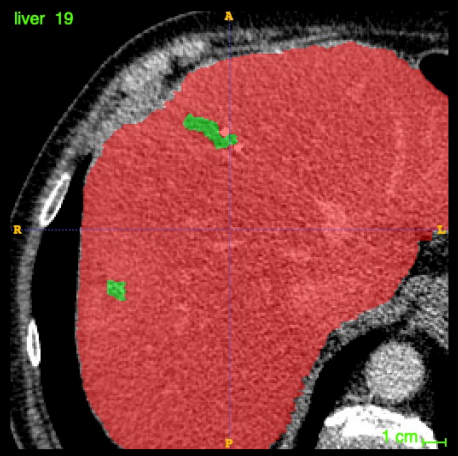
AI prediction
trained on synthetic tumors
with no annotation
DSC = 60% [55% - 65%]

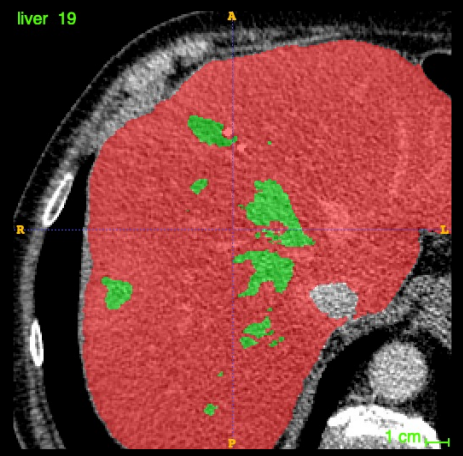
- Liver
- Liver tumor

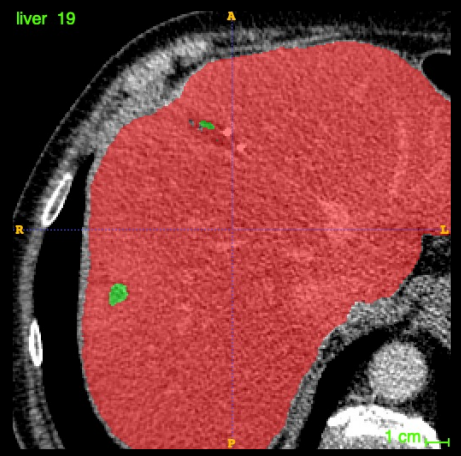
[Qualitative] Generating enormous small tumors for training AI models

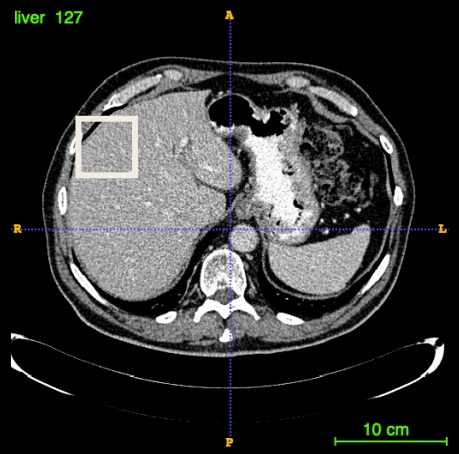


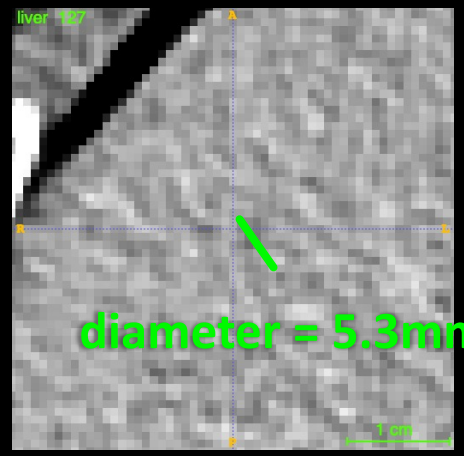


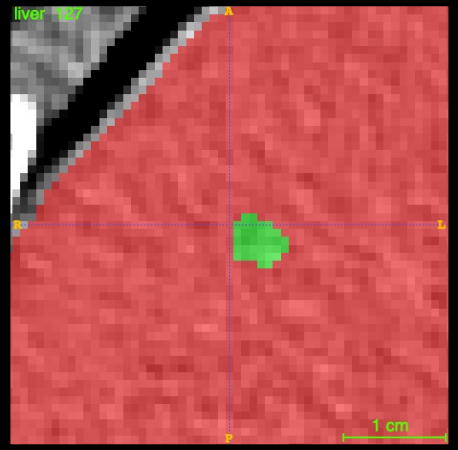


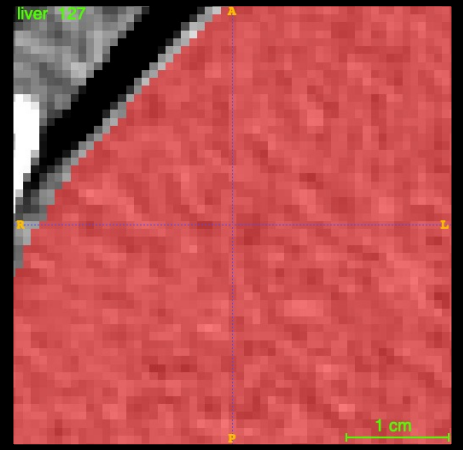


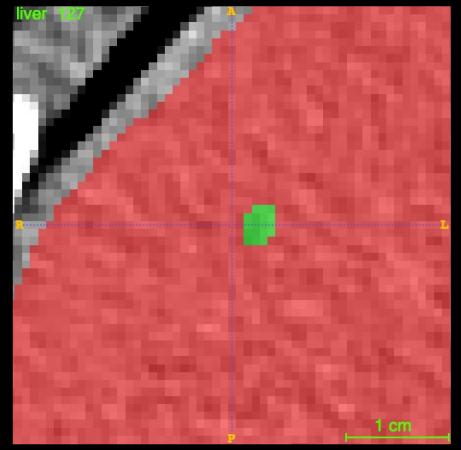






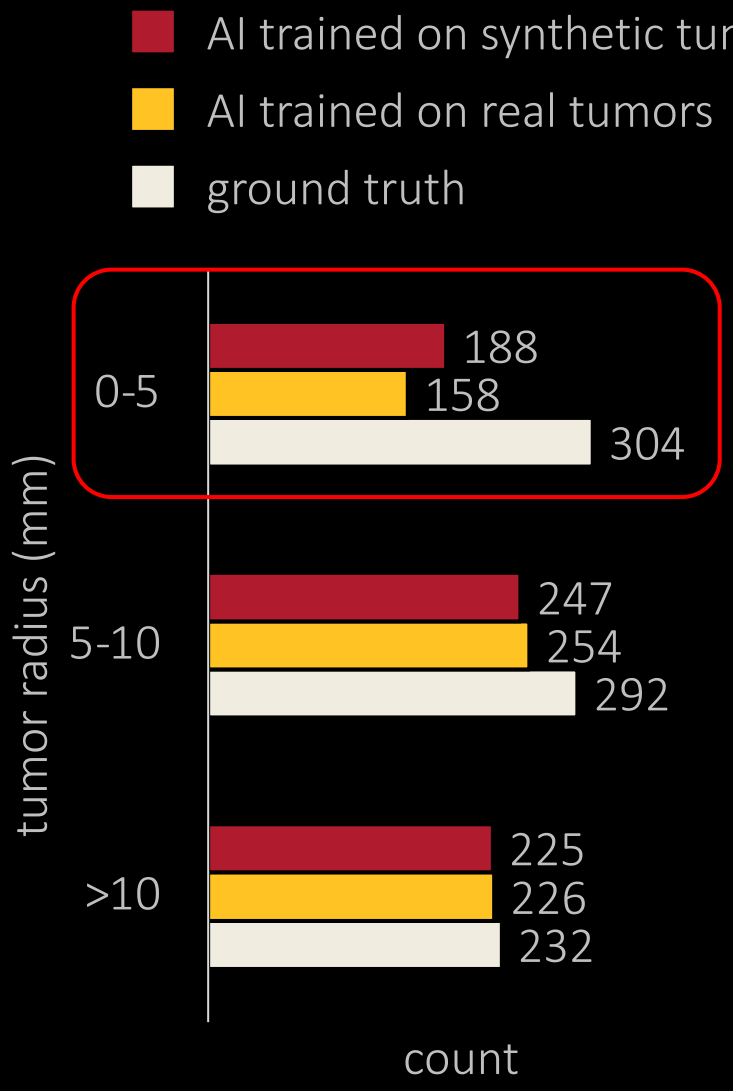






CT scan | zoom-in | ground truth | AI trained on real tumors | AI trained on synthetic tumors

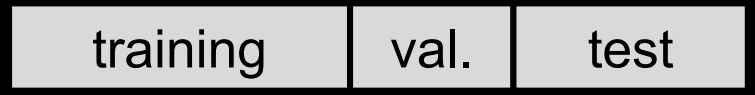
[Quantitative] Generating enormous small tumors for training AI models



Observation: Compared with real tumors,
AI trained on synthetic tumors improves Sensitivity from 52% to 62% for detecting small tumors (0-5mm).

- Needed for early detection
  - Early signs of cancer can be subtle
  - 1/2 of liver cancer are missed by radiologists
- Needed for AI development
  - CT scans with early cancer are limited
  - Annotations for early cancer are hard
- Needed for medical education
  - Junior radiologists have an Accuracy of 20%
  - Senior radiologists have an Accuracy of 78%

[Application] Not only for training, but also for validation



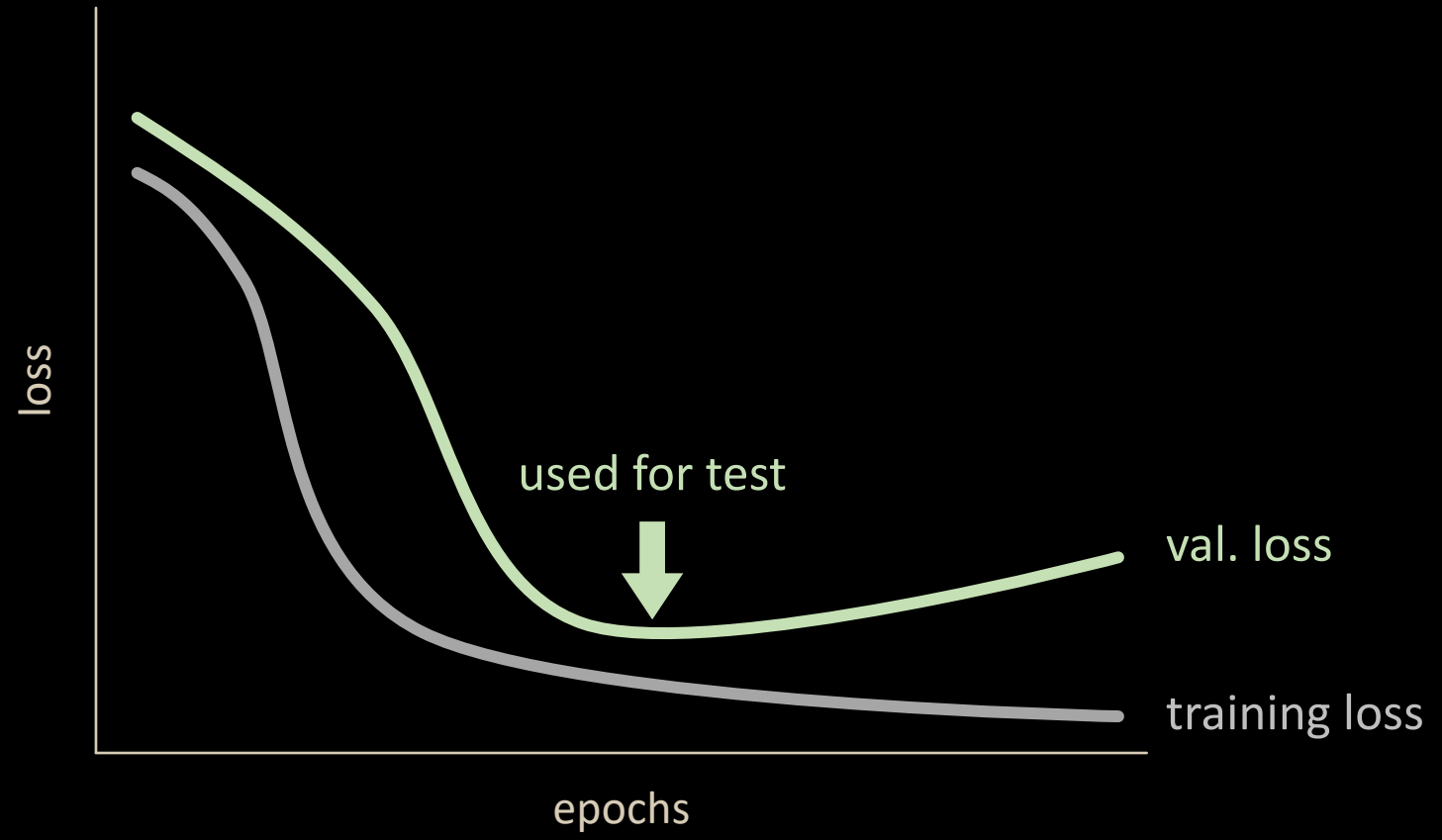
Problem

The total number of data is fixed

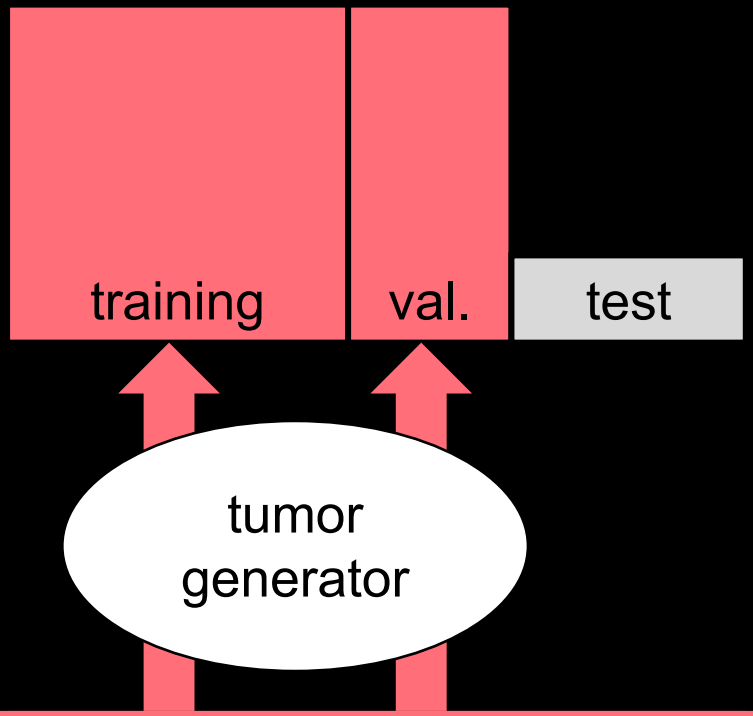
More for training, less for validation

Validation may not be reliable

Overfitting / underfitting!



[Application] Not only for training, but also for validation



Over **80 million** CT scans are performed in the United States each year

a stream of normal data

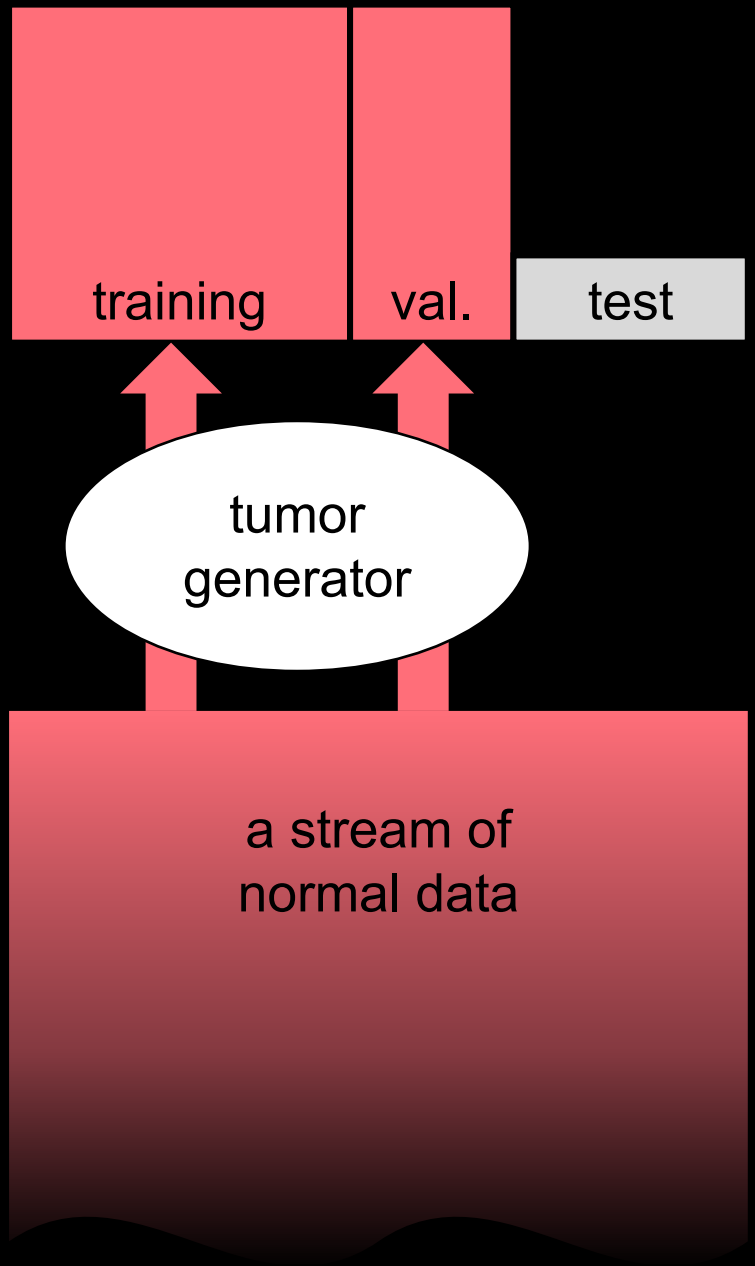
Hospital 1

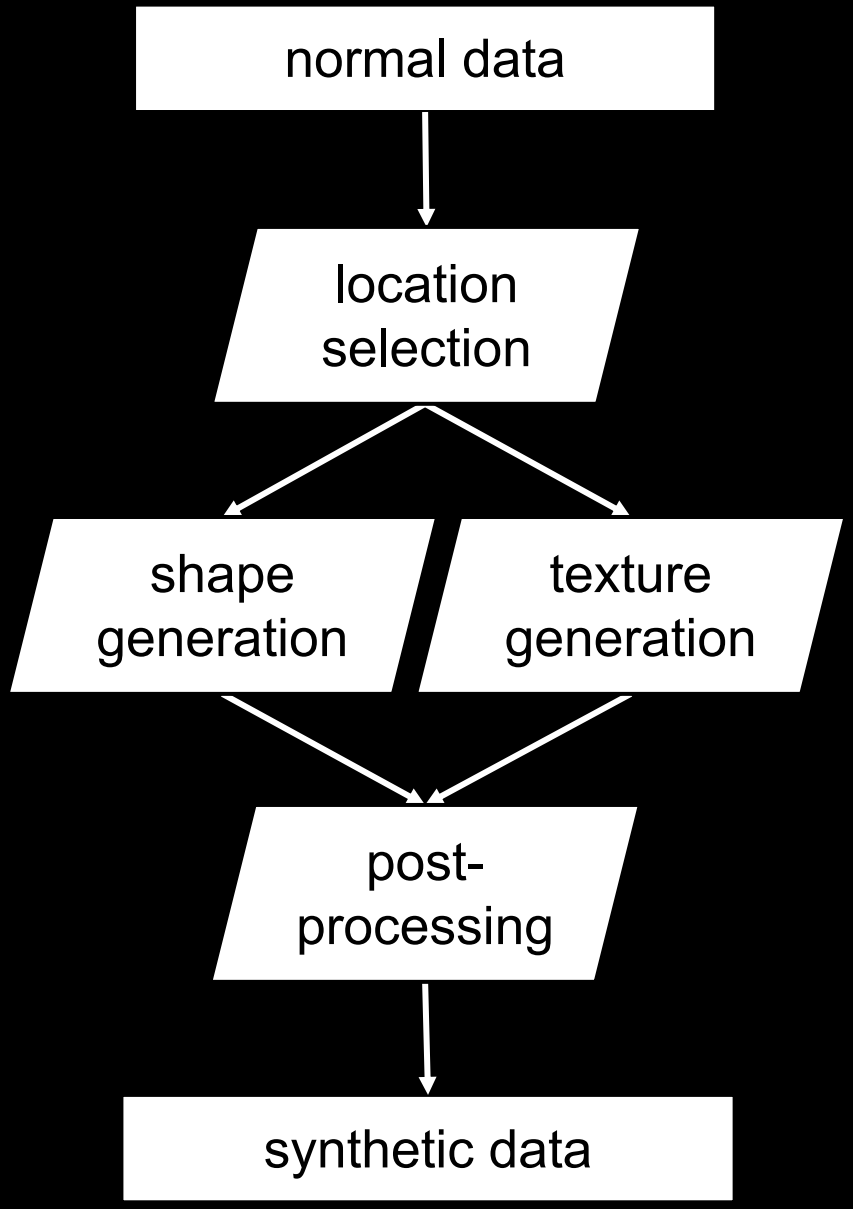
Hospital 2

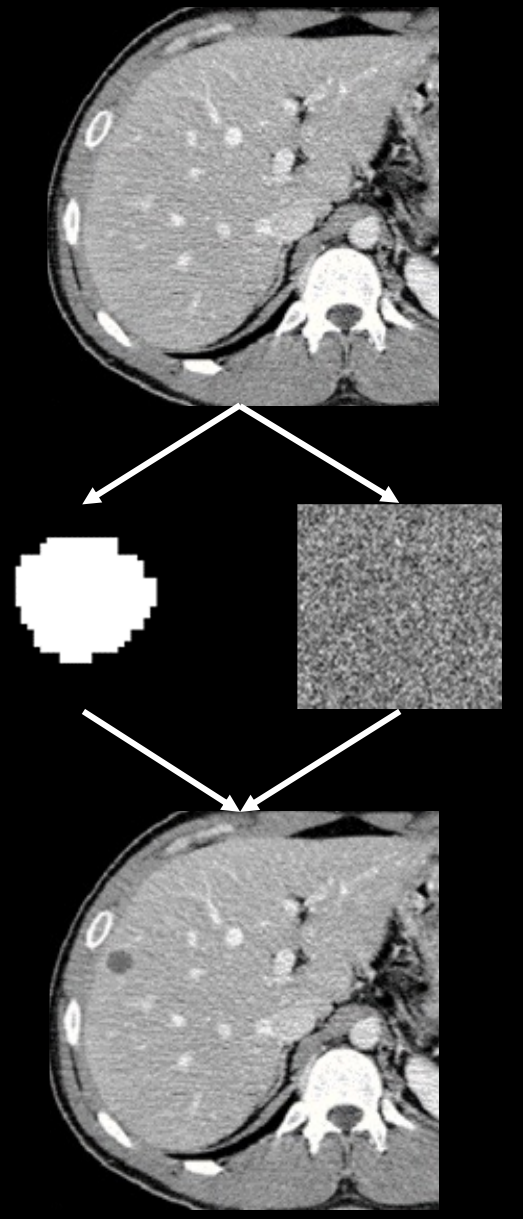
Hospital 3

Hospital 4

[Application] Not only for training, but also for validation

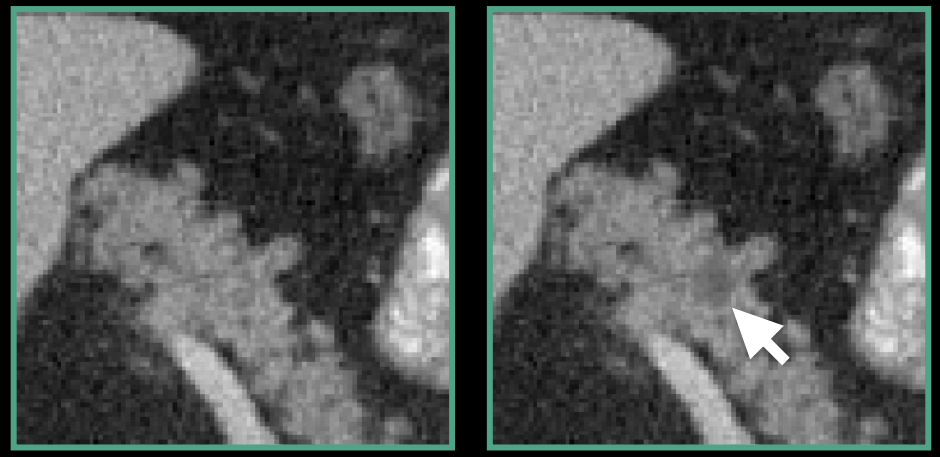




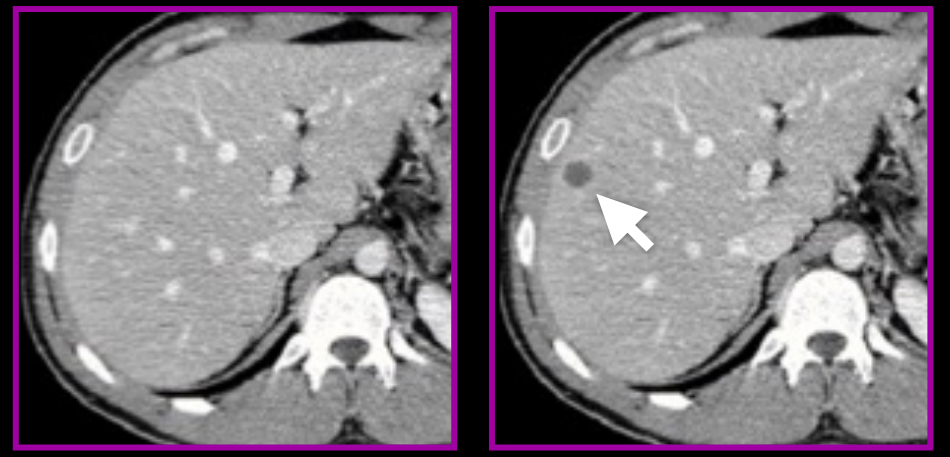


Towards generalizable tumor synthesis

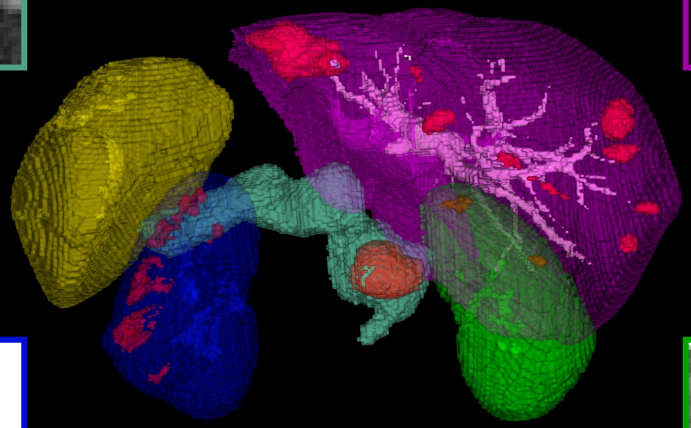
Early tumors



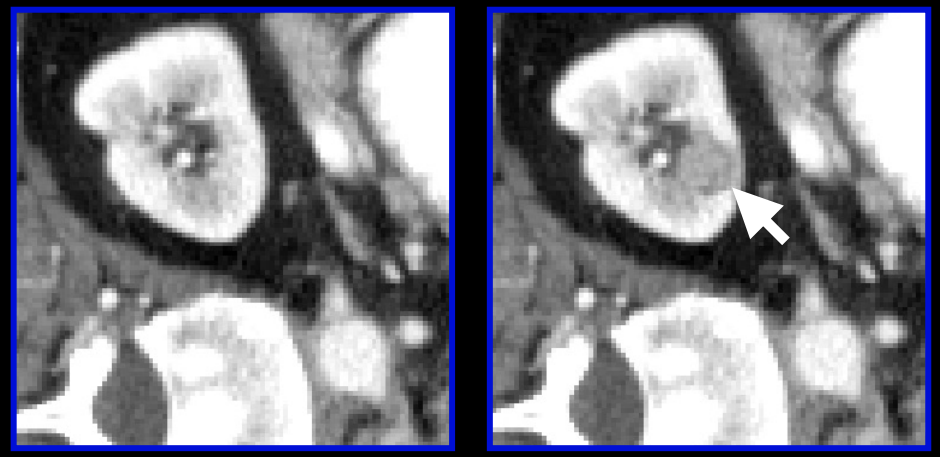
pancreas



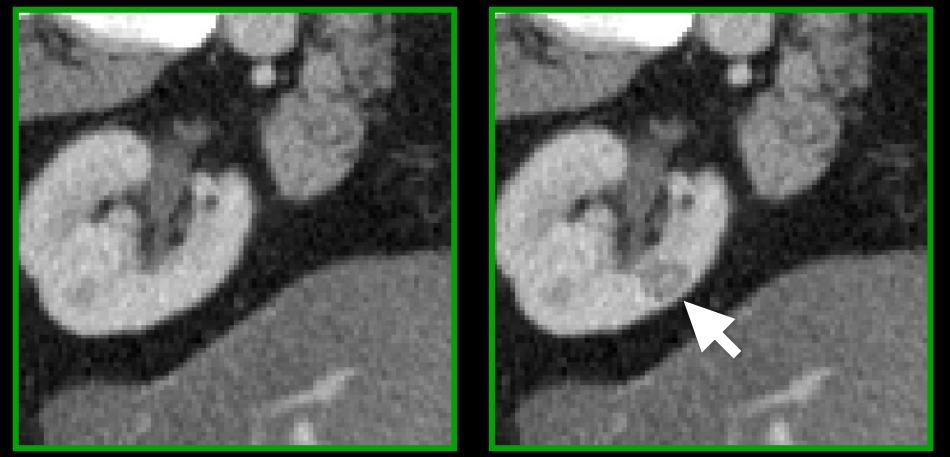
liver



left kidney



right kidney



Medical Image Analysis: Scaling Annotations, Datasets, and Algorithms

- **Tumor synthesis for the liver and pancreas**
- **Now, generalizable tumor synthesis is challenging**
  - Training, validation, and *evaluation (?)*
  - Similarity in early tumors

Medical Image Analysis: Scaling Annotations, Datasets, and Algorithms

“While GPT-4V demonstrates proficiency in distinguishing between medical image modalities and anatomy, it faces significant challenges in **disease diagnosis** and **generating comprehensive reports**.”

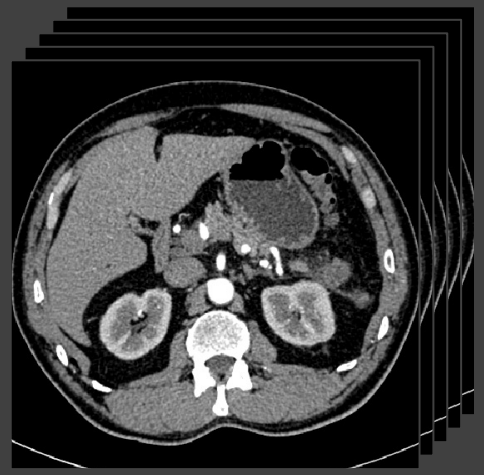
GPT-4V

Can GPT-4V(ision) Serve Medical Applications? Case Studies on GPT-4V for Multimodal Medical Diagnosis

Chaoyi Wu[1,2*], Jiayu Lei[2,3*], Qiaoyu Zheng[1,2*], Weike Zhao[1,2*], Weixiong Lin[1,2*], Xiaoman Zhang[1,2*], Xiao Zhou[2*], Ziheng Zhao[1,2*], Ya Zhang[1,2], Yanfeng Wang[1,2] and Weidi Xie[1,2]

[1]Shanghai Jiao Tong University [2]Shanghai AI Laboratory

[3]University of Science and Technology of China



Is there any abnormality (e.g., tumor) in this abdominal CT scan?

GPT-4V

I'm not a radiologist, and this platform isn't equipped to provide medical diagnoses. It's important to consult with a medical professional for a reliable interpretation of any medical image. If you have concerns, please reach out to a healthcare provider.

Can GPT-4V(ision) Serve Medical Applications?

Case Studies on GPT-4V for Multimodal Medical Diagnosis

Chaoyi Wu[1,2*], Jiayu Lei[2,3*], Qiaoyu Zheng[1,2*], Weike Zhao[1,2*], Weixiong Lin[1,2*], Xiaoman Zhang[1,2*], Xiao Zhou[2*], Ziheng Zhao[1,2*], Ya Zhang[1,2], Yanfeng Wang[1,2] and Weidi Xie[1,2]

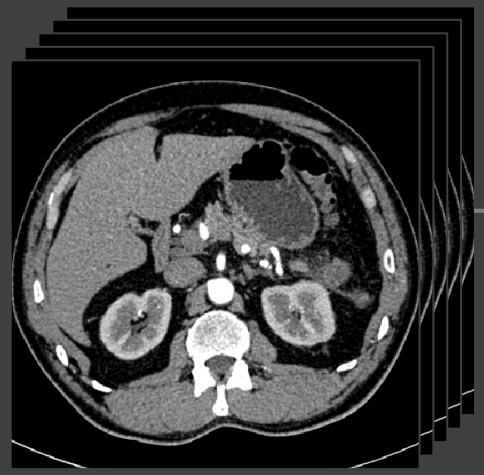
[1]Shanghai Jiao Tong University [2]Shanghai AI Laboratory

[3]University of Science and Technology of China

featured in

ChimeraX

at UCSF



Vision Encoder

Universal Model

Segment the liver.

Text Encoder

featured in

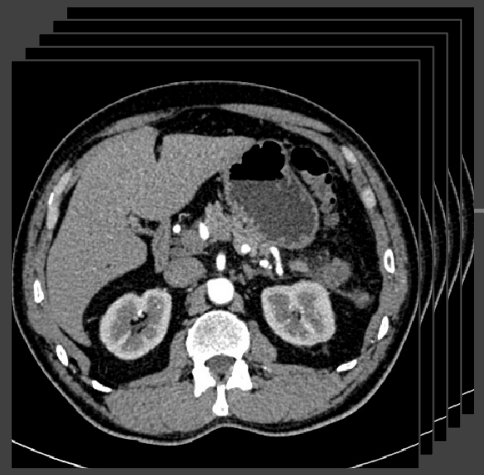
MONAI

at NVIDIA

featured in

ChimeraX

at UCSF

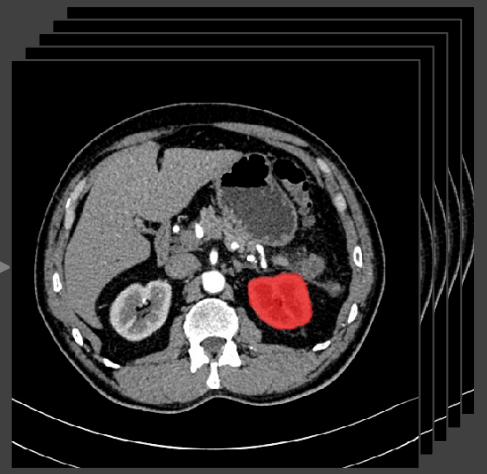


Vision Encoder

Universal Model

Text Encoder

Segment the left kidney.



featured in

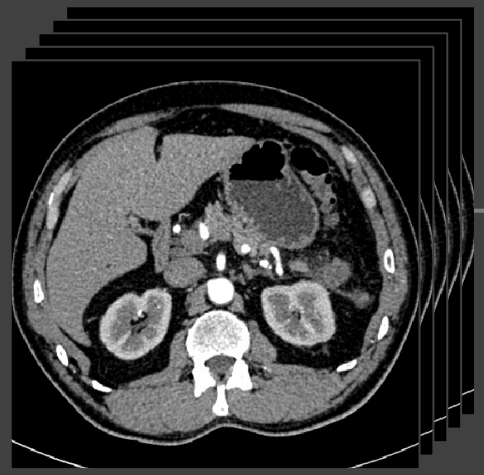
MONAI

at NVIDIA

featured in

ChimeraX

at UCSF

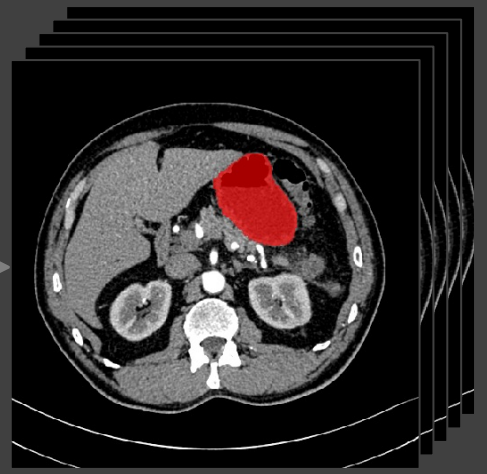


Vision Encoder

Universal Model

Text Encoder

Segment the stomach.



featured in

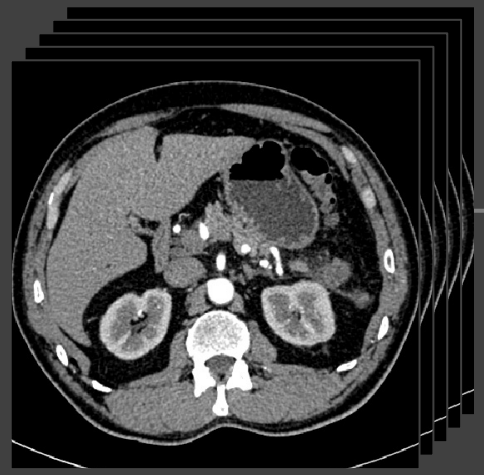
MONAI

at NVIDIA

featured in

ChimeraX

at UCSF

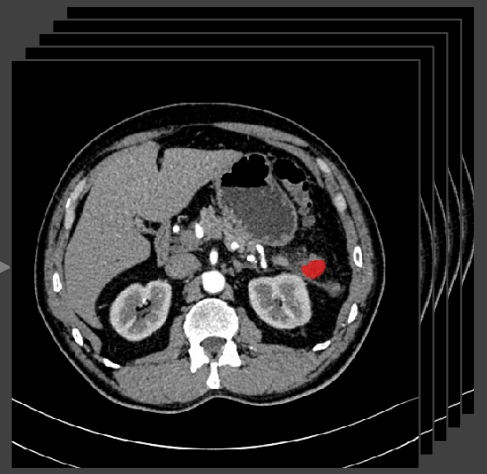


Vision Encoder

Universal Model

Text Encoder

Segment pancreatic tumors, if any.



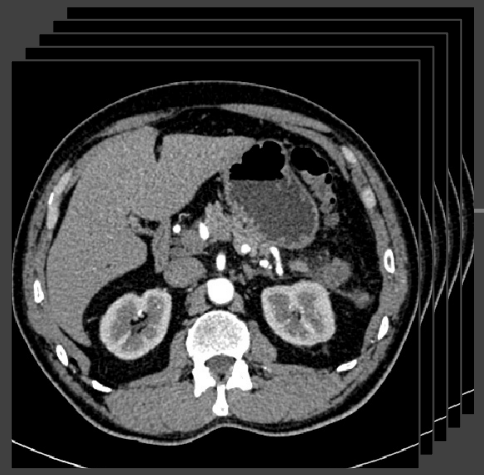
featured in

MONAI

at NVIDIA

featured in
ChimeraX
at UCSF

featured in
MONAI
at NVIDIA



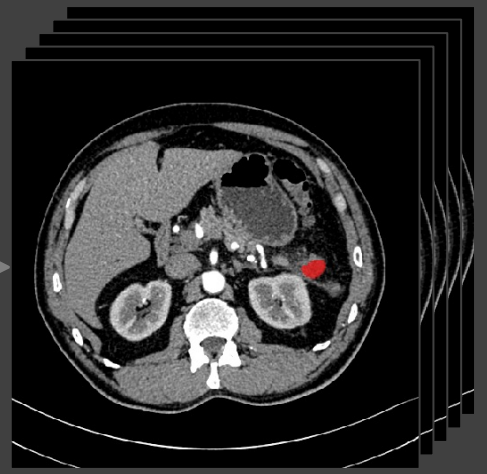
Vision Encoder

Universal Model

BTCV → LiTS → KiTS → AMOS → MSD

Segment pancreatic tumors, if any.

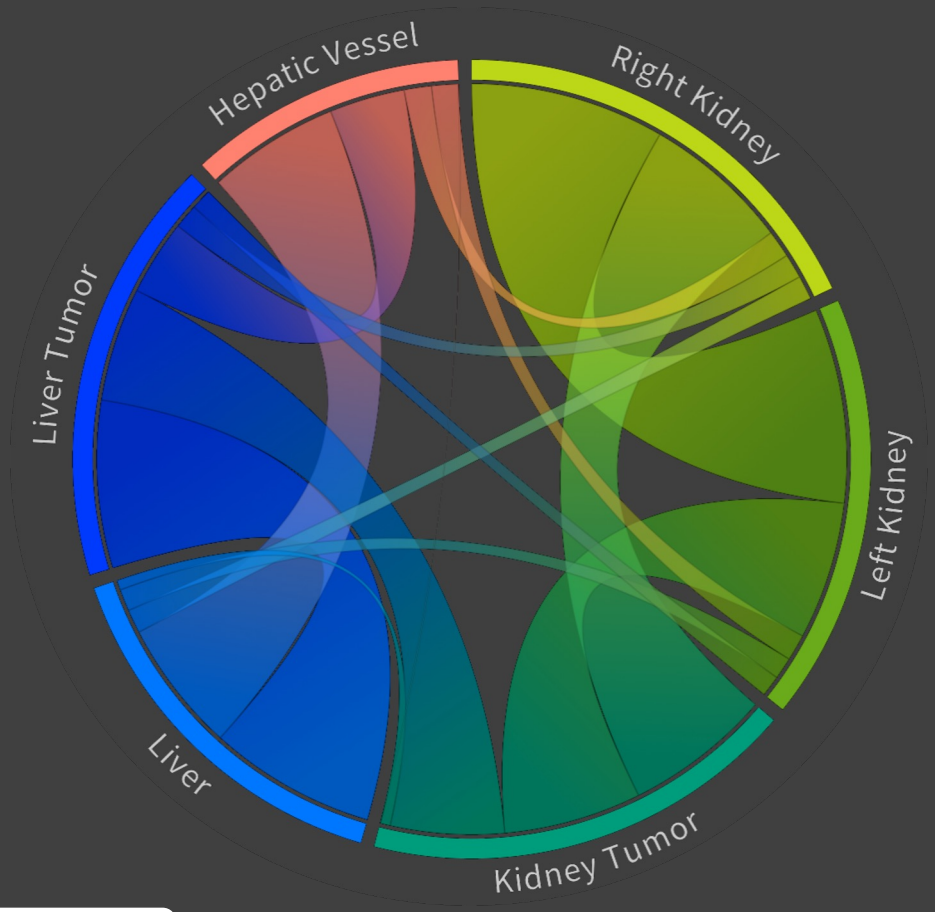
Text Encoder



One-hot embedding

1. *No semantic meaning*
2. *Not extendable to new classes*

liver:	[1,0,0,0,0,0]
liver tumor:	[0,1,0,0,0,0]
left kidney:	[0,0,1,0,0,0]
right kidney:	[0,0,0,1,0,0]
kidney tumor:	[0,0,0,0,1,0]
hepatic vessel:	[0,0,0,0,0,1]



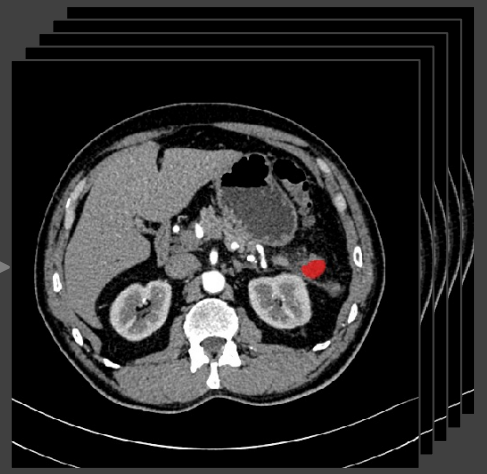
Language embedding

1. *Hierarchical relationship*
2. *Flexible to new classes*

e.g.,
Contrastive Language-Image Pre-training (CLIP)

Segment pancreatic tumors, if any.

Text Encoder



Info | Teams | Submit | Leaderboard | Statistics

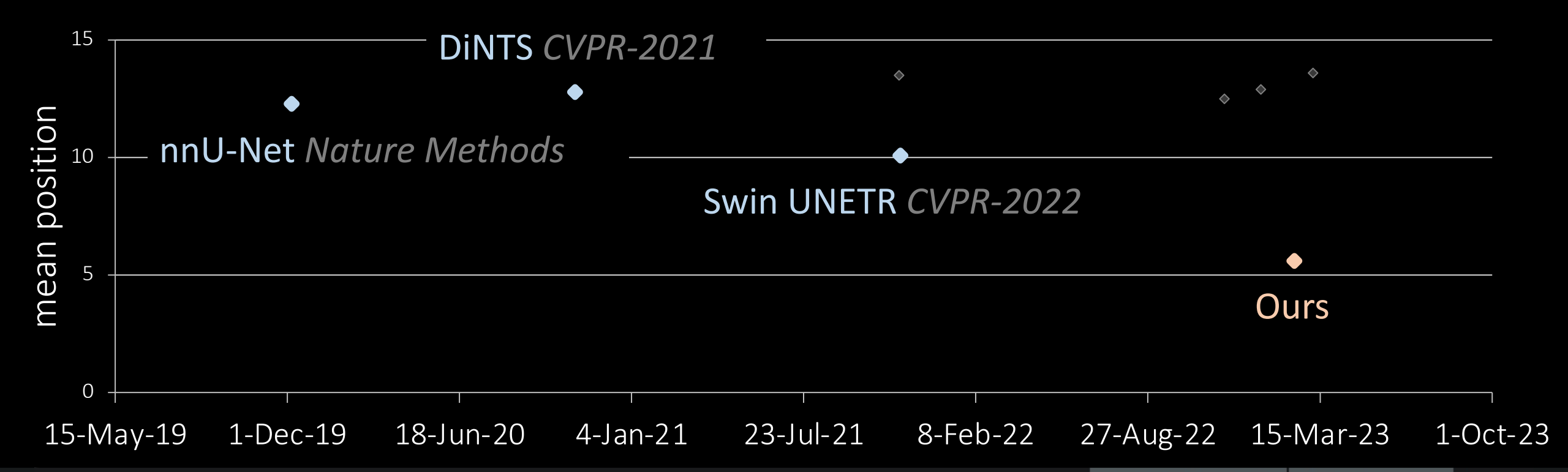
Challenge

Challenge Leaderboard

Search:

Additional metrics | Show all metrics

#	User (Team)	Created	Mean Position
1st	zongwei.zhou (universal_model)	13 Feb. 2023	5.6
2nd	Swin_UNETR	12 Nov. 2021	10.1
3rd	ahatamiz2	12 Nov. 2021	10.1
4th	Isensee	6 Dec. 2019	12.3
5th	AndyL	24 Nov. 2022	12.5
6th	heyufan1995	30 Oct. 2020	12.8
7th	qsyeung	5 Jan. 2023	12.9
8th	vishwesh.nath	11 Nov. 2021	13.5

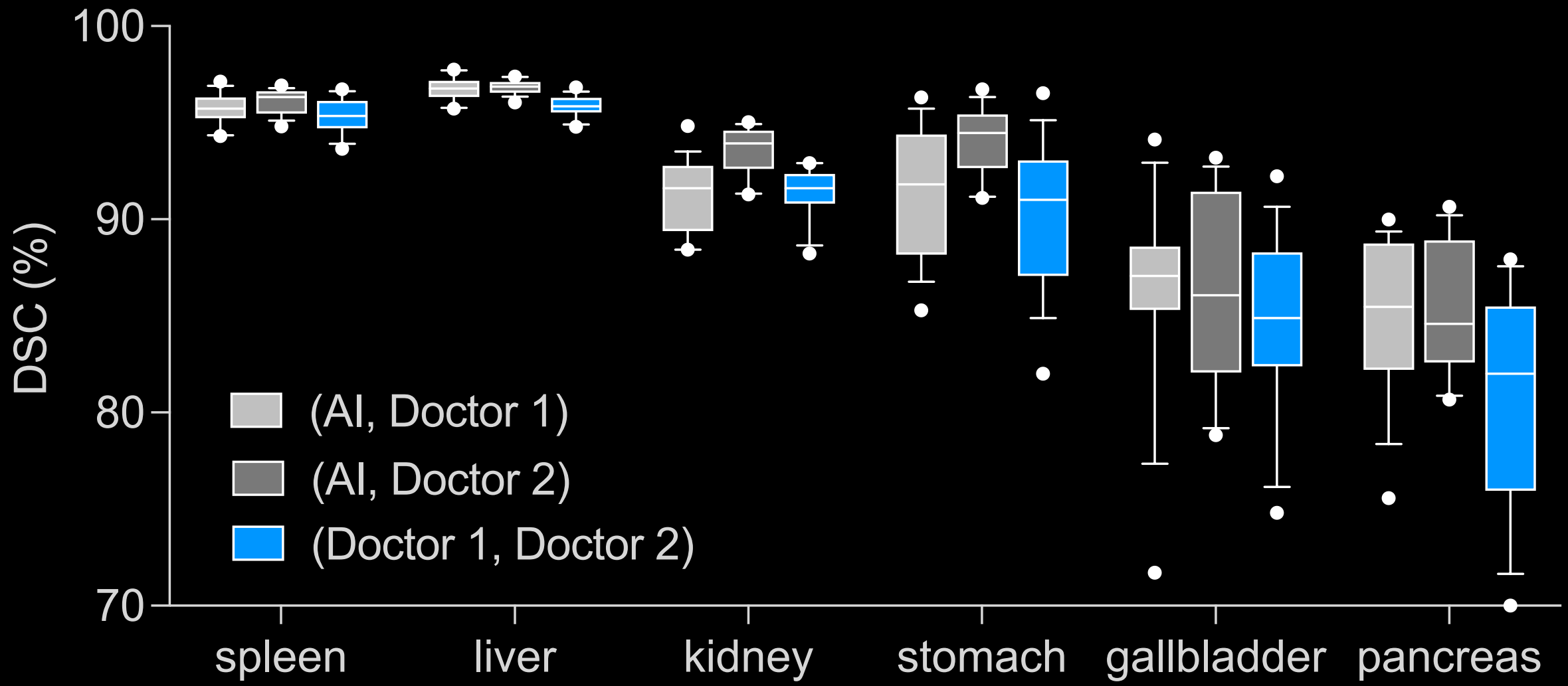


Additional metrics | Show all metrics

#	User (Team)	Created	Mean Position
1st	zongwei.zhou (universal_model)	13 Feb. 2023	5.6
2nd	Swin_UNETR	12 Nov. 2021	10.1
3rd	ahatamiz2	12 Nov. 2021	10.1
4th	Isensee	6 Dec. 2019	12.3
5th	AndyL	24 Nov. 2022	12.5
6th	heyufan1995	30 Oct. 2020	12.8
7th	qsyeung	5 Jan. 2023	12.9
8th	vishwesh.nath	11 Nov. 2021	13.5

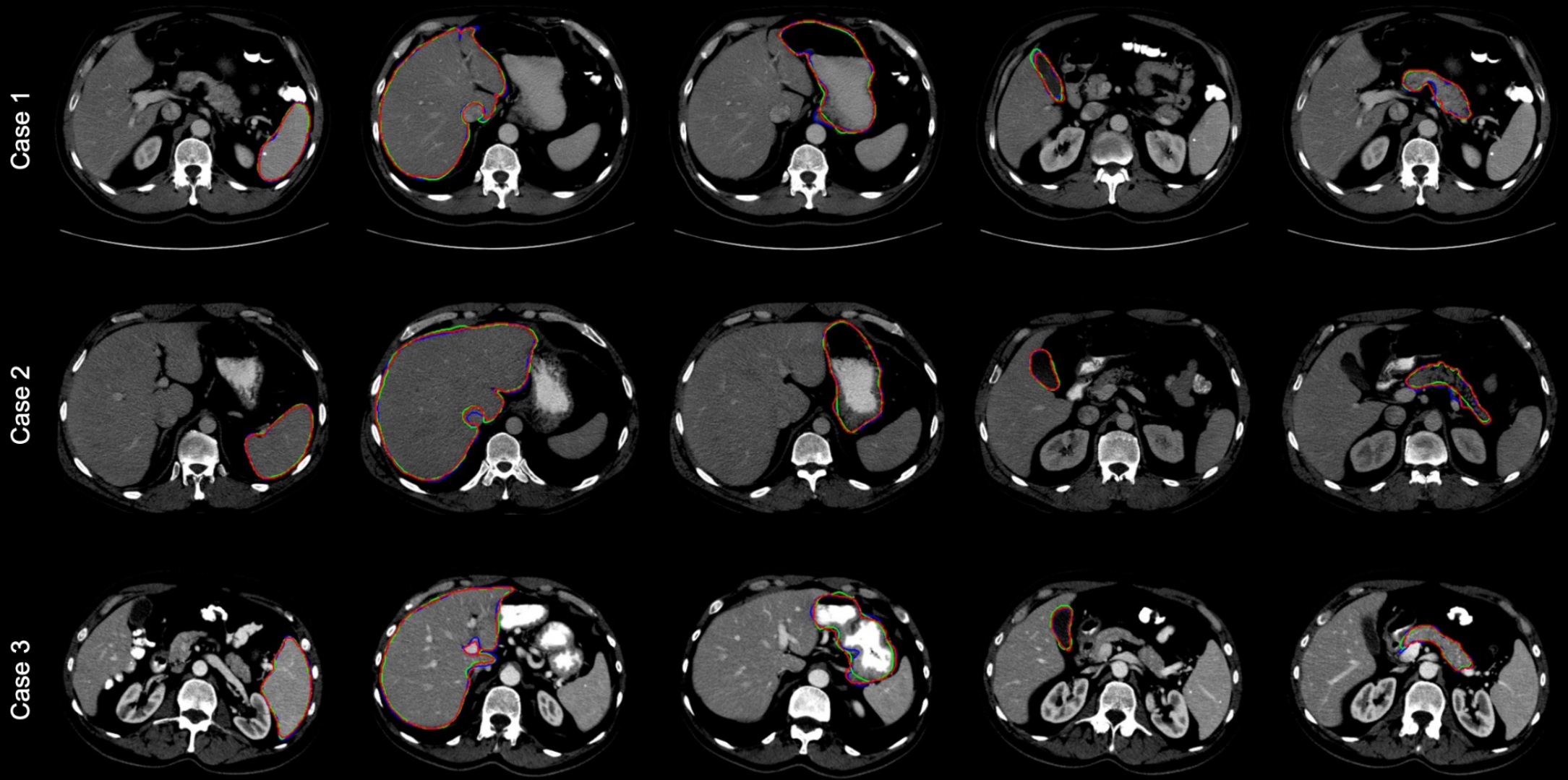
The AI predictions for six organs are comparable to expert annotators

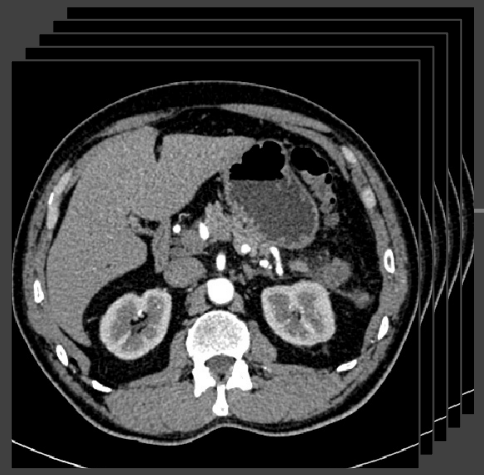
If we spend a lot more money to ask radiologists to annotate these six organs, it might turn out that the AI can do a similar quality annotation



The AI predictions for six organs are comparable to expert annotators

If we spend a lot more money to ask radiologists to annotate these six organs, it might turn out that the AI can do a similar quality annotation





Vision Encoder

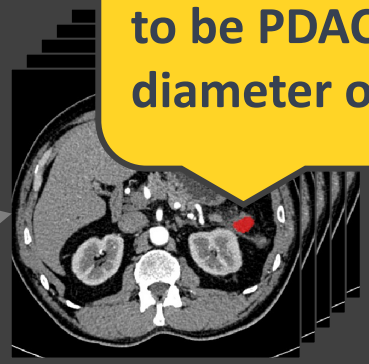
Universal Model

Please segment the tumor in the tail of the pancreas and then measure its size.

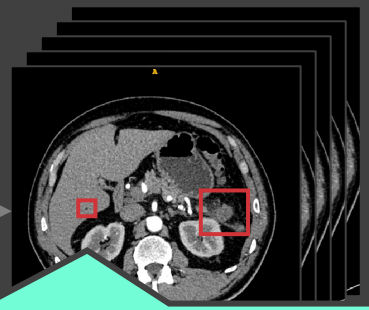
Take a look at these CT scans and mark the suspected tumor region.

Generate a report

Text Encoder



This tumor is likely to be PDAC with a diameter of 25mm.



Two potential tumors are framed in bounding boxes.

Medical Image Analysis: Scaling Annotations, Datasets, and Algorithms

- **Hierarchical annotations for organs/tumors**
- **Towards Foundation Models in healthcare**
  - High-performance & generalizable
  - Accommodating varied annotations and datasets
  - Vision-language, Multi-task capability

Scaling annotations

Efficient annotation
Human in the loop
Pathology reports

AbdomenAtlas-8K

8,448 annotated CT volumes

Code & Dataset
https://github.com/MrGiovanni/AbdomenAtlas
NeurIPS 2023

Scaling datasets

Multiple modalities
Diverse institutes
IRB approval

Tumor Synthesis

Annotation-free deep learning

Code & Turing Test
https://github.com/MrGiovanni/SyntheticTumors
CVPR 2023

Scaling algorithms

Vision-language
Lifelong learning
Reader study

Universal Model

25 organs and 7 cancers

Code & Model
https://github.com/ljwztc/CLIP-Driven-Universal-Model
ICCV 2023

Thank you!

Reference

Scaling Annotations

C. Qu, T. Zhang, H. Qiao, J. Liu, Y. Tang, A. Yuille, and Z. Zhou*. "Annotating 8,000 Abdominal CT Volumes for Multi-Organ Segmentation in Three Weeks." NeurIPS 2023.

Y. Chou, B. Li, D. Fan, A. Yuille, Z. Zhou*. “Acquiring Weak Annotations for Tumor Localization in Temporal and Volumetric Data.” MIR 2023.

T. Xiang, Y. Zhang, Y. Lu, A. Yuille, C. Zhang, W. Cai, Z. Zhou*. “SQUID: Deep Feature In-Painting for Unsupervised Anomaly Detection.” CVPR 2023.

Scaling Datasets

Q. Hu, Y. Chen, J. Xiao, S. Sun, J. Chen, A. Yuille, Z. Zhou*. “Label-Free Liver Tumor Segmentation.” CVPR 2023.

B. Li, Y. Chou, S. Sun, H. Qiao, A. Yuille, Z. Zhou*. “Early Detection and Localization of Pancreatic Cancer by Label-Free Tumor Synthesis.” MICCAI Workshop 2023.

Scaling Algorithms

J. Liu, Y. Zhang, J. Chen, Y. Lu, Y. Yuan, A. Yuille, Y. Tang*, Z. Zhou*. “CLIP-Driven Universal Model for Organ Segmentation and Tumor Detection.” ICCV 2023.

Y. Zhang, X. Li, H. Chen, A. Yuille, Y. Liu*, Z. Zhou*. “Learning without Forgetting for Continual Abdominal Multi-Organ and Tumor Segmentation.” MICCAI 2023.

M. Kang, B. Li, Z. Zhu, Y. Lu, E. Fishman, A. Yuille, Z. Zhou*. “Label-Assemble: Leveraging Multiple Datasets with Partial Labels.” ISBI 2023.