ScaleMAI: Accelerate the Development of Scalable Medical Artificial Intelligence

Zongwei Zhou, PhD
Dept. Computer Science, Johns Hopkins University
zzhou82@jh.edu | (480)738-2575

JOHNS HOPKINS UNIVERSITY

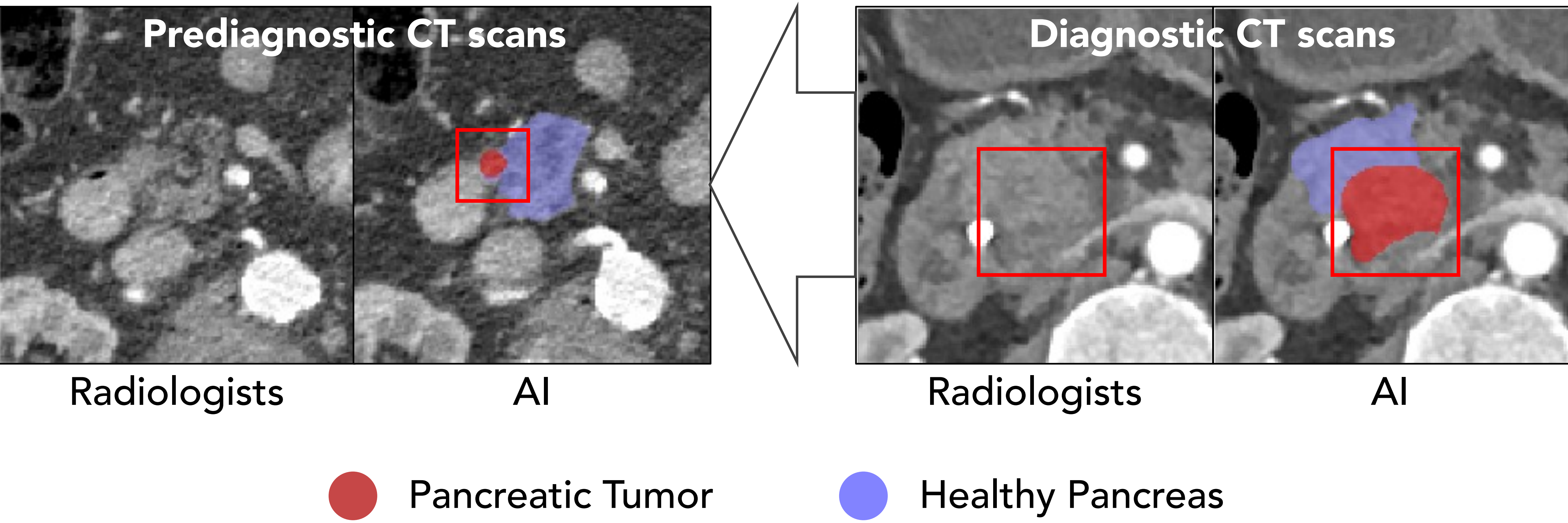
A Successful Story

- Early detection of cancer enables surgery and will save many lives.
- Radiologists can detect pancreatic cancer from CT scans, but the sensitivity of early pancreatic cancer is only **33-44%**.
- Our AI has achieved very high performance in early detection.

	Sensitivity early tumors ≤2 cm	Sensitivity all-size tumors	Specificity
Radiologists	33–44%	76–92%	82–96%
Our AI	94%	97%	99%

A Successful Story

8 months earlier ...

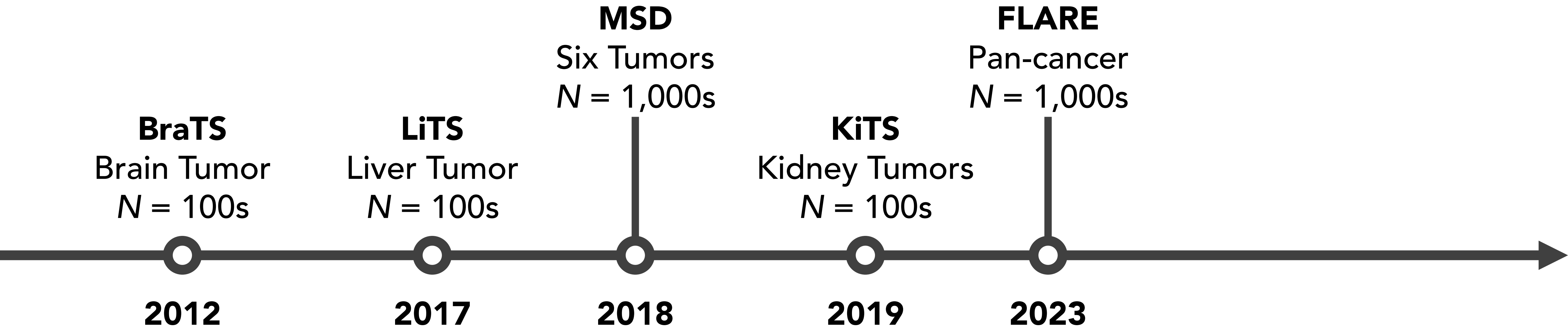


Challenges of Getting Big Data

- We don’t have enough “AI-ready” data to train these models.
- AI-ready? Voxel-wise annotation is very time consuming and requires experts (e.g., the JHU project required **25 person years**).
- In FELIX, joint project between CS and Radiology, the radiologist team has collected and annotated more than 5,000 CT scans (over 2.5 million images). This is largest dataset in the world to our knowledge dedicated to pancreatic cancer.
- But scaling this effort to all cancer types is not feasible.

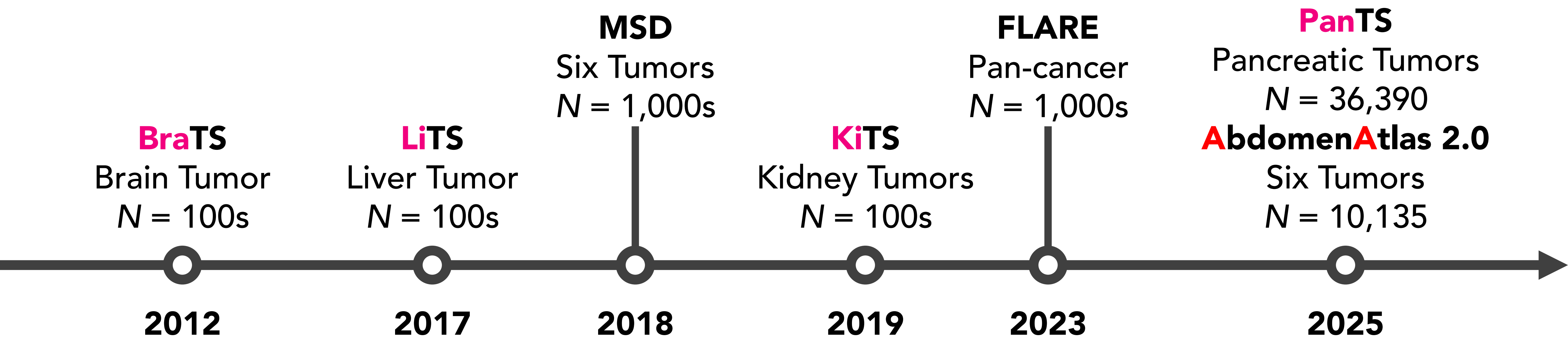
Annotated Tumor Datasets Should Be Open to More Researchers

- There's a huge data gap in medical AI right now, particularly when you have rare diseases, uncommon conditions (e.g., cancer).



Annotated Tumor Datasets Should Be Open to More Researchers

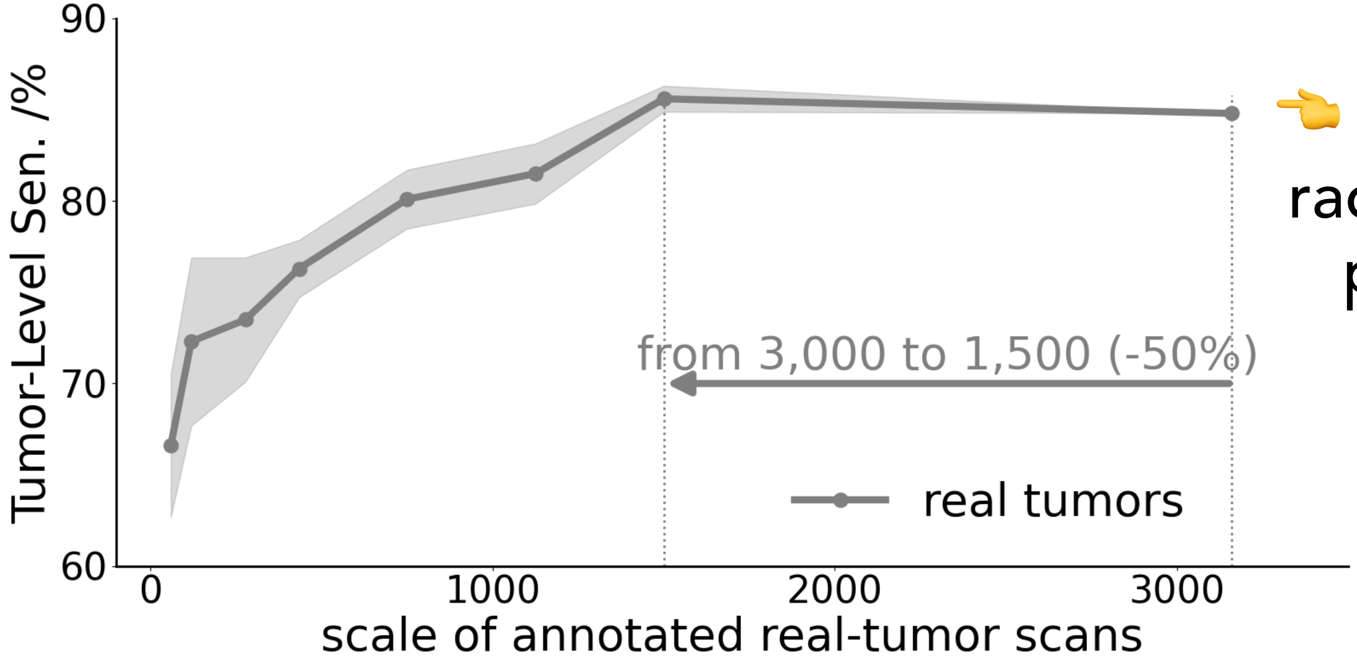
- There's a huge data gap in medical AI right now, particularly when you have rare diseases, uncommon conditions (e.g., cancer).



1. **Can We Reduce the Need of Voxel-Wise Annotations?**
2. **Can We Reduce the Cost of Voxel-Wise Annotations?**

Scaling Laws in Tumor Detection

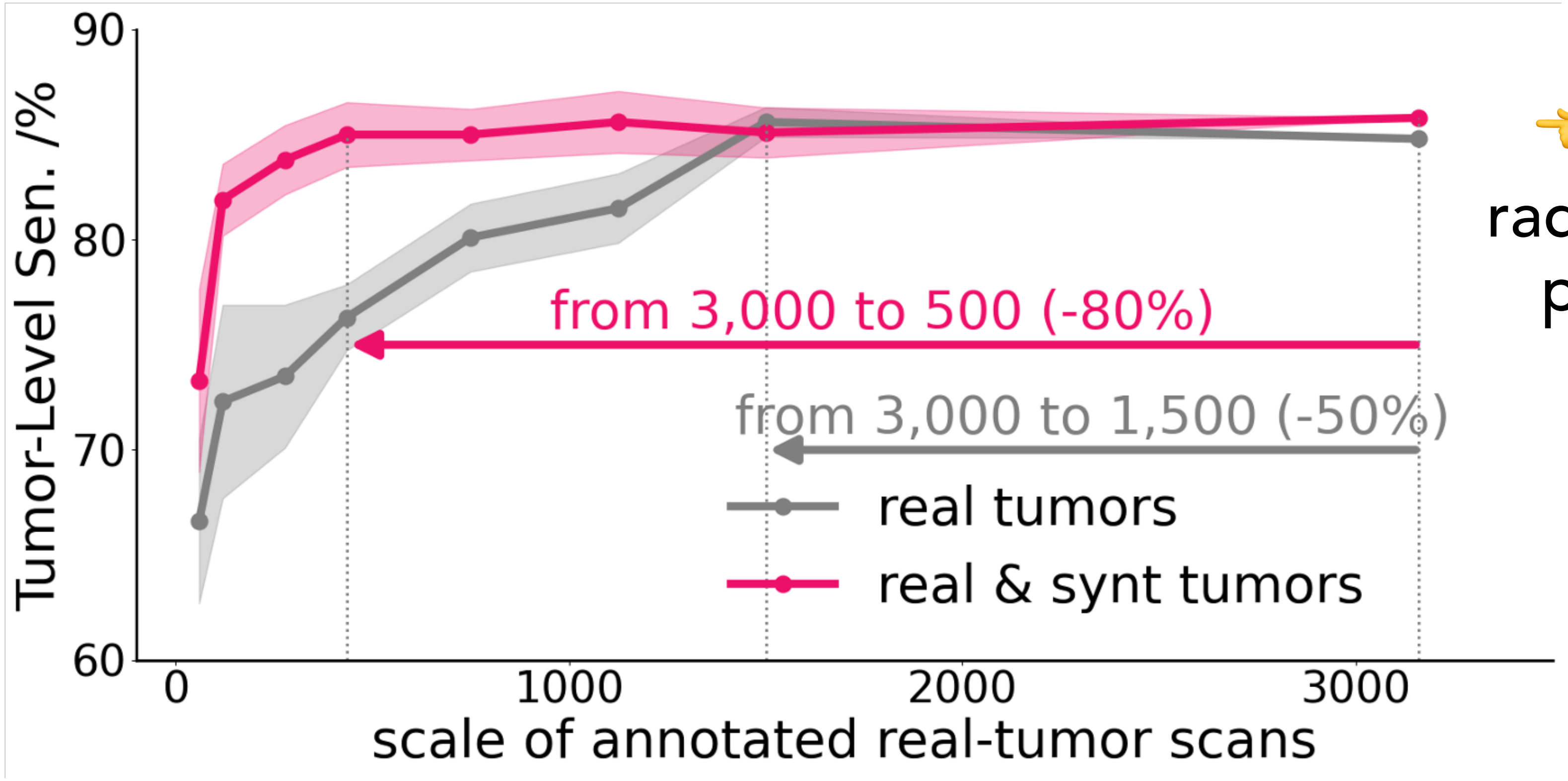
- Detection performance reaches a plateau at n = 1,500 with more data.
- But annotating 1,500 CT scans remain challenging, can we do better?
- *Note: this is an internal validation!*



👈 comparable to radiologist-level performance

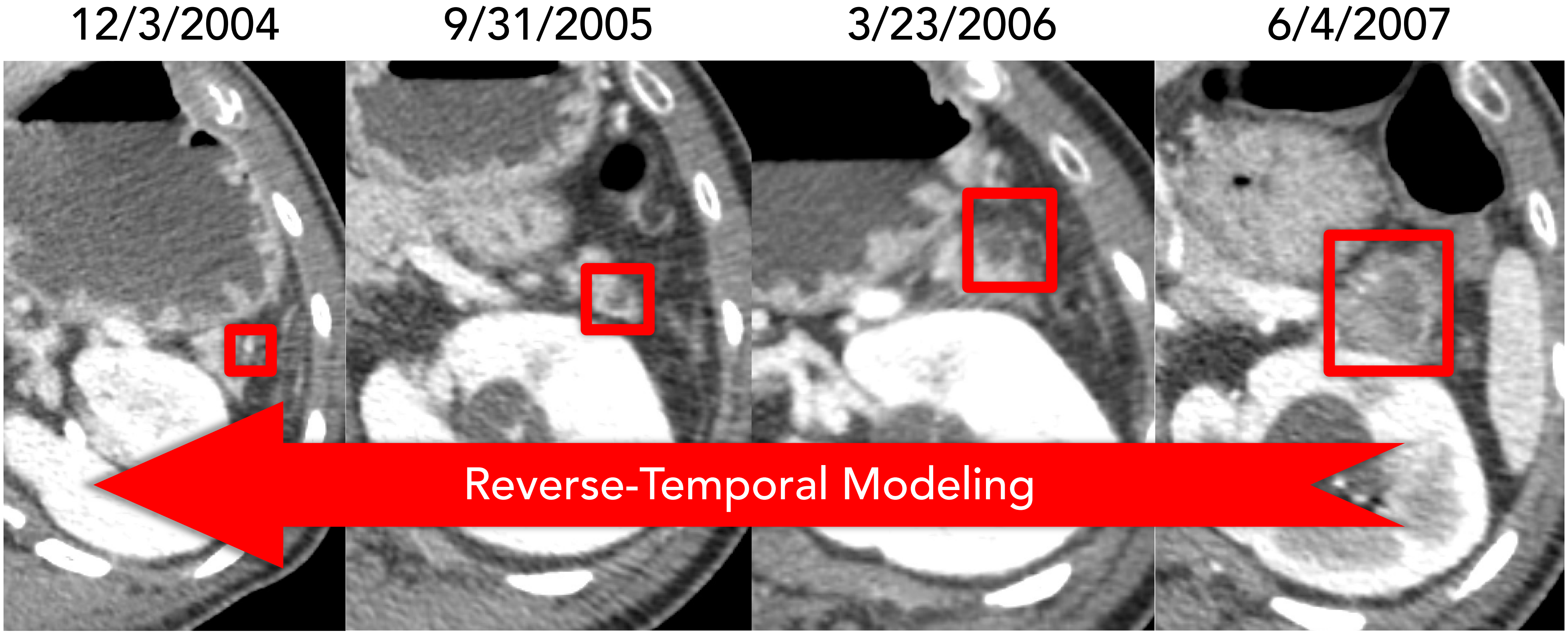
Scaling Laws in Synthetic Tumors

- Synthetic data reduces the need for voxel-wise annotated real data from 1,500 down to 500. (Q. Chen et al., ICCV 2025).

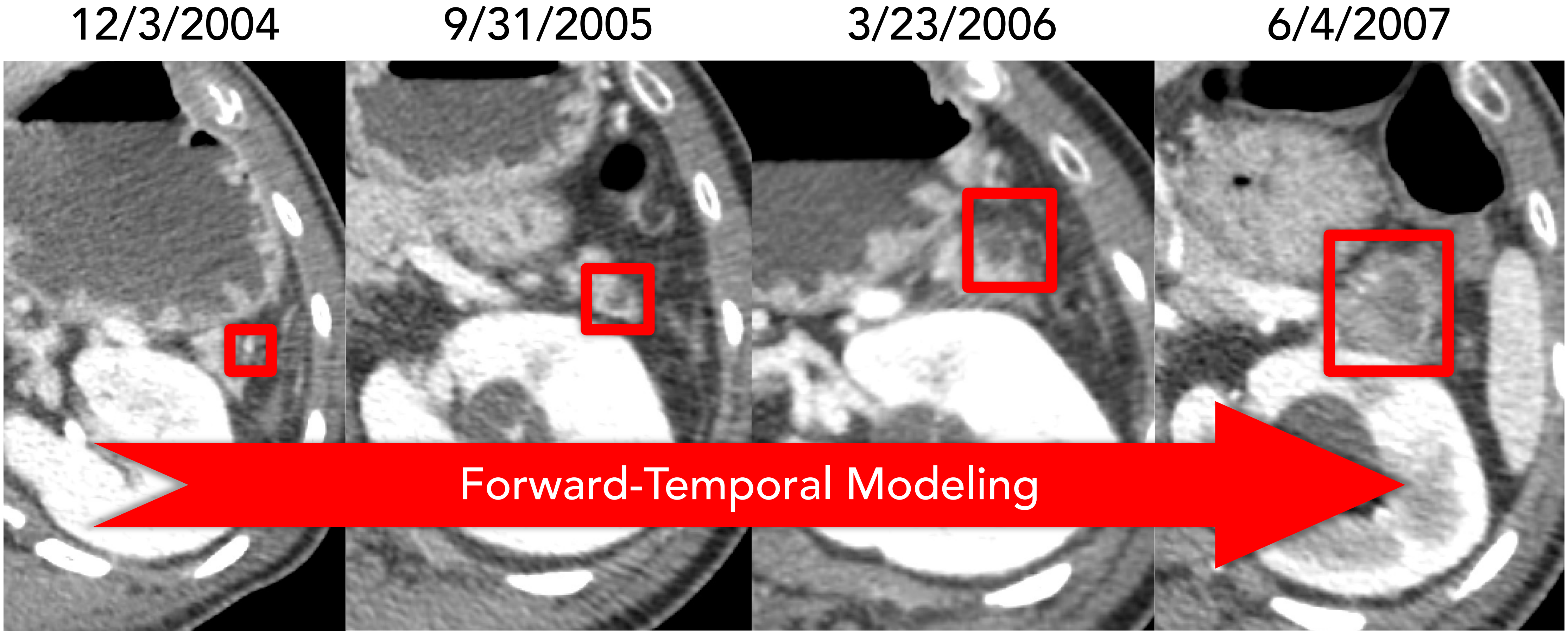


👈 exceeding radiologist-level performance

Synthetic Tumors as Time Machine

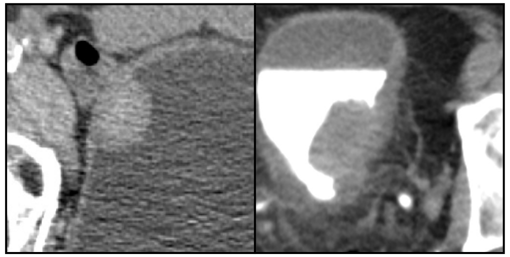


Synthetic Tumors as Time Machine

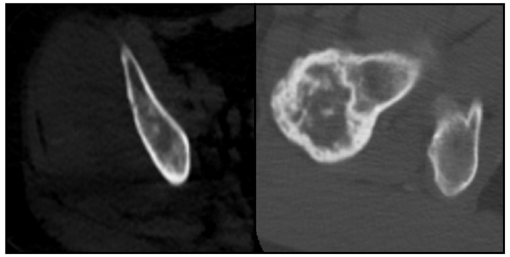


Visual Turing Test for Radiologists

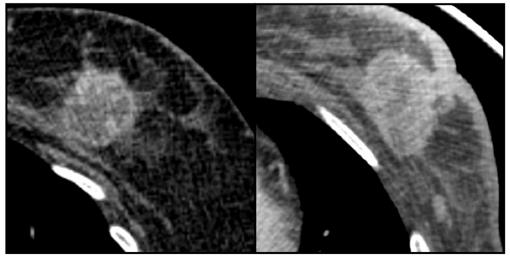
(a) real or fake test



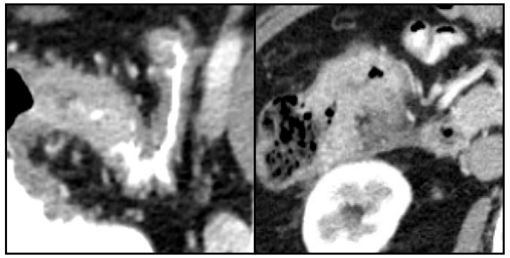
(b) bladder tumor



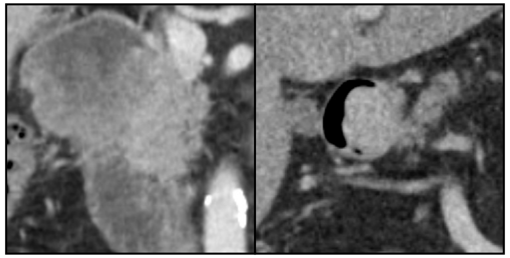
(c) bone tumor



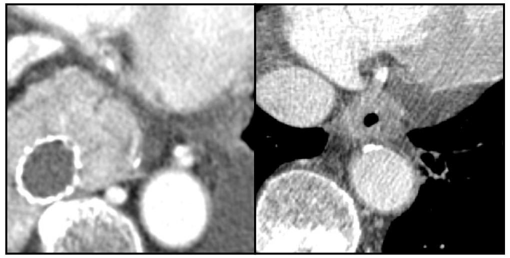
(d) breast tumor



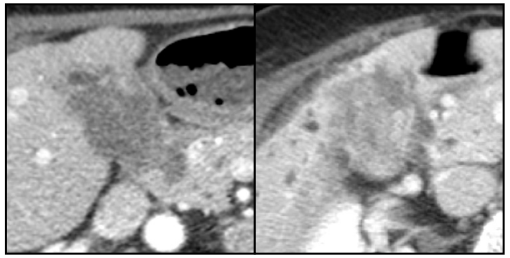
(e) colon tumor



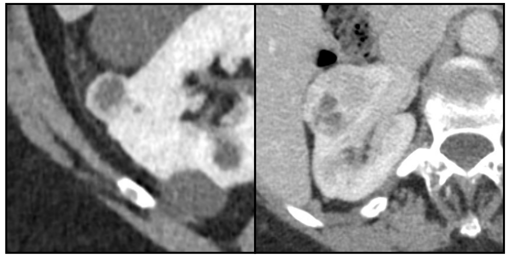
(f) duodenum tumor



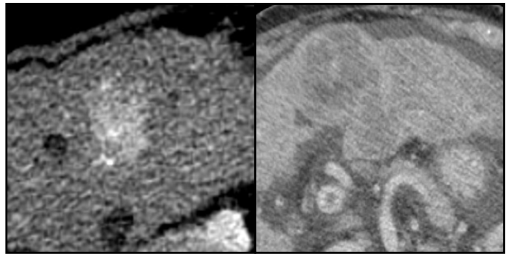
(g) esophagus tumor



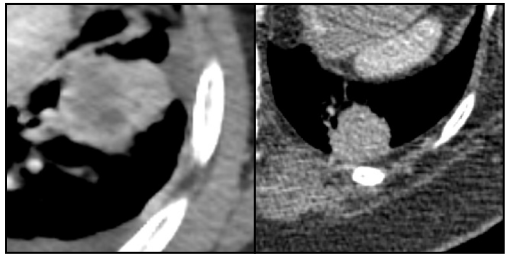
(h) gallbladder tumor



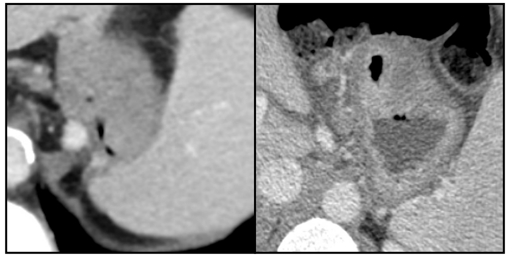
(i) kidney tumor



(j) liver tumor



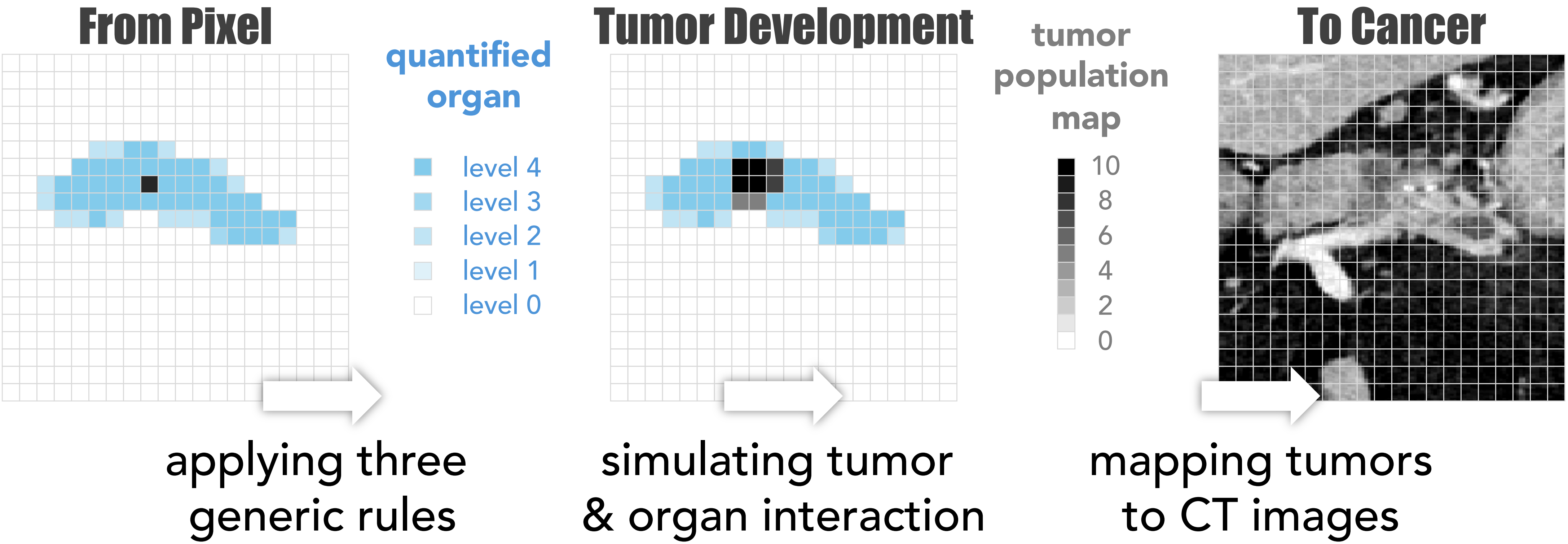
(k) lung tumor



(l) stomach tumor

Tumor/Vessel/Duct/Organ Synthesis

- We developed "game of life" to simulate tumor development (Lai et al., MICCAI 2024) and applied diffusion models to create synthetic tumors.

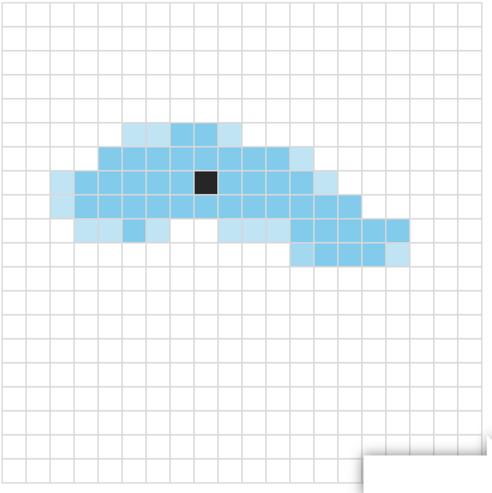


Tumor/Vessel/Duct/Organ Synthesis

Cellular Automata

a mathematical model that uses simple rules to simulate complex systems

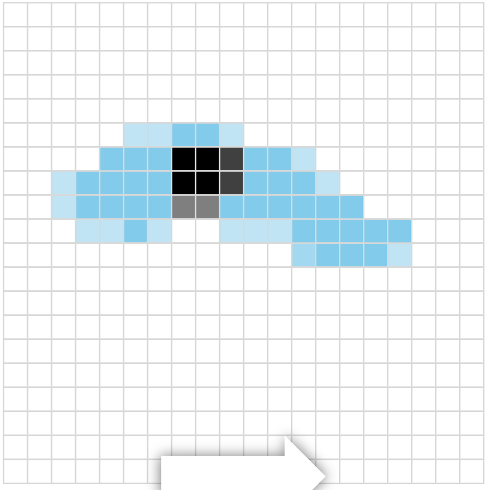
From Pixel



quantified organ

- level 4
- level 3
- level 2
- level 1
- level 0

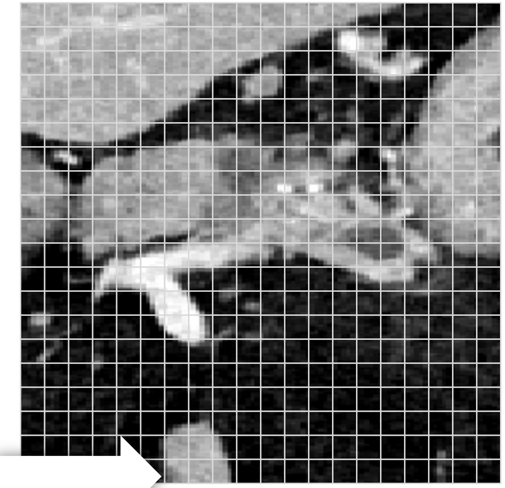
Tumor Development



tumor population map

10
8
6
4
2
0

To Cancer



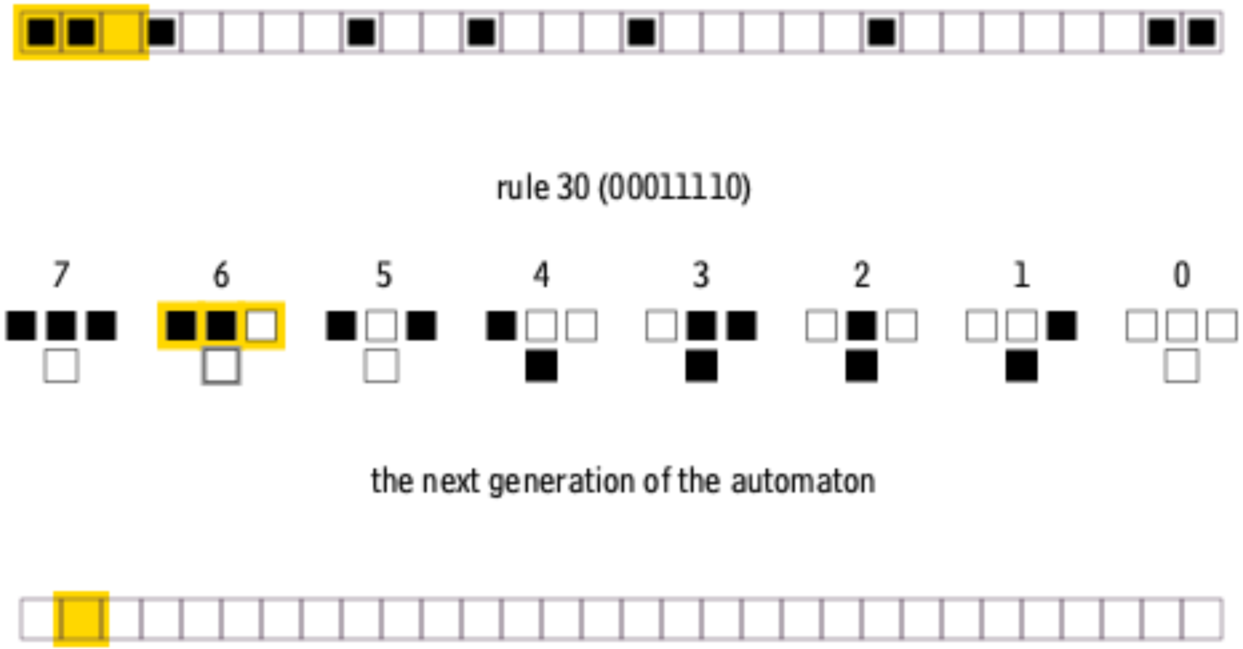
applying three generic rules

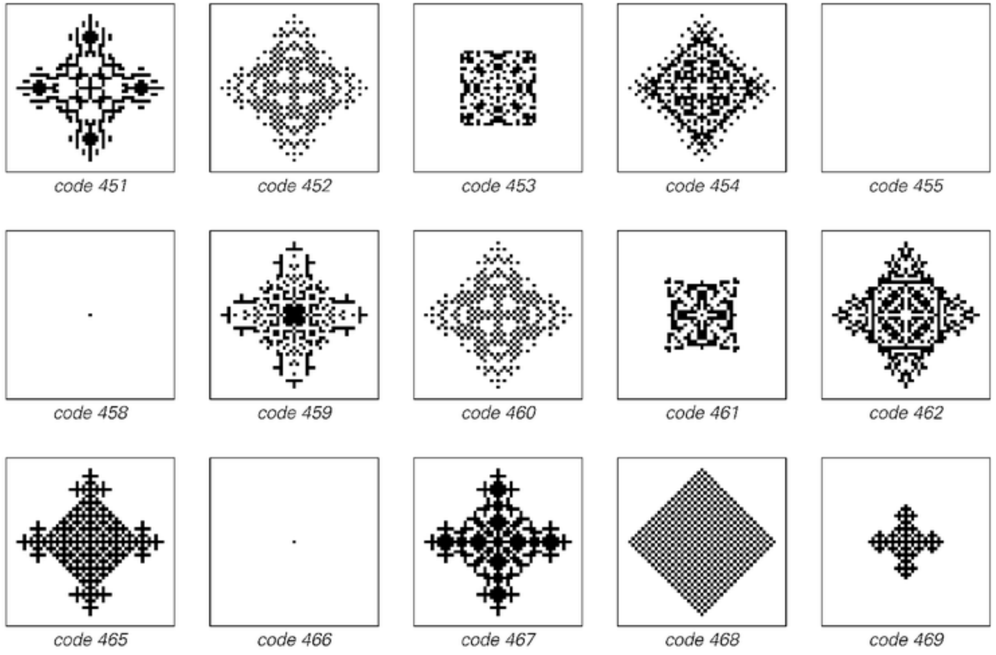
simulating tumor & organ interaction

mapping tumors to CT images

Cellular Automata

- Cellular Automata (CA) is a mathematical model that uses simple rules to simulate complex systems.
- CAs consist of grids with evolving cell states influenced by neighbors.

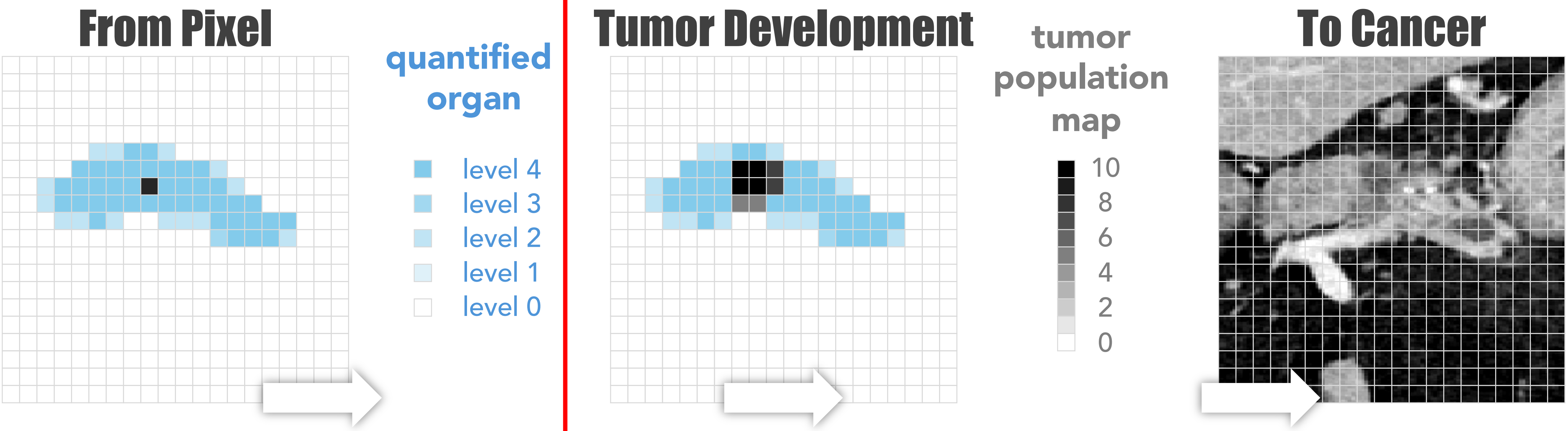




Tumor/Vessel/Duct/Organ Synthesis

Diffusion Models

conditioned on tumor/vessel/duct/organ shapes simulated by cellular automata

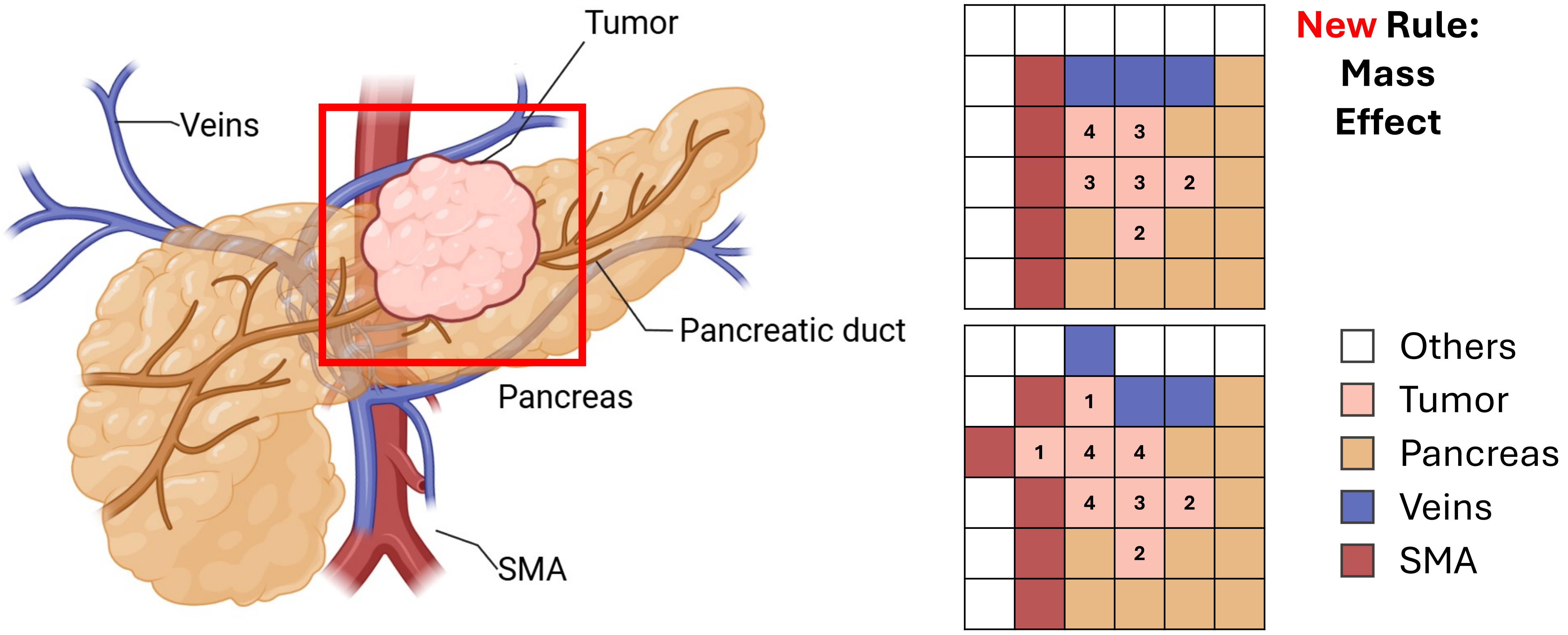


applying three generic rules

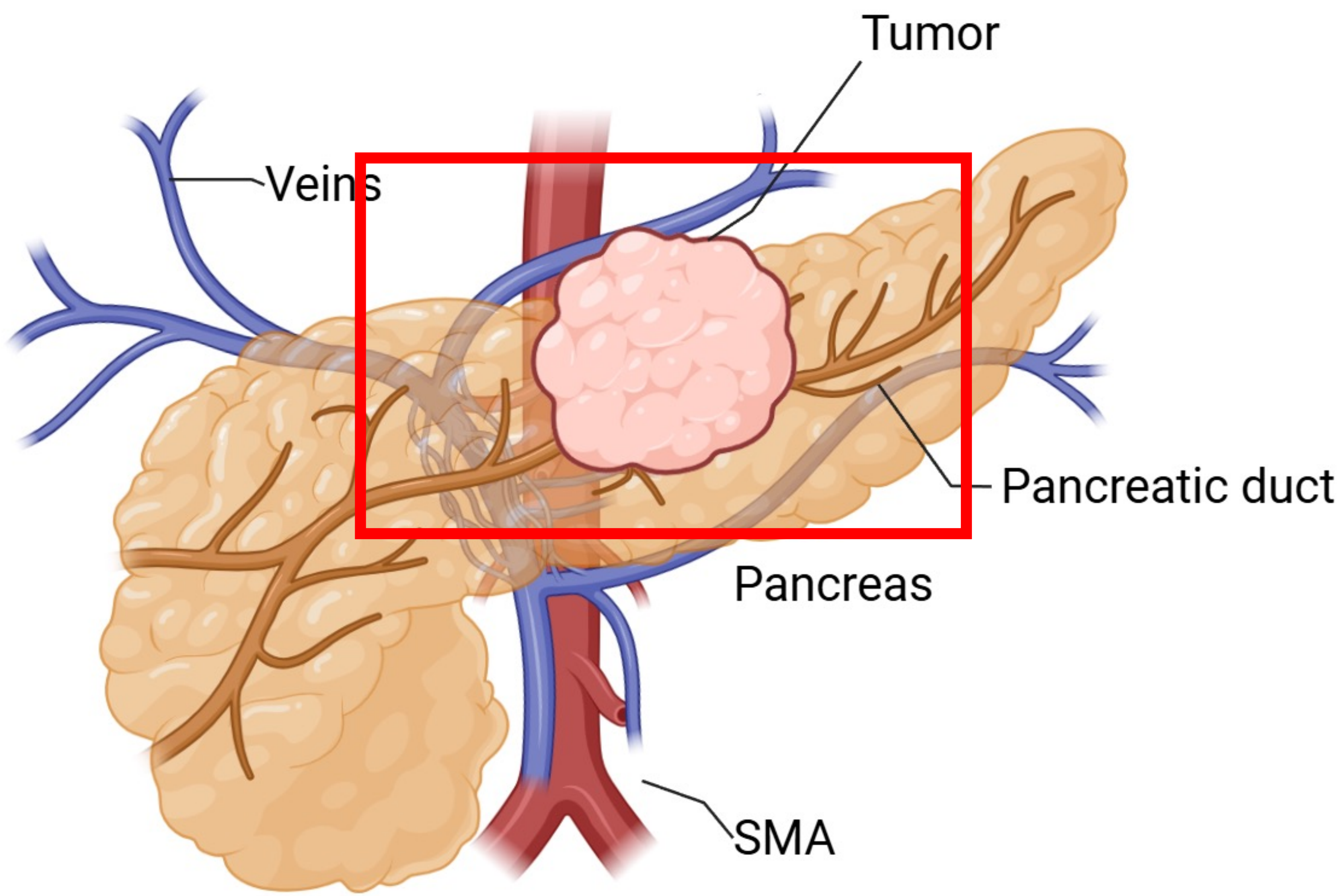
simulating tumor & organ interaction

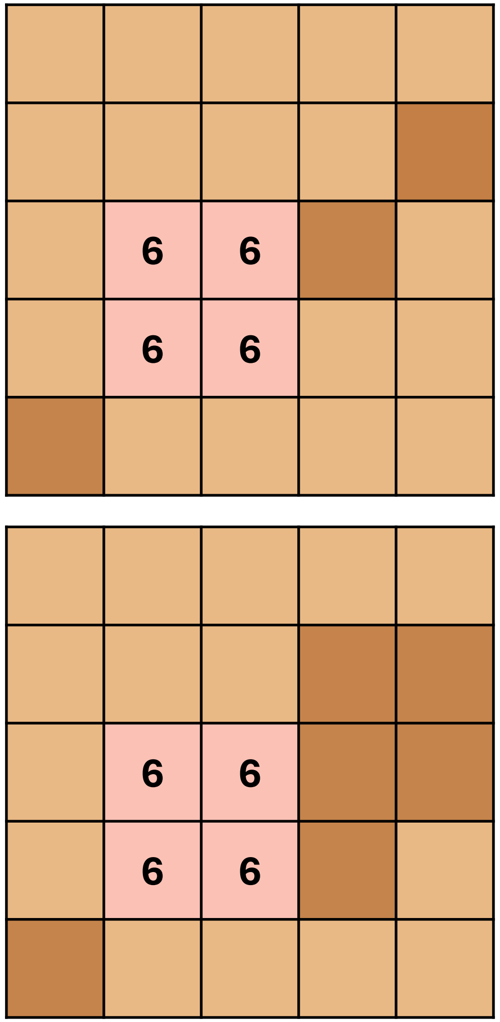
mapping tumors to CT images

Tumor/Vessel/Duct/Organ Synthesis

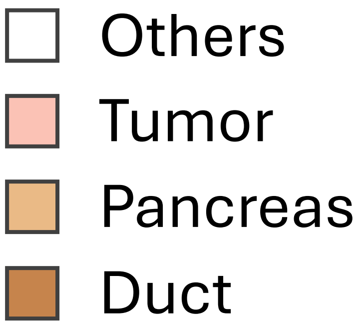


Tumor/Vessel/Duct/Organ Synthesis



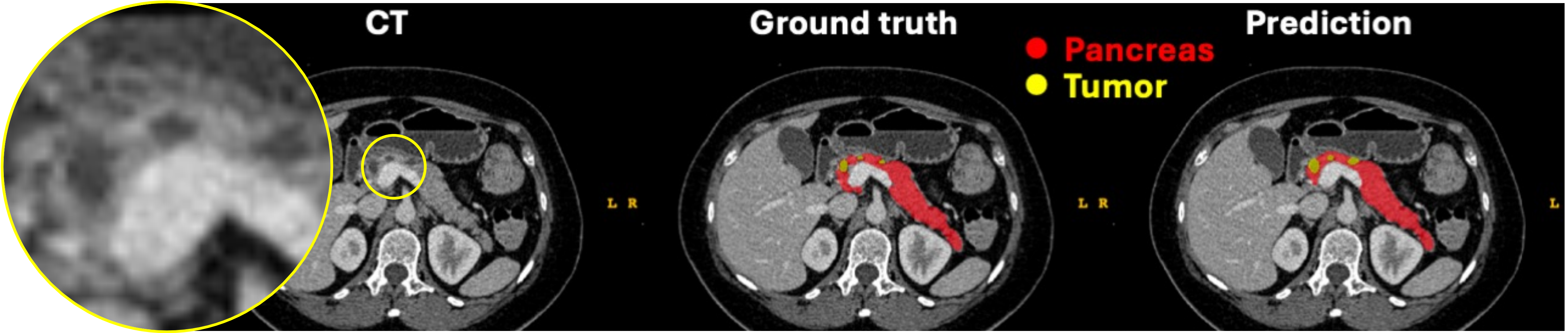


New Rule:
Duct Dilation



Result I. Synthetic Data for Small Tumors

- Synthetic data improves sensitivity of detecting small tumors (≤2 cm) by **5% (89% → 94%)** (Q. Chen et al., CVPR 2024; Q. Hu et al., CVPR 2023)
- The smallest lesion we detected was 2 mm.





GitHub.com/MrGiovanni/SyntheticTumors

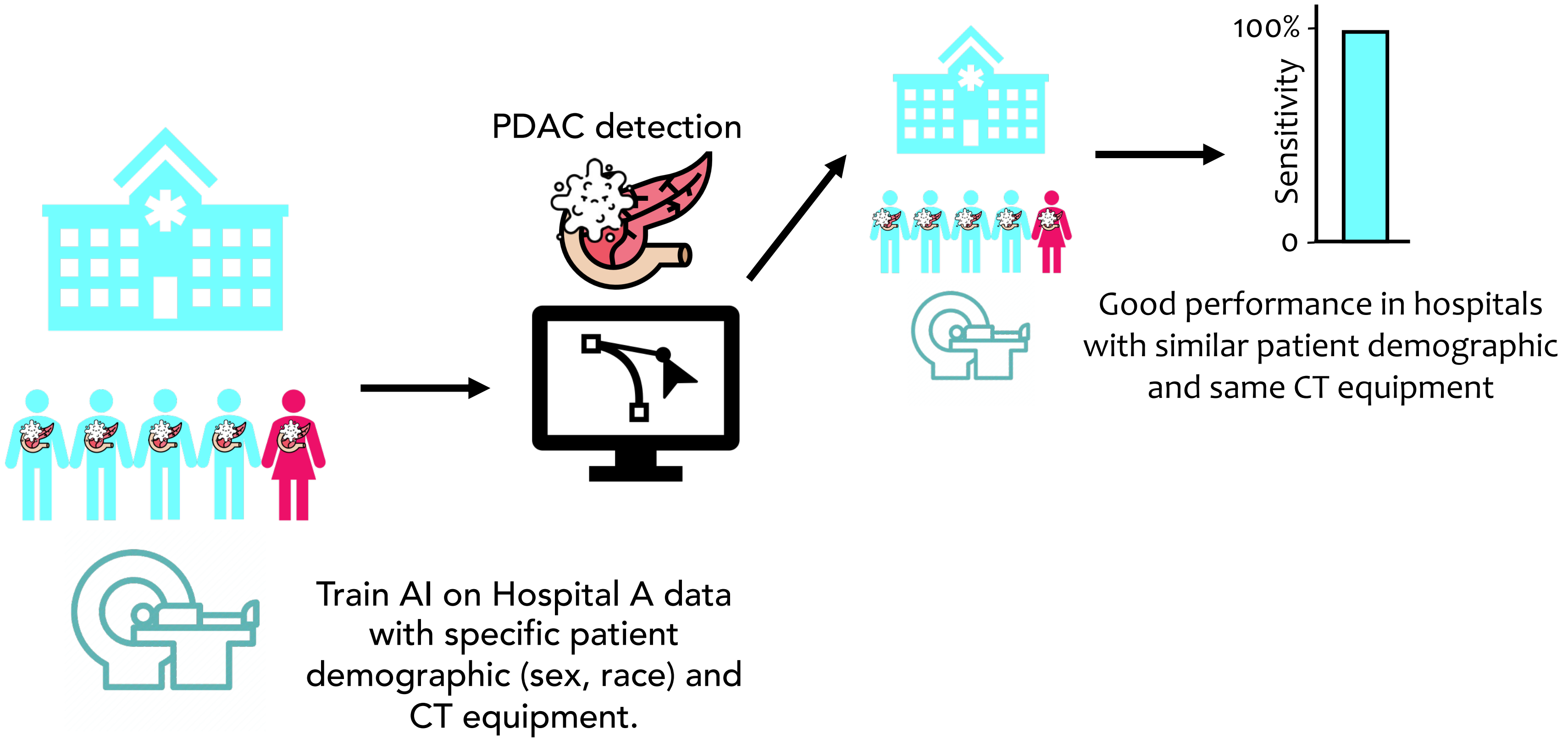
Result II. Synthetic Data for Augmentation

- Real-world tumor datasets are often biased.
  - E.g., Tumor location: about 65% of pancreatic tumors arise in the head of the pancreas with the remaining roughly one-third in the body or tail.

- Targeted data augmentation enabled by error analysis and synthetic data (X. Li et al., TMI 2025 In Submission)
  - The AI often misses small tumors in the body or tail of the pancreas.
  - We can add more synthetic small tumors in these regions during training.

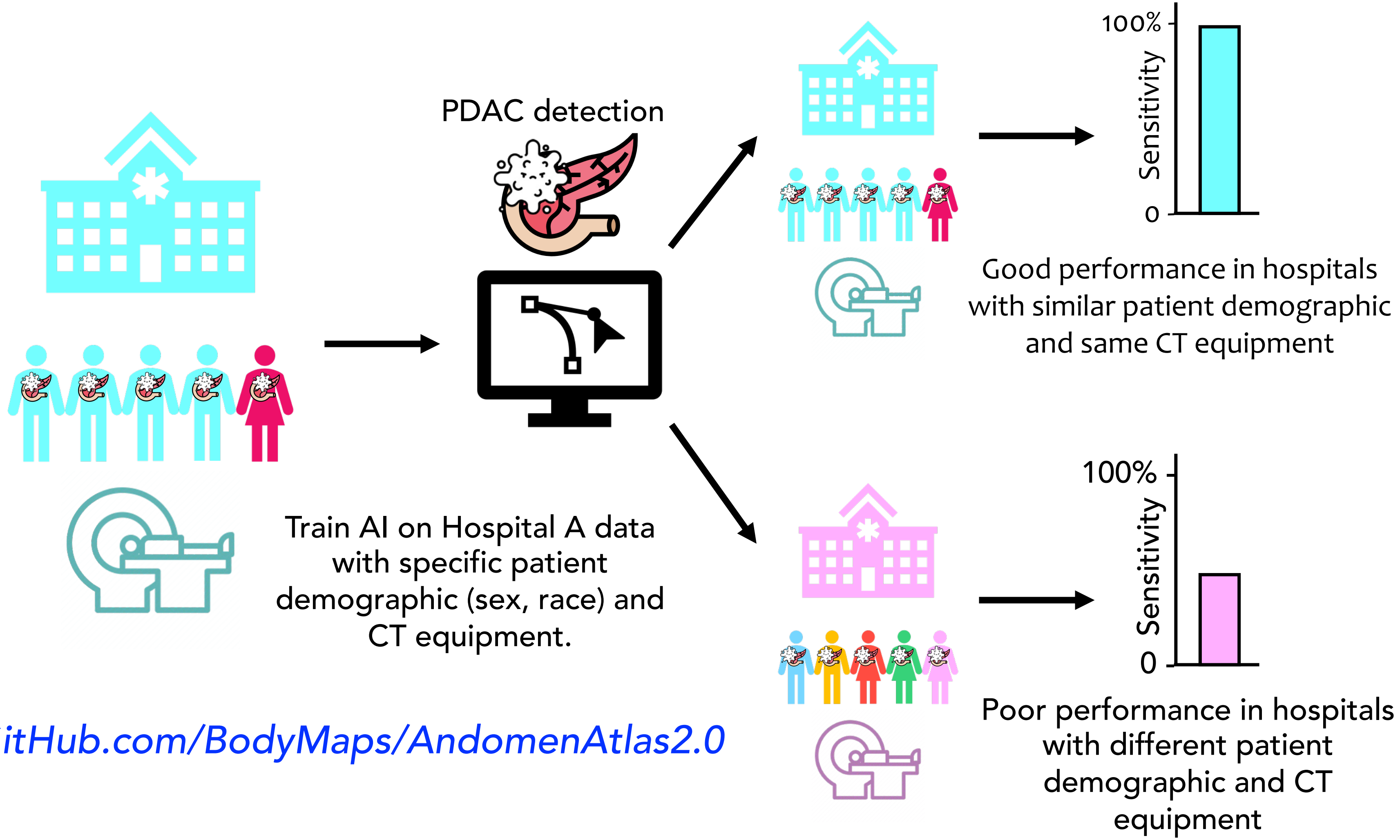
GitHub.com/MrGiovanni/TextoMorph

Result III. Generalized to Different Hospitals



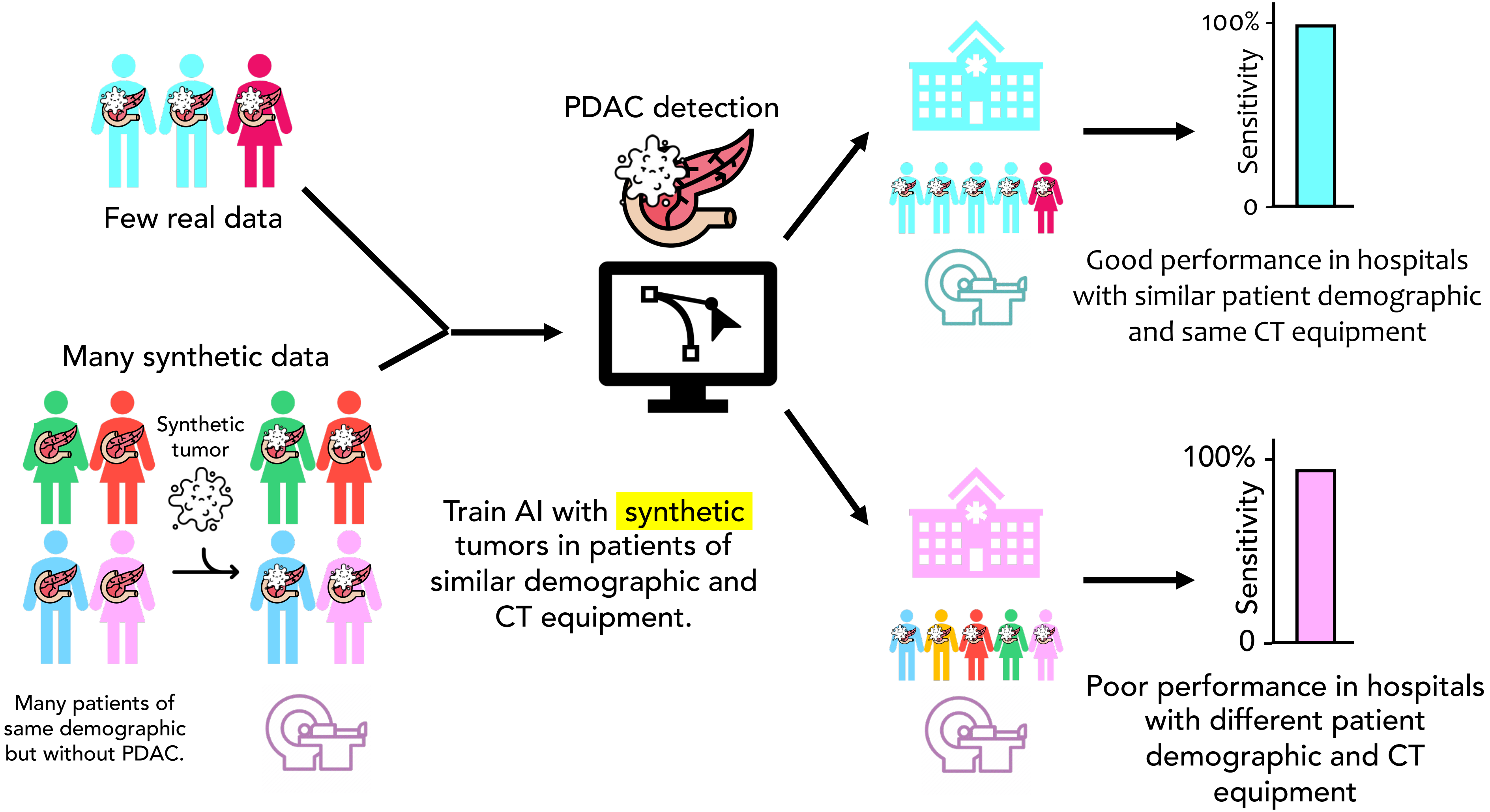
GitHub.com/BodyMaps/AndomenAtlas2.0

Result III. Generalized to Different Hospitals



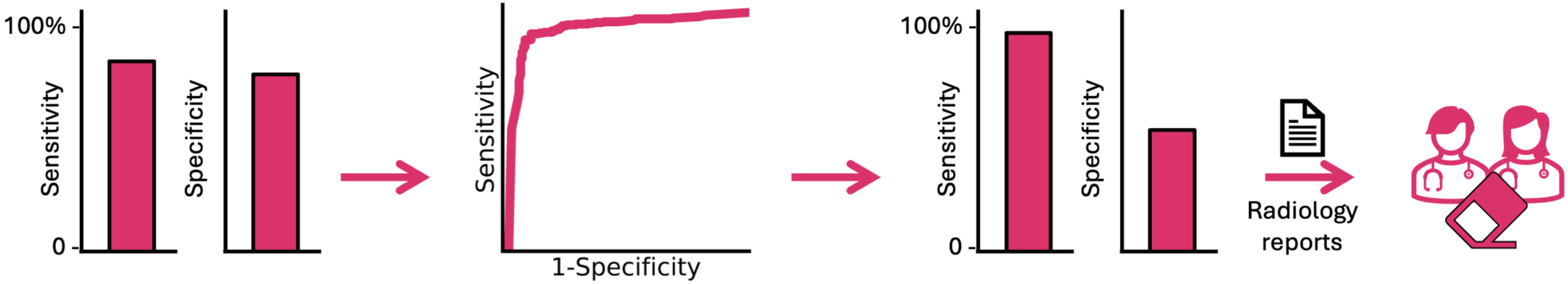
GitHub.com/BodyMaps/AndomenAtlas2.0

Result III. Generalized to Different Hospitals



Result IV. Efficient Tumor Annotations

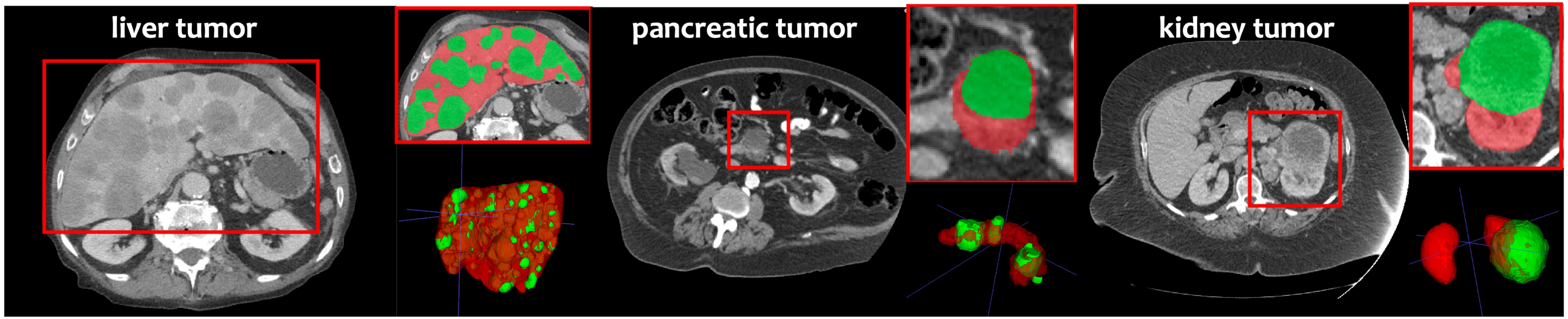
- Make the AI highly sensitive, offering a strong starting point for radiologist review and edit at least **80×** faster (Zhou et al., ISBI 2024).



Adjust AI prediction threshold to make the AI overly sensitive,
i.e., false negatives -- false positives ++

Result IV. Efficient Tumor Annotations

- Make the AI highly sensitive, offering a strong starting point for radiologist review and edit at least **80×** faster (Zhou et al., ISBI 2024).
- (I) Editing an AI-generated tumor takes **~1 minute**. (rarely needed)
- (II) Removing a false positive takes **<5 seconds**.
- In contrast, manual annotation from scratch takes **4–5 minutes**.



AbdomenAtlas 2.0 (Chen et al., ICCV 2025)

- A large, multi-center (*real*) tumor dataset with voxel-wise annotations.
- 10,135 CT scans + 8,562 tumor masks.
- Get early access 👉https://www.zongweiz.com/dataset
- Much larger than existing public datasets.
- (I) **46x** LiTS (201 liver scans).
- (II) **22x** MSD-Pancreas (420 pancreatic scans).
- (III) **15x** KiTS (599 kidney scans).

🤝 **A team of 23 board-certified radiologists**

PanTS

Pancreatic Tumor Segmentation



GitHub.com/MrGiovanni/PanTS

```
git clone https://github.com/MrGiovanni/PanTS.git; cd PanTS
bash download_PanTS_data.sh
bash download_PanTS_label.sh
http://www.cs.jhu.edu/~zongwei/dataset/PanTSMini_Label.tar.gz
```

PanTS is a large-scale, multi-institutional dataset, containing **36,390** three-dimensional CT volumes from **145** medical centers, with expert-validated, voxel-wise annotations of over **993,000** anatomical structures, including ***pancreatic tumors, pancreas head, body, and tail, and 24 surrounding anatomical structures such as vascular/skeletal structures and abdominal/thoracic organs.***

(Li et al., NeurIPS 2025)

GitHub.com/MrGiovanni/PanTS

Summary (Technical – Synthetic Tumors)

1. Generative models can create realistic synthetic tumors. Radiologists check the realism of synthetic tumors and provide feedback.
2. Synthetic tumors supplement the strong per-voxel annotations in dataset like the JHU dataset to provide additional training data.
3. Synthetic tumors can be created at very small scale. This helps because current datasets are short of very small tumors (<2 cm, <1 cm) because they are so hard to detect.
4. Synthetic tumors improve AI generalization across hospitals. With only normal CT scans from these hospitals, the AI can be fine-tuned using synthetic tumors combined with these scans.

Summary

1. ***Can We Reduce the Need of Voxel-Wise Annotations?***
   - The data-scaling experiment shows that:
   - *N* = 1,500 annotated real-world tumor scans without synthetic data.
   - *N* = 500 annotated real-world tumor scans with synthetic data.
   - *But this estimation is only based on in-distribution (@JHU) testing.*
2. ***Can We Reduce the Cost of Voxel-Wise Annotations?***
   - The use of synthetic data can speed up the annotations by 80x.
   - We publicly released PanTS (*N* = 36,390) for pancreatic tumor segmentation.
   - We publicly released AbdomenAtlas 2.0 (*N* = 10,135) for multicancer segmentation.

- *Note: synthetic tumors are used for training AI only – not for testing.*

Key References

- Bassi, Pedro RAS, Wenxuan Li, ..., Alan Yuille, and **Zongwei Zhou**. "Touchstone benchmark: Are we on the right way for evaluating ai algorithms for medical segmentation?" *NeurIPS, 2024.*
- Bassi, Pedro RAS, …, **Zongwei Zhou**. "Learning Segmentation from Radiology Reports." *MICCAI, 2025* (Runner-up, Best Paper Award).
- Bassi, Pedro RAS, …, **Zongwei Zhou**. "RadGPT: Constructing 3D Image-Text Tumor Datasets." *ICCV, 2025.*
- Chen, Qi, …, and **Zongwei Zhou**. "Towards generalizable tumor synthesis." *CVPR, 2024.*
- Lai, Yuxiang, …, **Zongwei Zhou**. "From pixel to cancer: Cellular automata in computed tomography." *MICCAI, 2024.*
- Li, Wenxuan, …, **Zongwei Zhou**. "How well do supervised 3d models transfer to medical imaging tasks?" *ICLR, 2025* (Oral).
- Li, Wenxuan, …, **Zongwei Zhou**. "PanTS: The Pancreatic Tumor Segmentation Dataset." NeurIPS, 2025.
- Li, Wenxuan, …, **Zongwei Zhou**. "Abdomenatlas: A large-scale, detailed-annotated, & multi-center dataset for efficient transfer learning and open algorithmic benchmarking." *MEDIA, 2024.*
- Lubonja, Ariel, …, **Zongwei Zhou**. "Auditing Significance, Metric Choice, and Demographic Fairness in Medical AI Challenges." *MLMI, 2025.*
- Xia, Yingda, Qihang Yu, …, **Zongwei Zhou**, …, Alan Yuille. "The felix project: Deep networks to detect pancreatic neoplasms." *medRxiv, 2022.*