



Large-Scale Multi-Cancer Detection by Learning Segmentation from Reports

Pedro R. A. S. Bassi

Postdoctoral Researcher (2025-present)

Dept. Computer Science, Johns Hopkins University

psalvad2@jh.edu



The Medical Image Computing
and Computer Assisted
Intervention Society

Best Paper Award

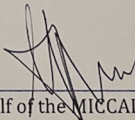
— RUNNER UP —

Pedro R. A. S. Bassi

For the paper titled:

"Learning Segmentation from Radiology Reports"




On behalf of the MICCAI Society Board
September 26, 2025

This talk briefly covers multiple papers
(9, first author)

Awards:

- **MICCAI 2025 Best Paper Award**
(runner-up), top 2 of 1,027 papers
- **RSNA 2025 Certificate of Merit Award**
- **MICCAI 2026 Best Paper Award
Nomination**

github.com/mrgiovanni/R-Super

AI Can Help Early Tumor Detection

- Early detection drastically improves survival
 - Detected early: often >90% 5-year survival
 - Detected late: often <20% 5-year survival
- 400M CT scans/year, multiple reasons
- Tumors are often difficult to see
 - E.g., studies show that about 50% of early pancreatic tumors may be missed by radiologists
- AI can see details not perceptible to humans
- Segmentation models: detect and locate tumors
 - Interpretability and trustability
 - Surpass classification and VLMs

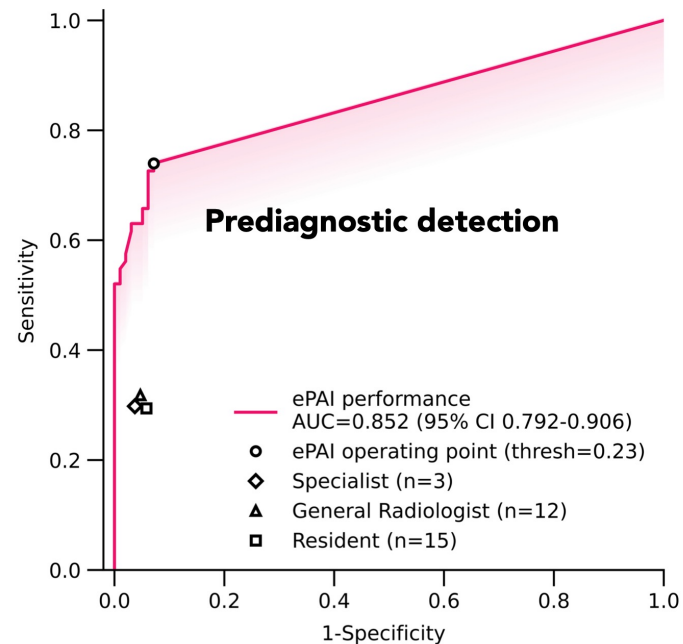
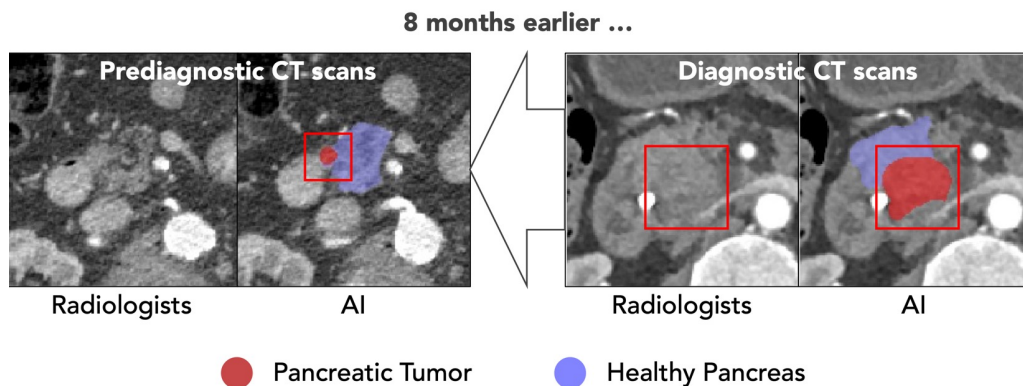
A Successful Story: Pancreatic Cancer

- Radiologists can detect pancreatic cancer from contrast-enhanced 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: Pancreatic Cancer

- 30 radiologists: 34% sensitivity 94% specificity
- AI: obtained 2x sensitivity at similar specificity

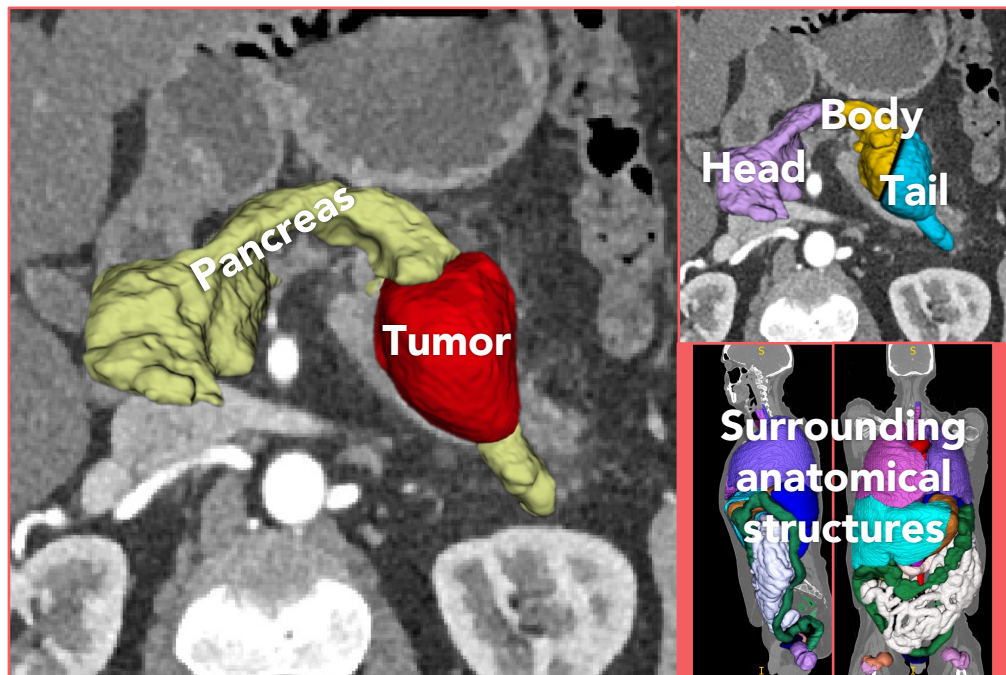




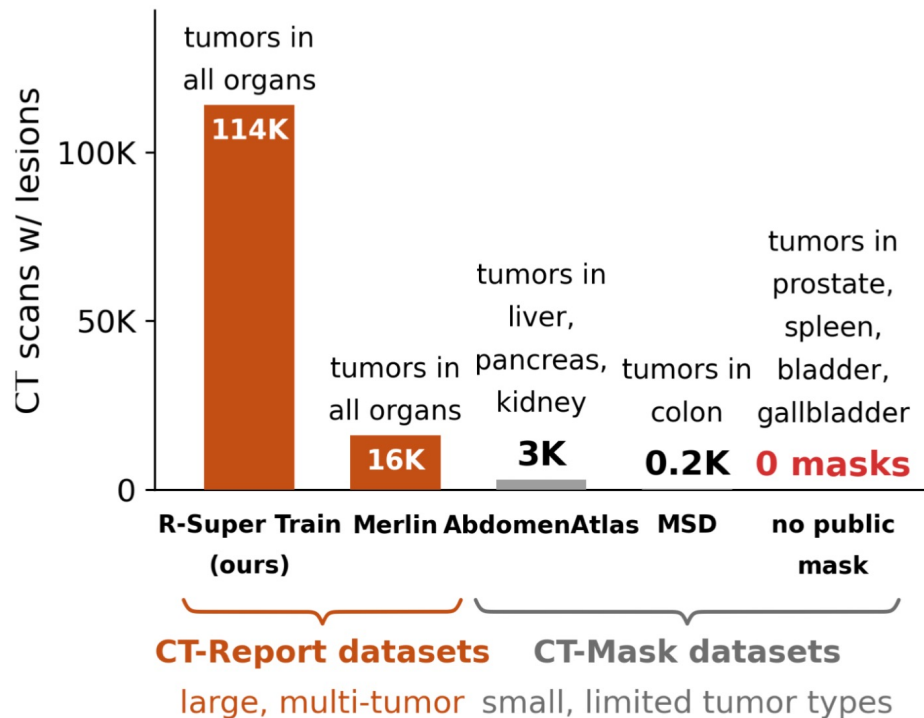
JOHNS HOPKINS
MEDICINE

Behind the Scenes

- >5,000 voxel-wise annotated CT scans at JHU, taking **25 person years**
- **\$6M**, five-year annotations
- Drawing masks is:
 - Labor-intensive
 - Costly
 - **Not** part of standard clinical workflow
 - 15-45 minutes per 3D CT
- **All for 1 tumor type, 1 hospital**

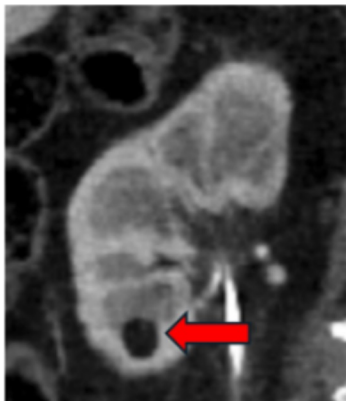


Problem: Few Per-voxel Tumor Masks



Problem: Few Per-voxel Tumor Labels

- Informative: tumor size, location, attenuation, and quantity

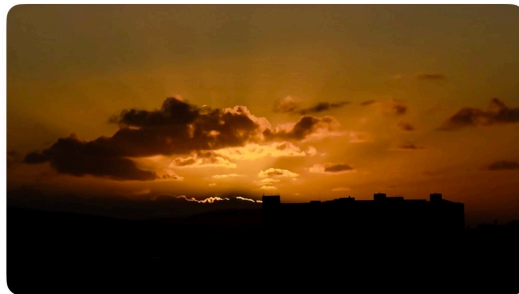


Report: [...] hypodense
cystic **lesion** in the right
kidney, measuring 11 x 10
mm [...]

≠

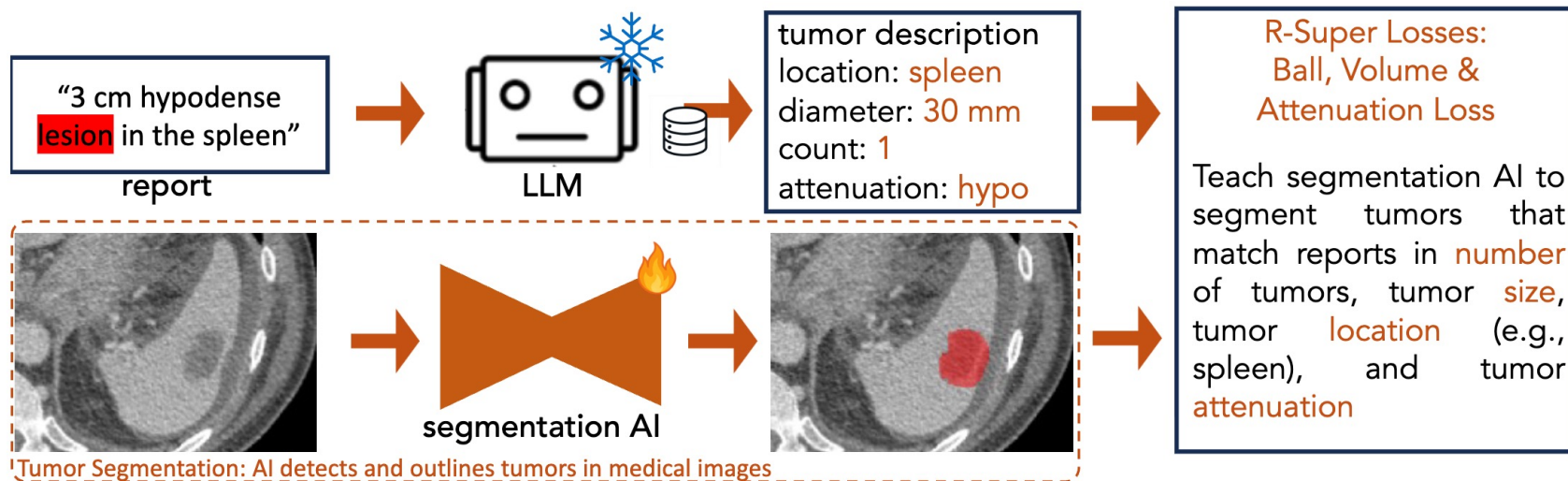


I'm using this post in my 4 PM VLM3D talk — just a quick reminder that social media captions are not always related to the images 😊



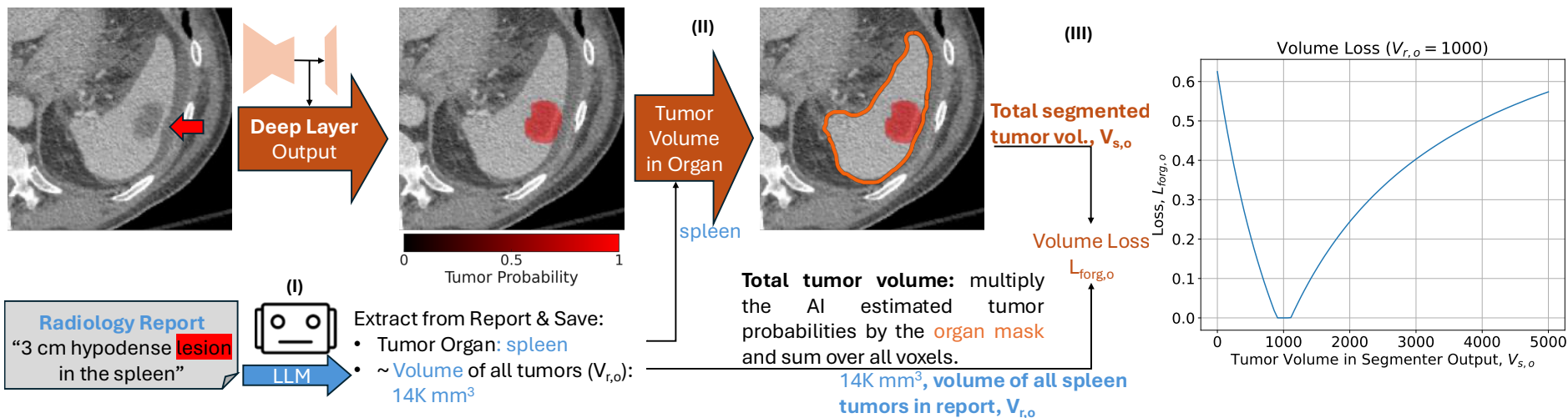
How to use this rich information to improve tumor detection AI?

Report Supervision: Overview



- Reports can supplement or substitute masks
- **Reports only used in training**
- Any architecture

Volume Loss: Matches Tumor Volume/Location

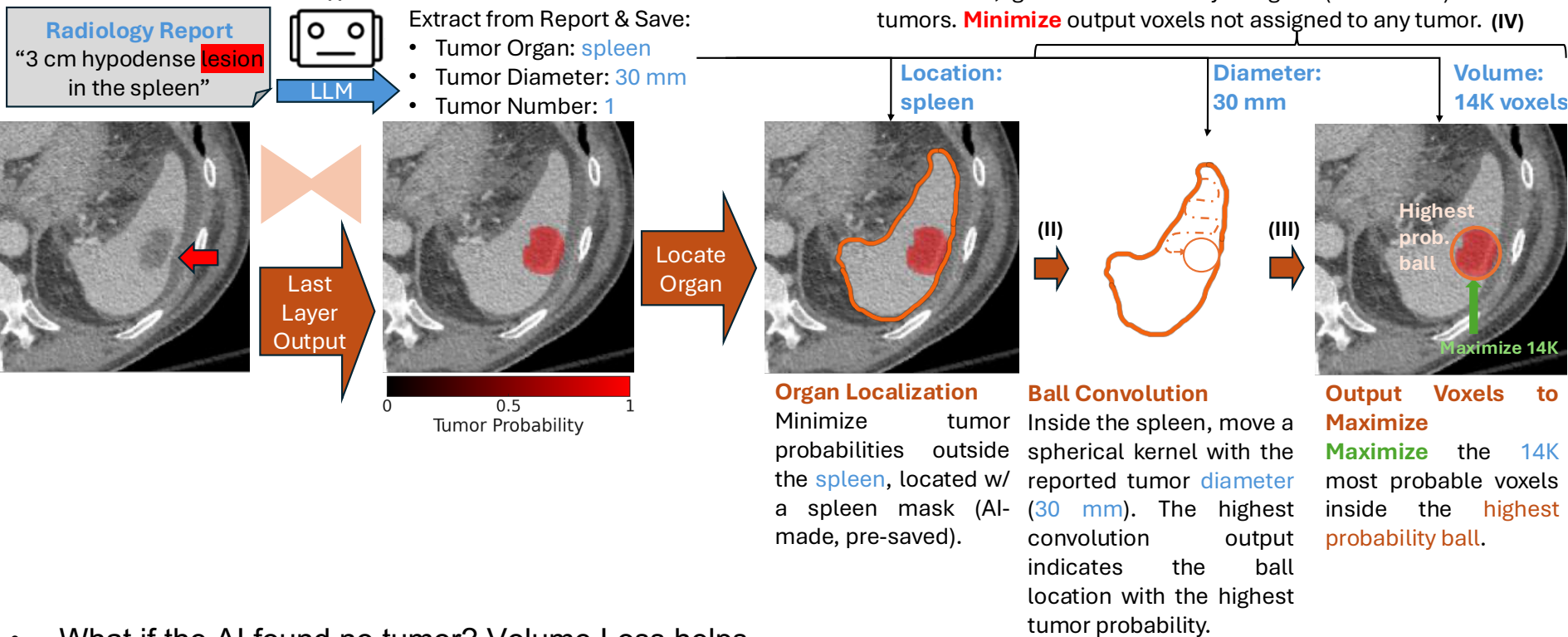


- Organ segmentation is easier than tumor segmentation: TotalSegmentator, Touchstone, AbdomenAtlas 3,...

Ball Loss: Matches Tumor Diameter/Location/Count

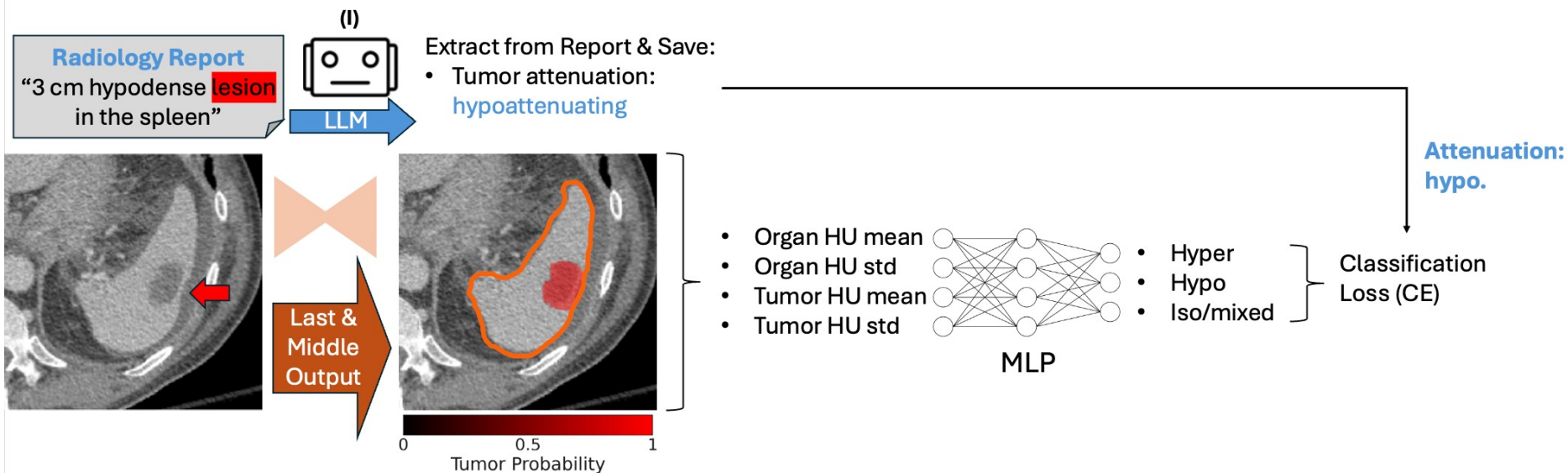
Ball loss: Which output voxels to **maximize** and **minimize**?

Repeat for all tumors **reported** in the organ, from large to small. For each tumor, ignore voxels already assigned (maximized) for other tumors. **Minimize** output voxels not assigned to any tumor. (IV)

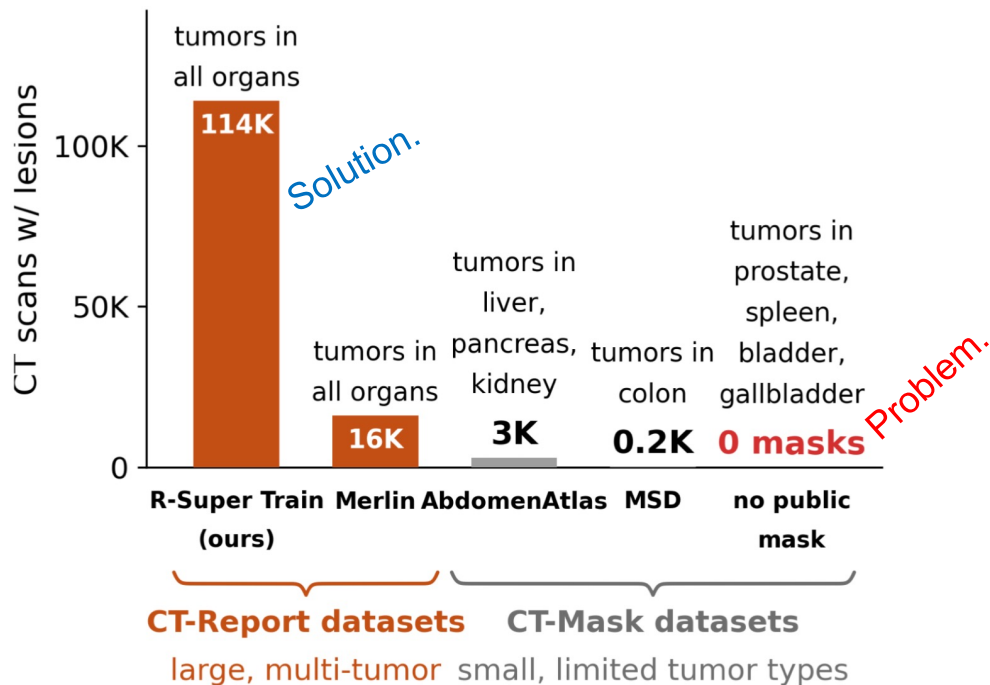


- What if the AI found no tumor? Volume Loss helps
- Can it enforce a false positive? In one sample yes, over the dataset no

Attenuation Loss

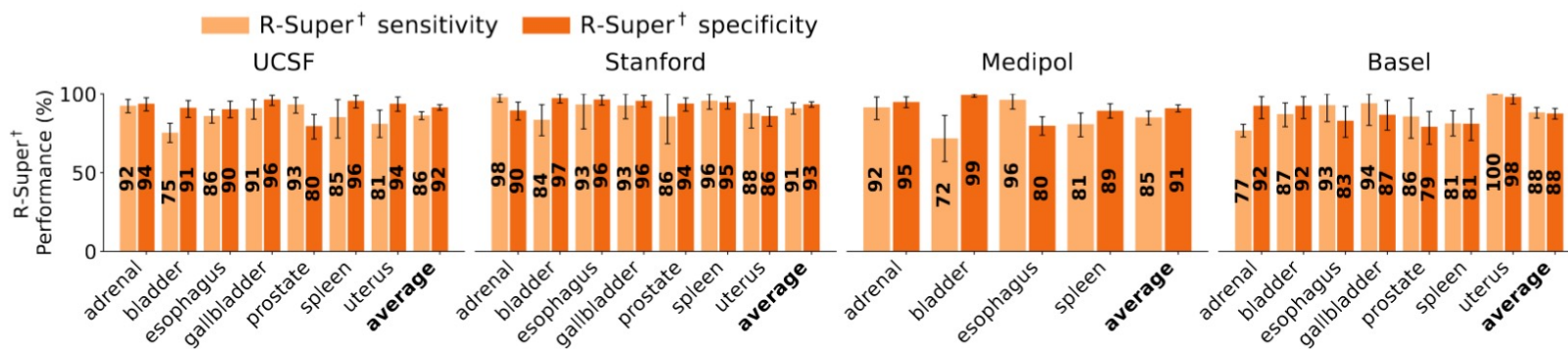
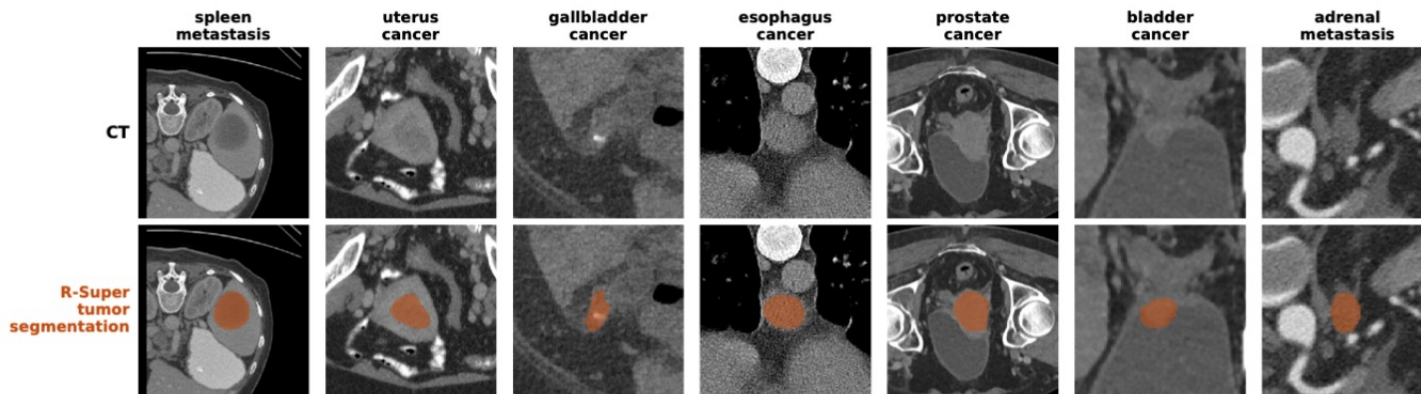


Detecting 7 Cancer Types Missing Masks

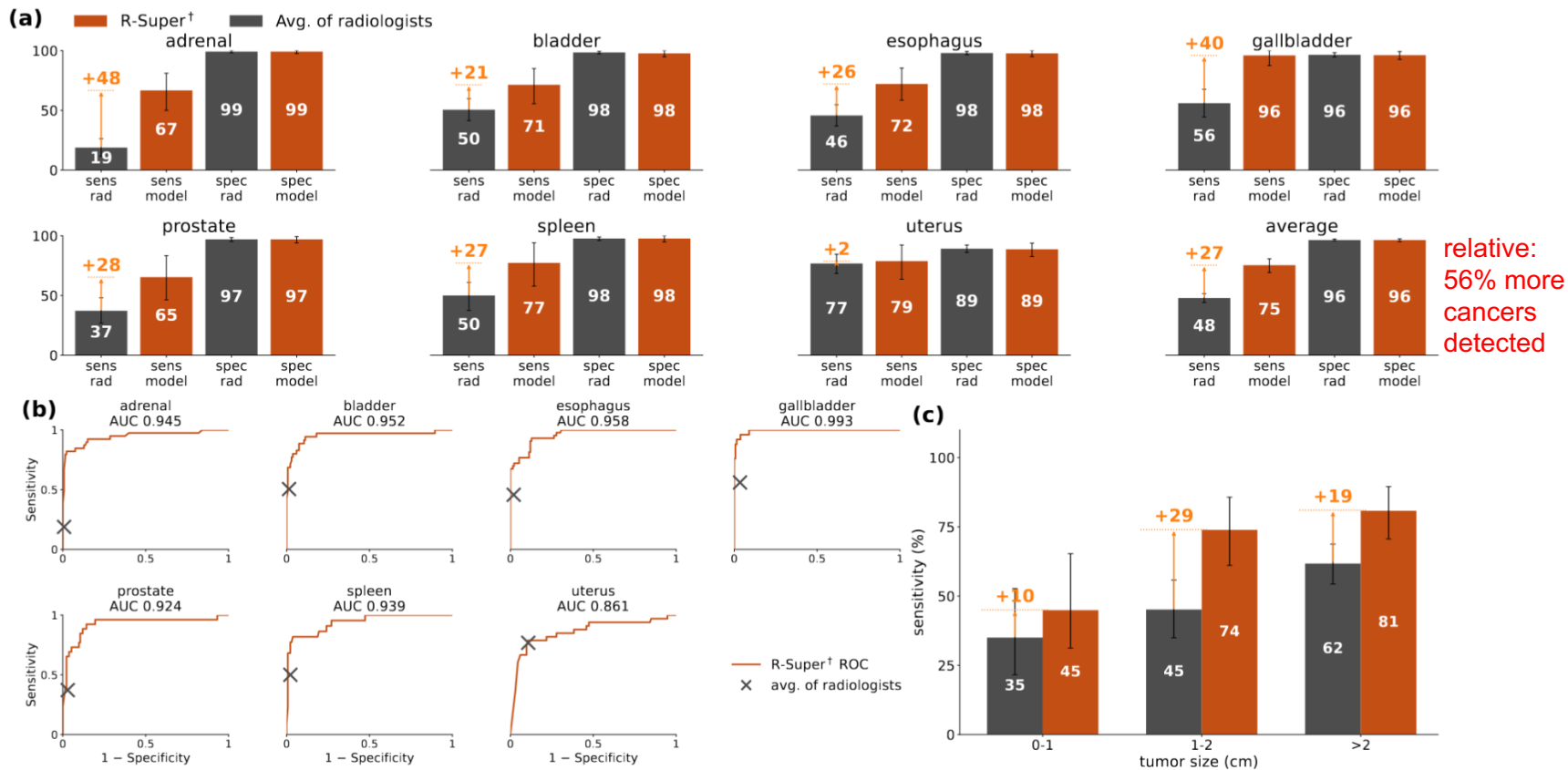


PACS (0.5M reports, 30 years) > LLM (98K tumor reports + 18K normal)

Detecting 7 Cancer Types Missing Masks

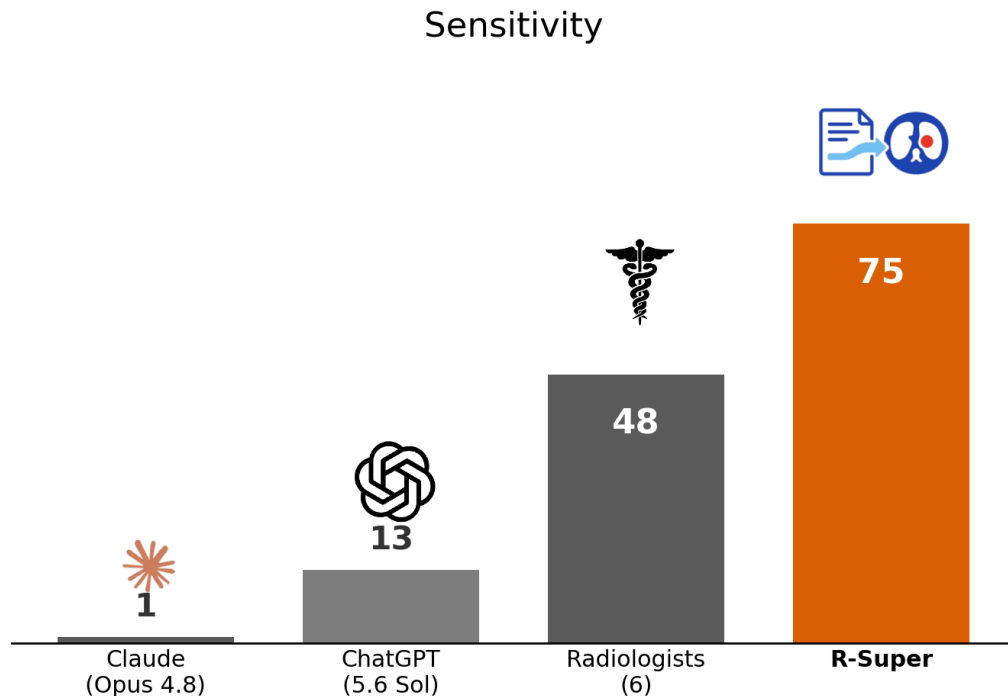


Surpassing 6 Radiologists in Detecting 7 Cancers



Surpassing Frontier VLMs

- Claude (Opus 4.8):
1% sensitivity, 99% specificity
- ChatGPT (5.6 Sol):
13% sensitivity, 96% specificity
- Radiologists (6):
48% sensitivity, 96% specificity
- **R-Super:**
75% sensitivity, 96% specificity
First AI above human performance



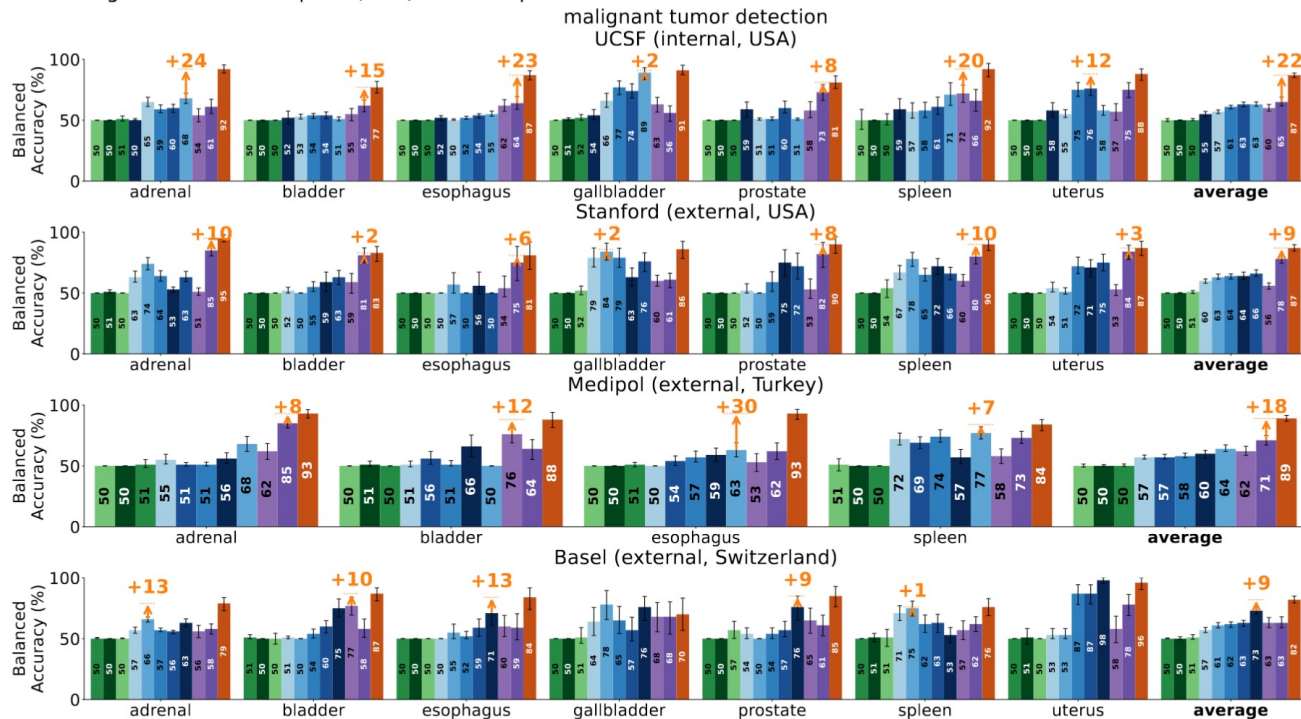
Surpassing Medical VLMs & Segmentation

VLM zero-shot classification: MedImageInsight Merlin

VLM report generation: Lingshu Merlin MedGemma

segmentation from masks: ULS FLARE23 Top 1 FLARE24 Top 1 FLARE25 Baseline FLARE23 Top 2

segmentation from reports (ours): R-Super

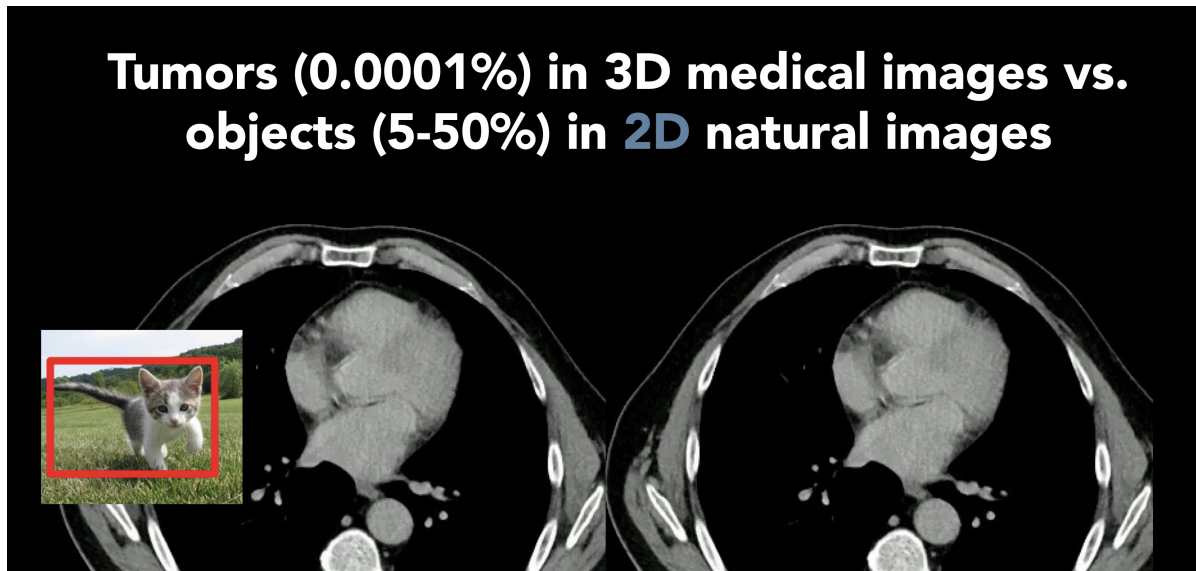


Alibaba

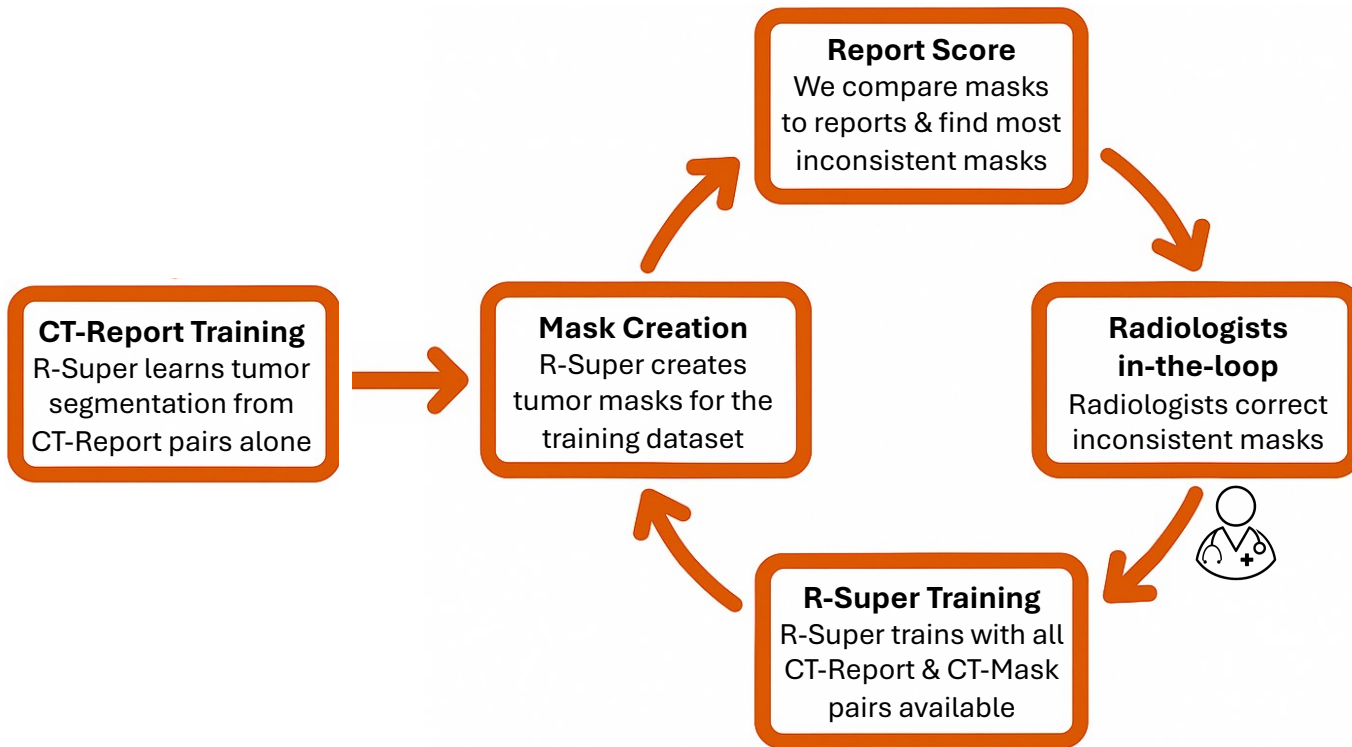


Why does R-Super surpass VLMs?

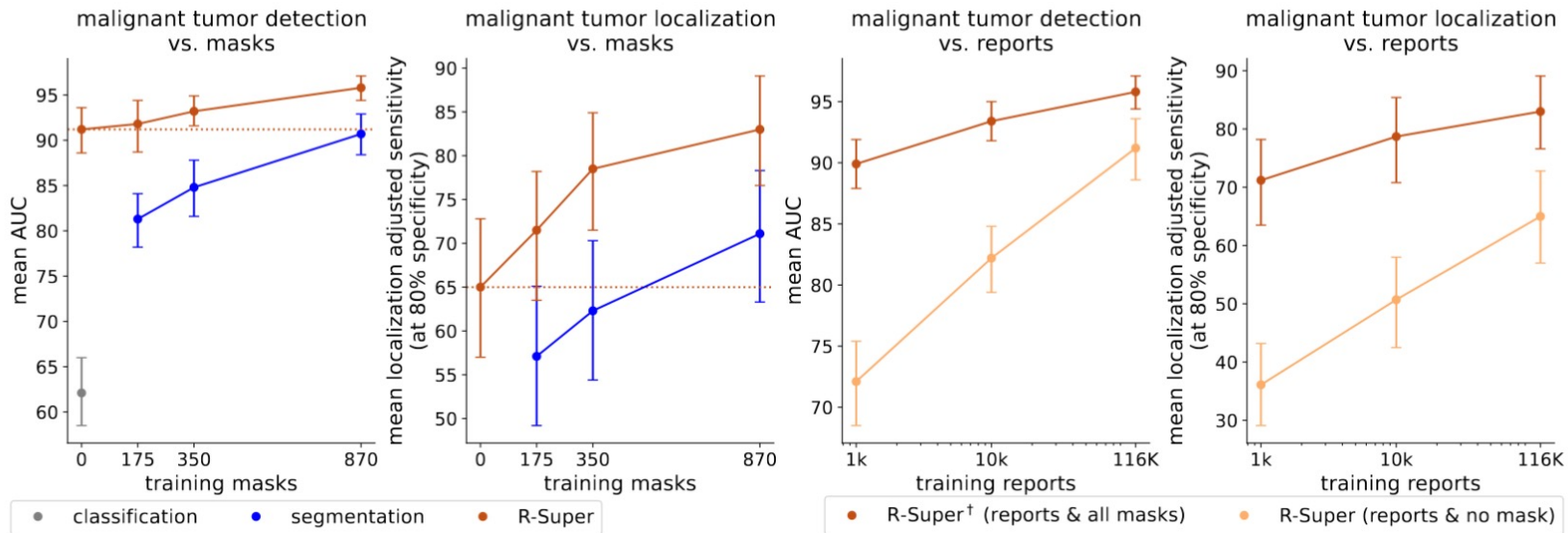
- Data? No, R-Super surpasses VLMs trained on the same data (+17% bal. acc. over Merlin)
- Reports are not usual captions
- CLIP: aligns whole image embedding to text
- Segmentation: localized supervision
- R-Super: reports directly supervise segmentation



Report or masks? Both?



Performance Scales with More Report or Masks



- Training with 127,000 reports surpasses training with 870 masks
- Training with reports and masks surpasses training with masks alone

DATASET OVERVIEW

Merlin Plus: A Longitudinal, Multi-Cancer CT Dataset

Image + mask + report pairs designed for scalable tumor AI

9 TUMOR ORGANS

radiologist-verified

DATASET SCALE

25,494

CT scans + reports

TUMOR ANNOTATION

1,153

per-voxel masks

CLINICAL REVIEW

4

radiologists

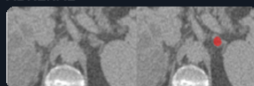
CREATION TIME

5 months

report-based active learning

TUMOR MASKS ACROSS UNDER-REPRESENTED ORGANS

ADRENAL



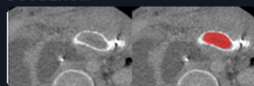
BLADDER



GALLBLADDER



DUODENUM



UTERUS



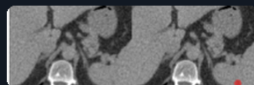
ESOPHAGUS



PROSTATE



SPLEEN



Red overlays show 3D tumor masks • examples reproduced from manuscript Fig. 1

LONGITUDINAL BY DESIGN

18,317 patients

At least 2 scans 3,830

At least 3 scans 1,524

9 CANCER / TUMOR ORGANS

MASK COUNT • TOTAL 1,153

Adrenal glands 262

Bladder 122

Prostate 80

Spleen 186

Uterus 115

Duodenum 78

Stomach 160

Gallbladder 96

Esophagus 54

DATASET OVERVIEW

CancerVerse: Longitudinal + Multimodal Multicancer Screening

A fully open dataset combining patient trajectories, clinical context, and voxel-wise tumor annotations

13 CANCER TYPES

pelvis • abdomen • chest

DATASET SCALE

24,168

CT scans

PATIENT COHORT

13,863

patients

CANCER COVERAGE

13

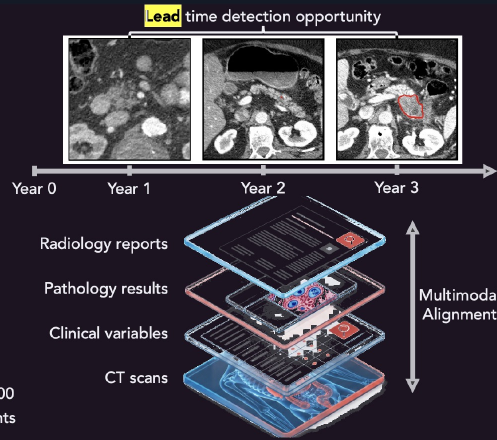
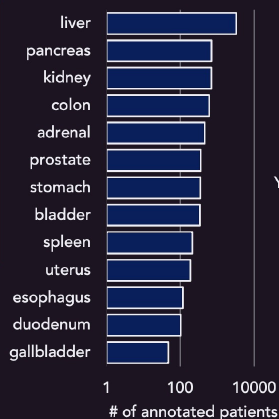
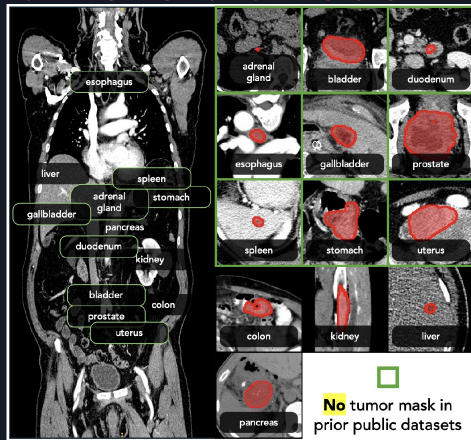
cancer types

HUMAN ANNOTATION

28 radiologists

>13 months

CANCERVERSE: MULTICANCER • LONGITUDINAL • MULTIMODAL



LONGITUDINAL COHORT

1-32 scans/patient

Median with follow-up

3 scans / 299 days

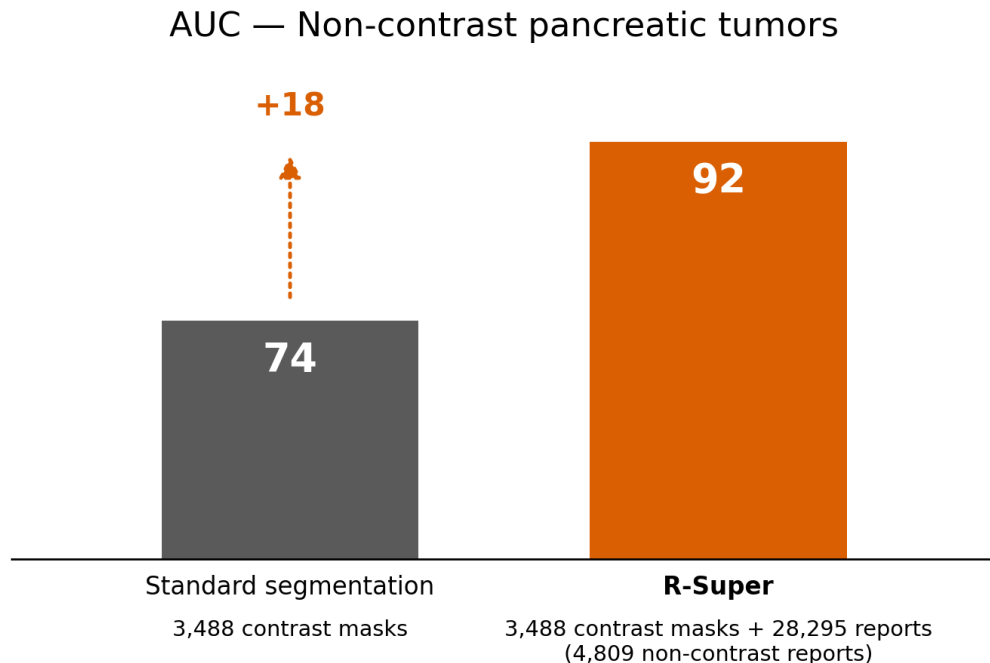
8,799

verified healthy patients

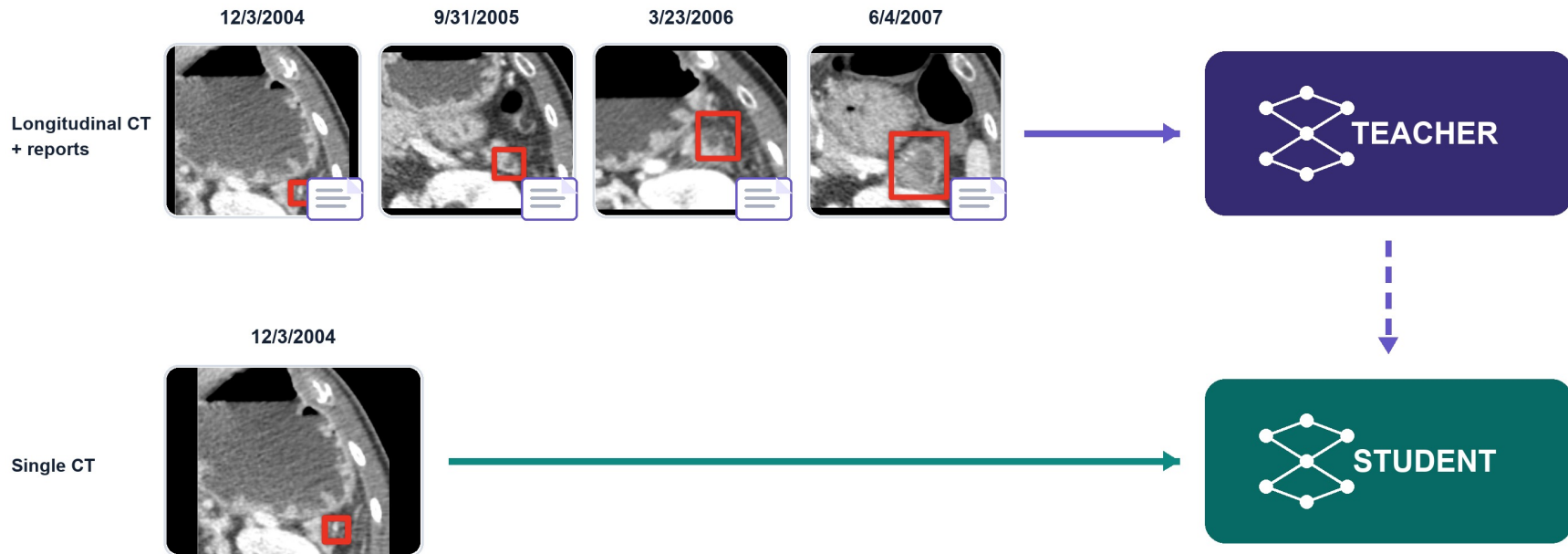
>1 year documented follow-up

Back to the Pancreas: Non-contrast CT

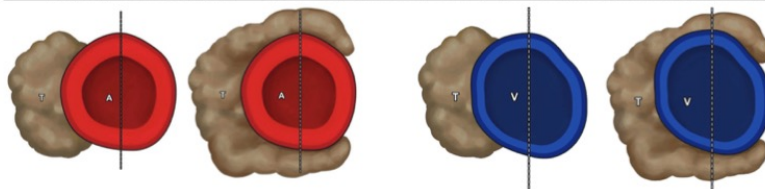
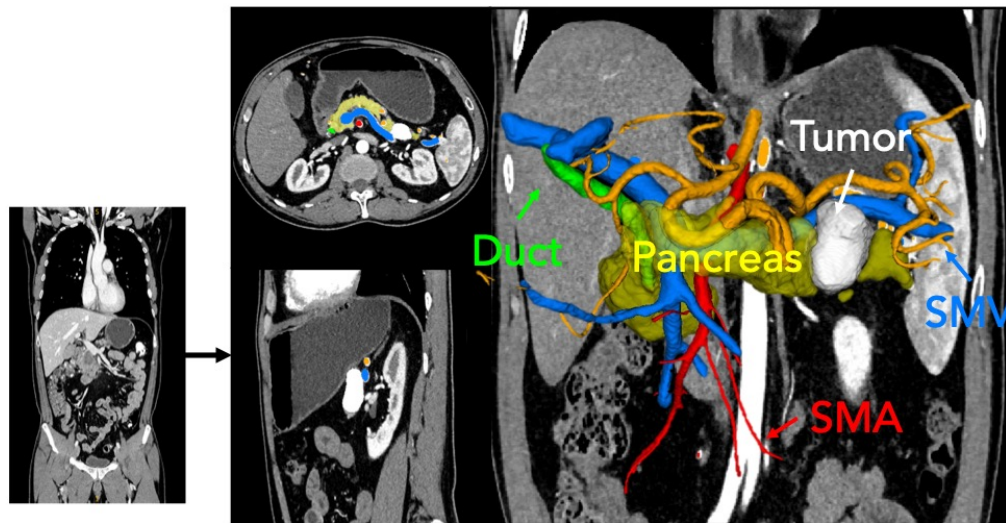
- Detecting pancreatic cancer on non-contrast CT is extremely difficult
- Important for incidental detection and screening
- Free contrast to non-contrast AI:
 - No new mask
 - No non-contrast mask



RT-Super: Learning from Longitudinal Reports



What about Report Generation?



Pancreas:

Pancreas is enlarged (volume: 84.6 cm³).
Mean HU value: 105.7 +/- 33.1.

Pancreas lesions:

Pancreas tumor 1:

Location: pancreas head/body.
Size: 2.9 x 2.2 cm (image 298).
Volume: 8.2 cm³.

Tumor Stage (T stage): T2.

Enhancement relative to pancreas:
Hypoattenuating (HU value is 52.6 +/- 26.8).



Conclusions

Can reports improve cancer detection and segmentation?

- Scalable: performance is not saturated
- Improves results for **small tumors**, for **unseen hospitals**, and for **non-contrast CT**
- Improves results with no mask, few masks (50), or many masks (3,500)
- Many Image-Report pairs create better segmentation AI than few Image-Mask pairs
- Unlocks the detection and segmentation of more tumor types
 - **First** public AI that can detect these 7 tumor types at/above human performance
 - **First** public Image-Mask pairs for multiple tumor types (CancerVerse & Merlin Plus)
- Segmentation can support report generation and VQA
 - Surpasses frontier medical and general VLMs
- Clinical implication: potential to radiologists in incidental cancer detection
 - AI flags potentially missed tumor, radiologist checks and may order specific exams or ask patient to return sooner



github.com/mrgiovanni/R-Super



Key References

- **Bassi, Pedro RAS**, ..., Alan Yuille, and Zongwei Zhou. "Learning Segmentation from Radiology Reports." *MICCAI, 2025. Best Paper Award* (2 of 1,027 papers).
- **Bassi, Pedro RAS**, ..., Alan Yuille, and Zongwei Zhou. "RadGPT: Constructing 3D Image-Text Tumor Datasets." *ICCV, 2025*.
- **Bassi, Pedro RAS**, ..., Alan Yuille, and Zongwei Zhou. "Report Supervision." *Medical Image Analysis, 2026*.
- **Bassi, Pedro RAS**, ..., Alan Yuille, and Zongwei Zhou. "RT-Super: Learning Tumor Segmentation from Longitudinal Images and Reports." *MICCAI, 2026*.
- **Bassi, Pedro RAS**, Wenxuan Li, ..., Alan Yuille, and Zongwei Zhou. "Merlin Plus: A Large-Scale, Multi-Cancer, Image-Mask-Report Dataset." *MICCAI, 2026*.
- **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*.
- Wenxuan Li, **Bassi, Pedro RAS**, ..., Alan Yuille, and Zongwei Zhou. "CancerVerse: A Fully Open Longitudinal and Multimodal Dataset for Multicancer Screening." *MICCAI, 2026*.