Learning Segmentation from Radiology Reports

Pedro R. A. S. Bassi
Postdoctoral Researcher
Dept. Computer Science, Johns Hopkins University
psalvad2@jh.edu

Learning Segmentation From Radiology Reports
Pedro R. A. S. Bassi, Wenxuan Li, Jieneng Chen, Zheren Zhu, Tianyu Lin,
Sergio Decherchi, Andrea Cavalli, Kang Wang, Yang Yang, Alan Yuille, Zongwei Zhou

MICCAI 2025

MICCAI 2025

github.com/mrgiovanni/R-Super

Awards:

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

Nov. 30 – Dec. 4
RSNA® 2025
Imaging the Individual

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
- 300M 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
  - E.g., Pancreatic tumors in non-contrast CT
- Segmentation models: detect and locate tumors
  - Interpretability and trustability

A Successful Story

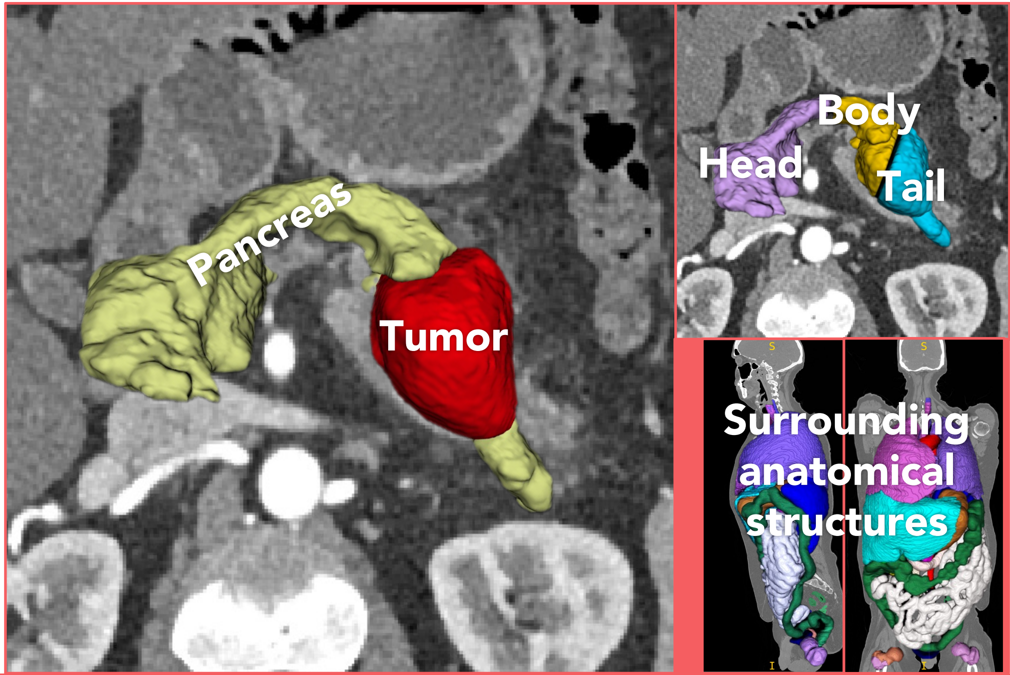
- Radiologists can detect pancreatic cancer from contrast-enhanced CT scans, but the sensitivity of early pancreatic cancer is only 33-44%
- Our AI has achieved very high performance in early detection

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

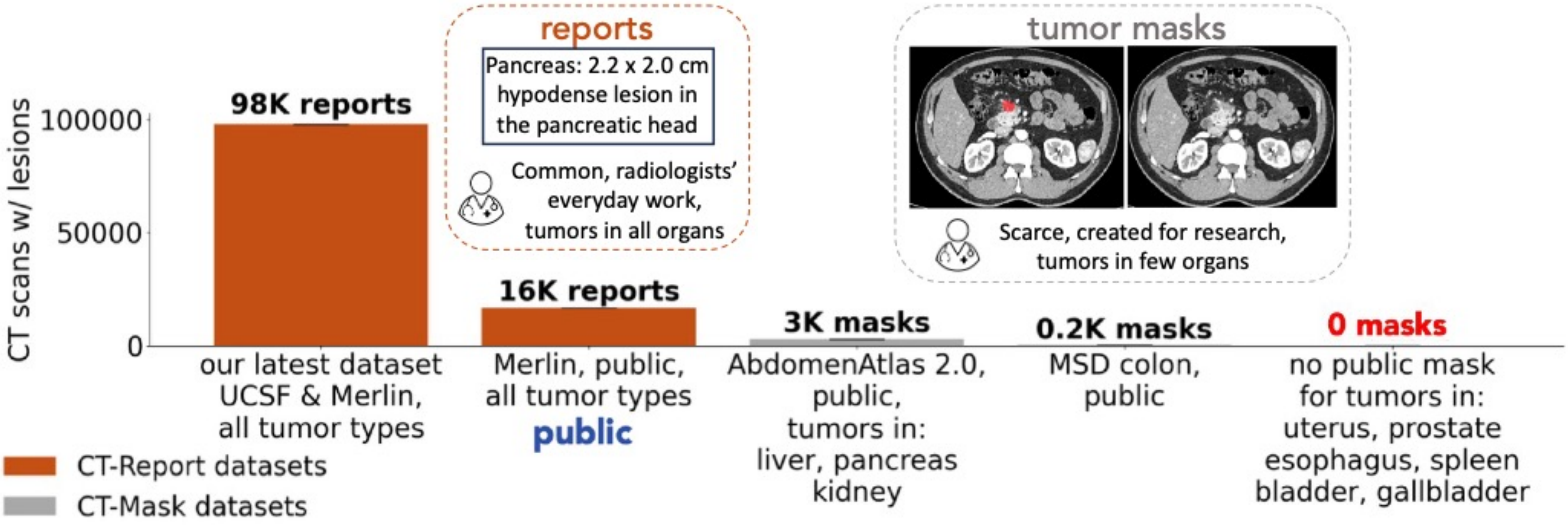
Behind the Scenes

JOHNS HOPKINS
MEDICINE

- >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
  - 15-30 minutes per 3D CT
- All for 1 tumor type, 1 hospital

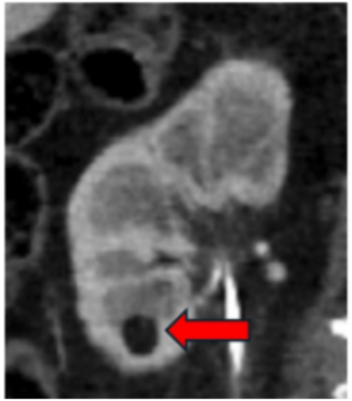


Problem: Few Per-voxel Tumor Masks



Problem: Few Per-voxel Tumor Labels

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

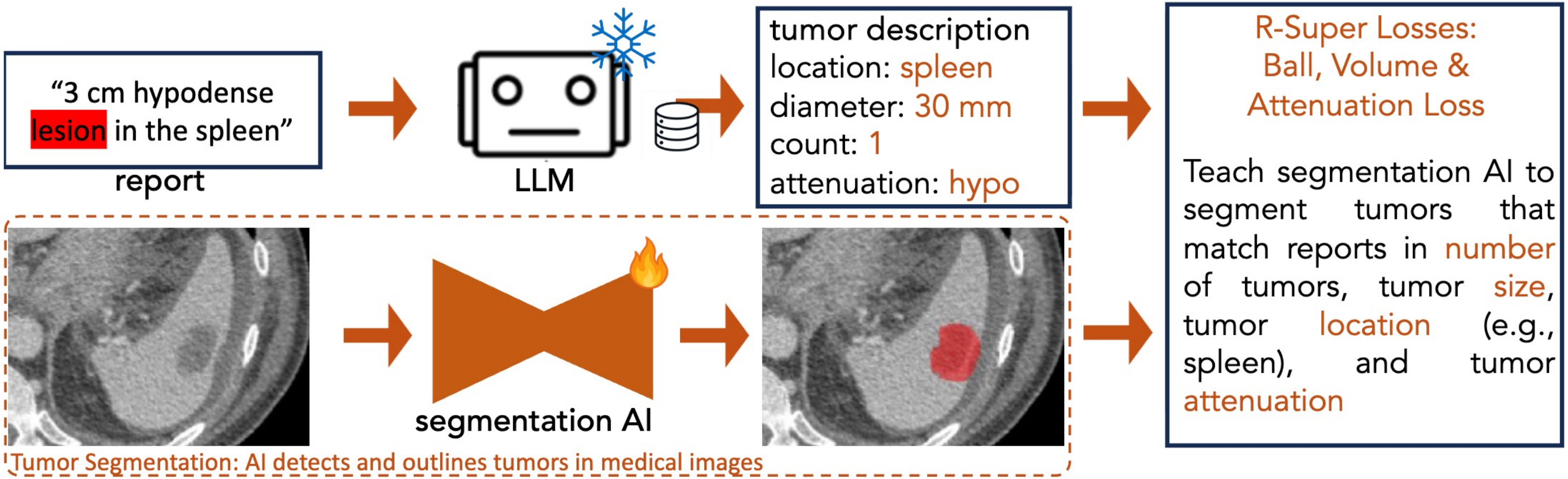


Report: [...] hypodense cystic lesion in the right kidney, measuring 11 x 10 mm [...]

≠

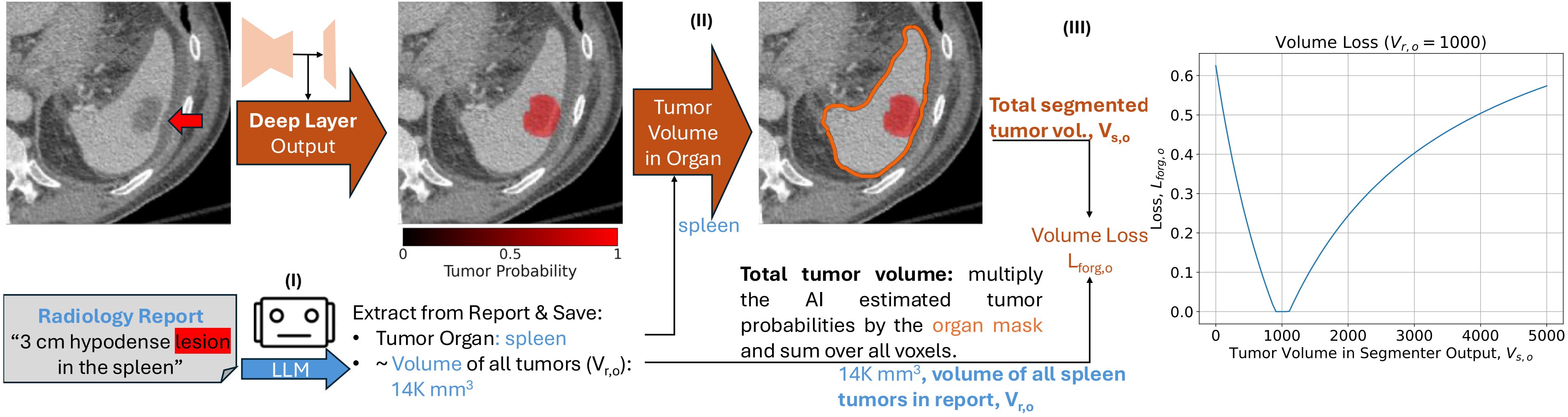
How to use this rich information to improve tumor segmentation AI?

Report Supervision: Overview



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

Volume Loss: Matches Tumor Volume/Location

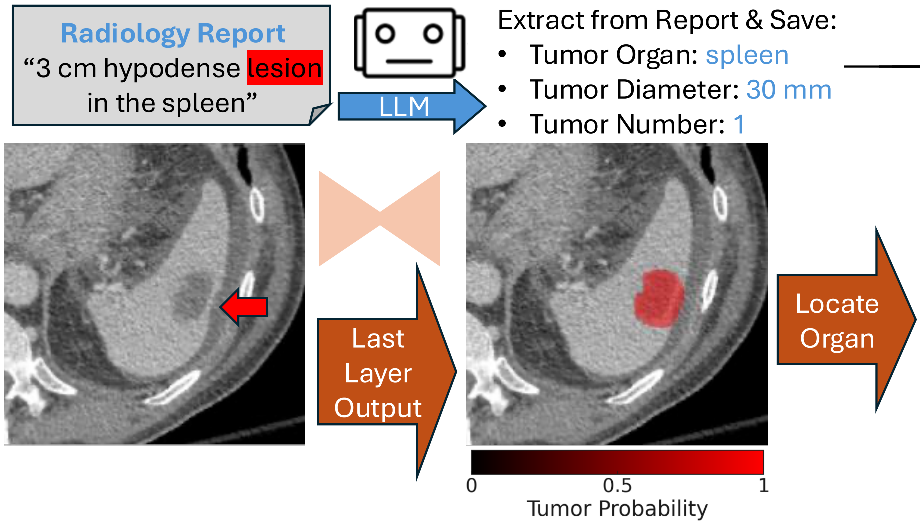


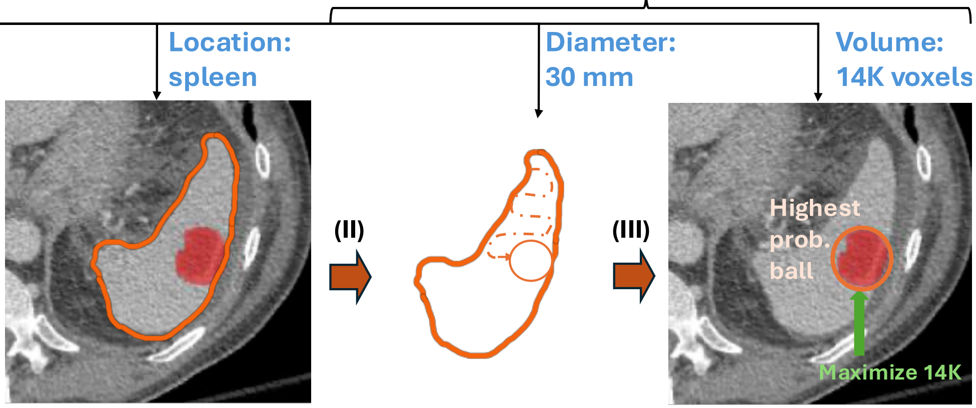
- Organ segmentation is easier than tumor segmentation: TotalSegmentator, Touchstone, AbdomenAtlas,…

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)**





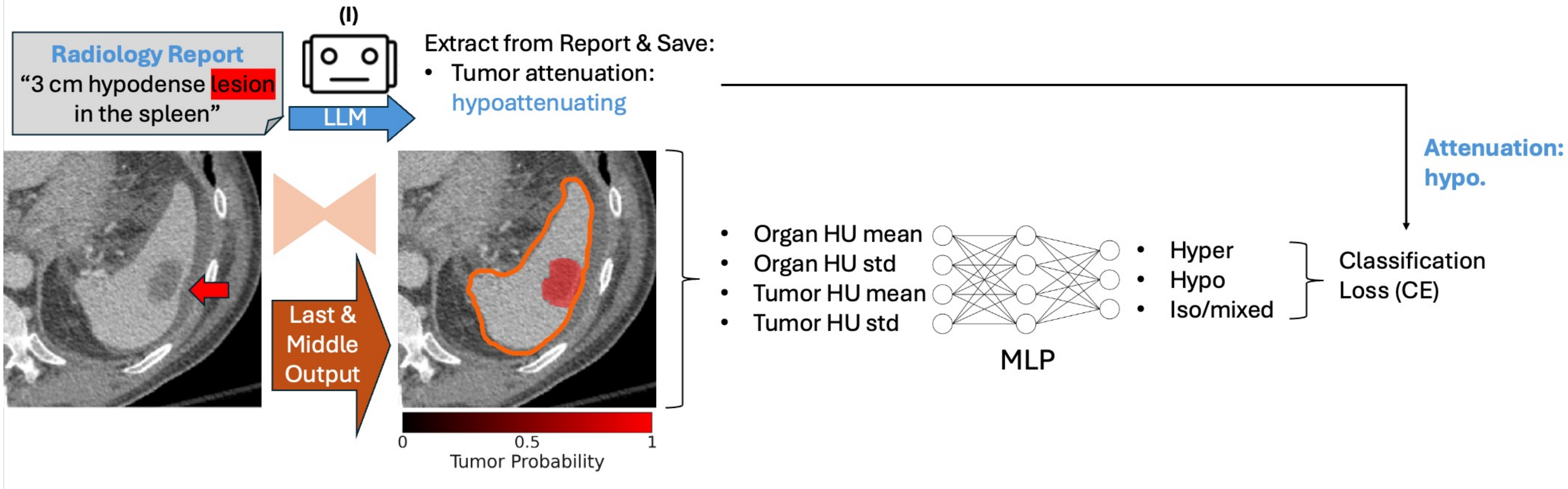
Organ Localization
Minimize tumor probabilities outside the spleen, located w/ a spleen mask (AI-made, pre-saved).

Ball Convolution
Inside the spleen, move a spherical kernel with the reported tumor diameter (30 mm). The highest convolution output indicates the ball location with the highest tumor probability.

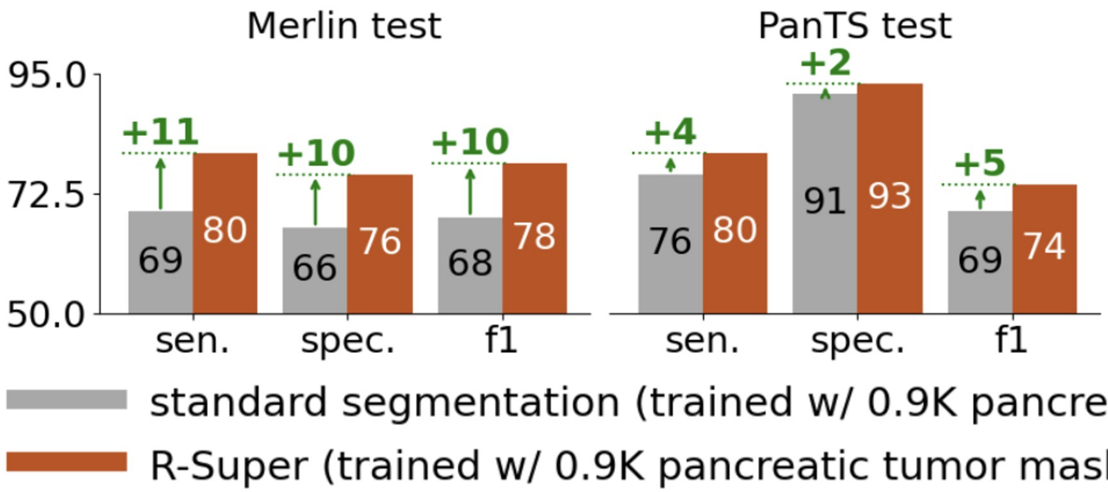
Output Voxels to Maximize
Maximize the 14K most probable voxels inside the highest probability ball.

- 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



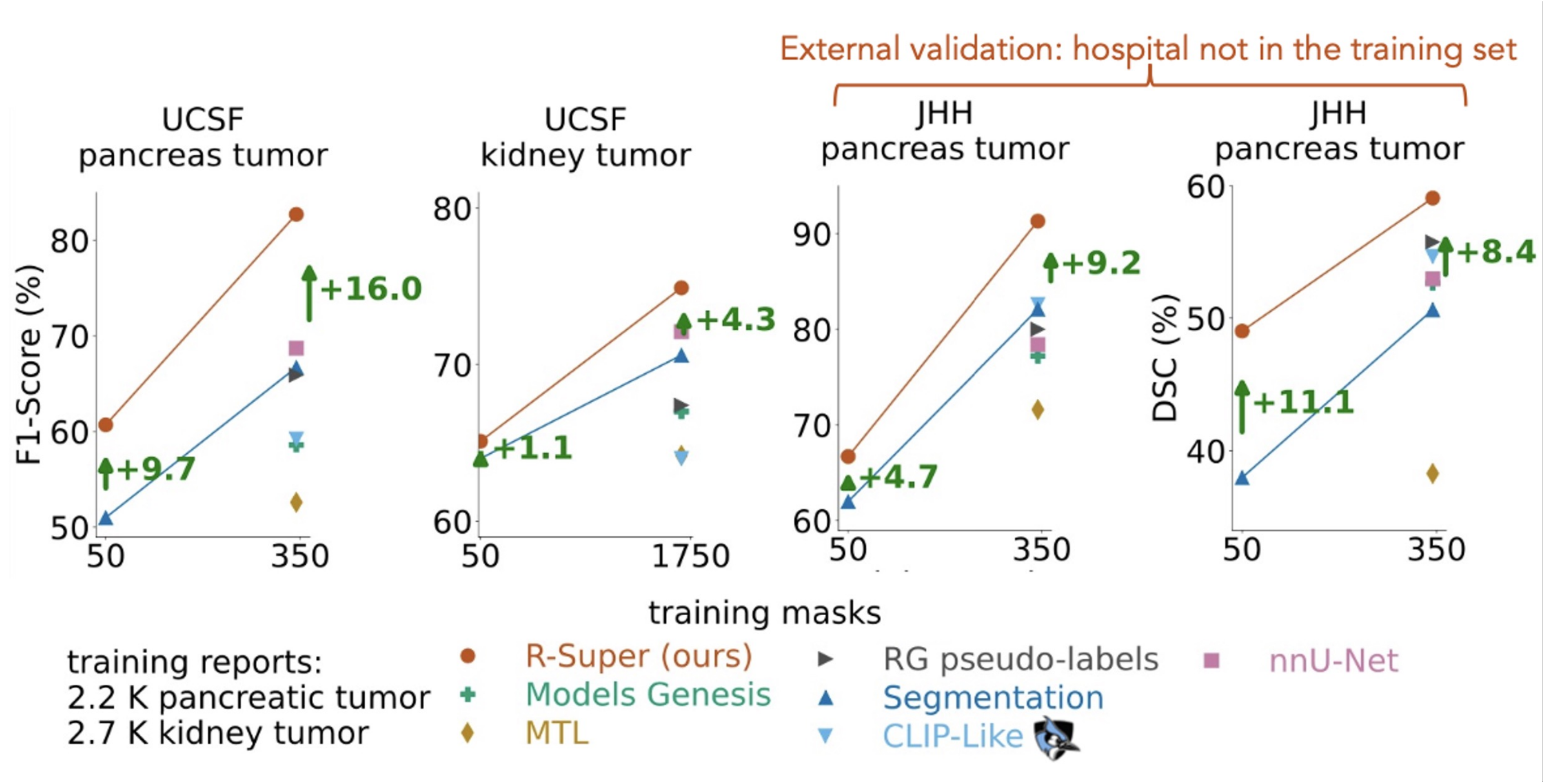
Results on Public Data: By Learning from Public Masks (N=1.1K) & Reports (N=1.8K), R-Super Improves Pancreatic Tumor Detection



Public Datasets:

- Merlin: 25K CT-Reports, 16K with lesions, 2K pancreatic.
- AbdomenAtlas 2.0: 9K CT-Masks, 3K with lesions, 344 pancreatic.
- PanTS: 9K CT-Masks, 1.1K with pancreatic lesions.

Results: Reports help with few / many masks



Why are R-Super results better?

training paradigm	learns from CT w/o mask	learns w/ reports	uses detailed report info.†	reports optimize segmentation directly	reports penalize FP/FN directly
Segmentation [14,11]	✗	✗	✗	✗	✗
Models Gen. [29,30]	✓	✗	✗	✗	✗
Multi-task Learn. [7]	✓	✓	✗	✗	✗
RG pseudo-labels [6]	✓	✓	✗	✓	✗
CLIP-Like [5]	✓	✓	✓	✗	✗
R-Super (ours)	✓	✓	✓	✓	✓

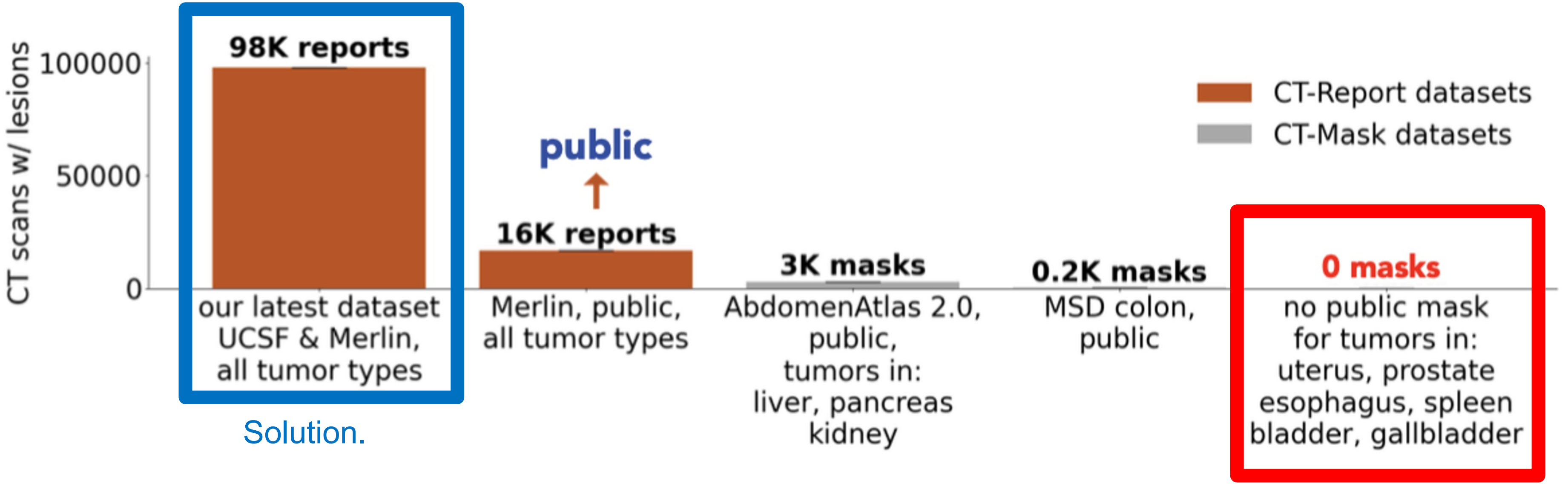
† AI learns from tumor count, sizes, and locations (organ/organ sub-segment) in reports.

Early Detection: R-Super Improves the Detection of Small (<2 cm) Tumors

	pancreas tumor												kidney tumor					
	JHH-Test								UCSF-Test				UCSF-Test					
train paradigm	mask	rep.	dsc	nsd	F1	AUC	Se	Sp	F1	AUC	Se	Sp	mask	rep.	F1	AUC	Se	Sp
CLIP-Like [5]	344	2.2K	11	19	68	90	100	71	50	74	54	75	1.7K	2.7K	40	65	57	48
Multi-task l. [7]	344	2.2K	15	26	54	83	87	60	42	61	60	50	1.7K	2.7K	46	65	57	63
RG Pseudo-l. [6]	344	2.2K	19	32	61	77	73	80	63	82	62	86	1.7K	2.7K	50	71	66	60
Models G. [29]	344	0	10	20	62	85	80	76	48	70	60	63	1.7K	0	43	65	60	54
nnU-Net [14]	344	0	7	17	55	74	60	82	60	78	70	74	1.7K	0	53	71	79	53
segmentation [11]	344	0	17	36	65	83	67	88	59	77	58	85	1.7K	0	39	69	44	68
R-Super (our)	344	2.2K	25	48	80	89	93	88	75	90	77	89	1.7K	2.7K	56	78	69	69

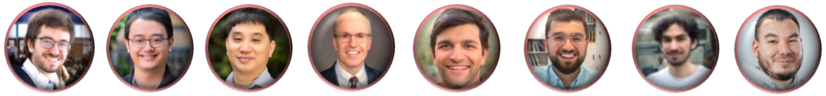
+26% / +15% / +8% in sensitivity / F1-Score / DSC for pancreatic tumors (external valid.)

Extension: Unlocking New Tumor Types



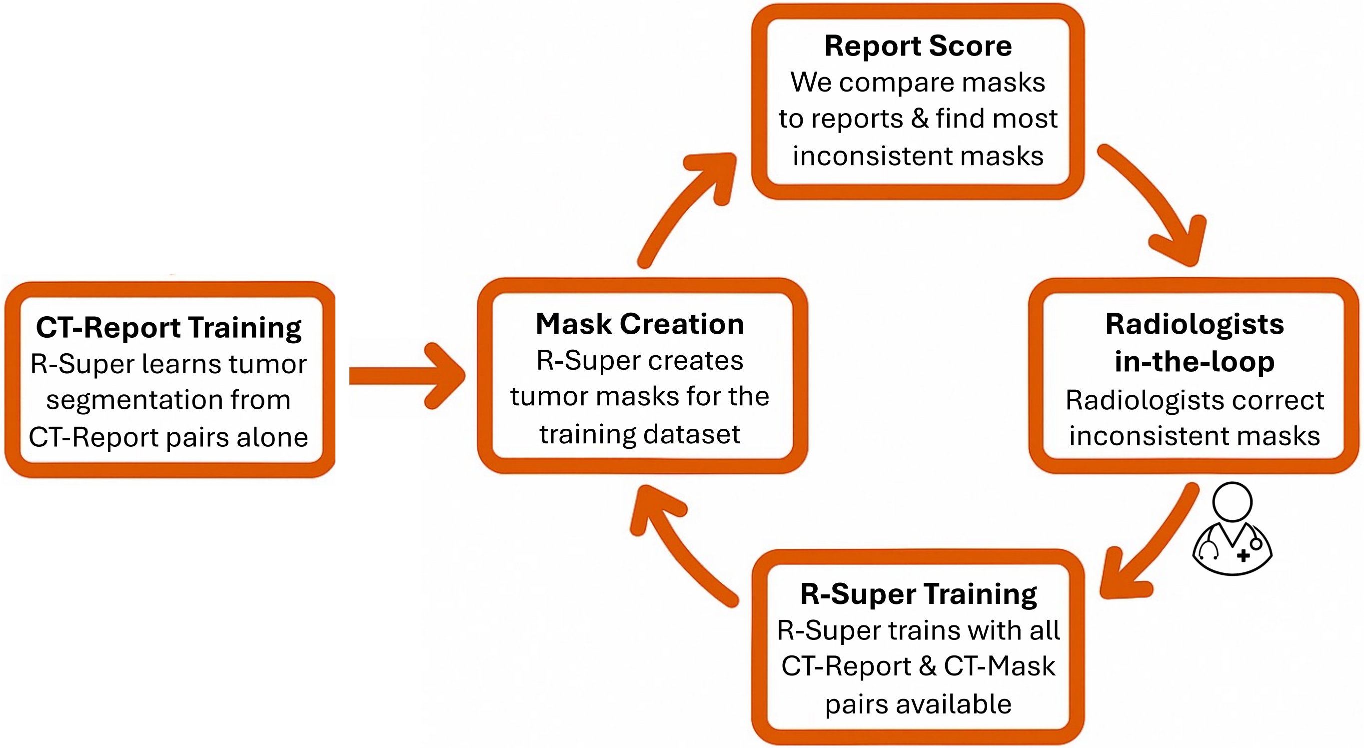
Solution.

Problem.

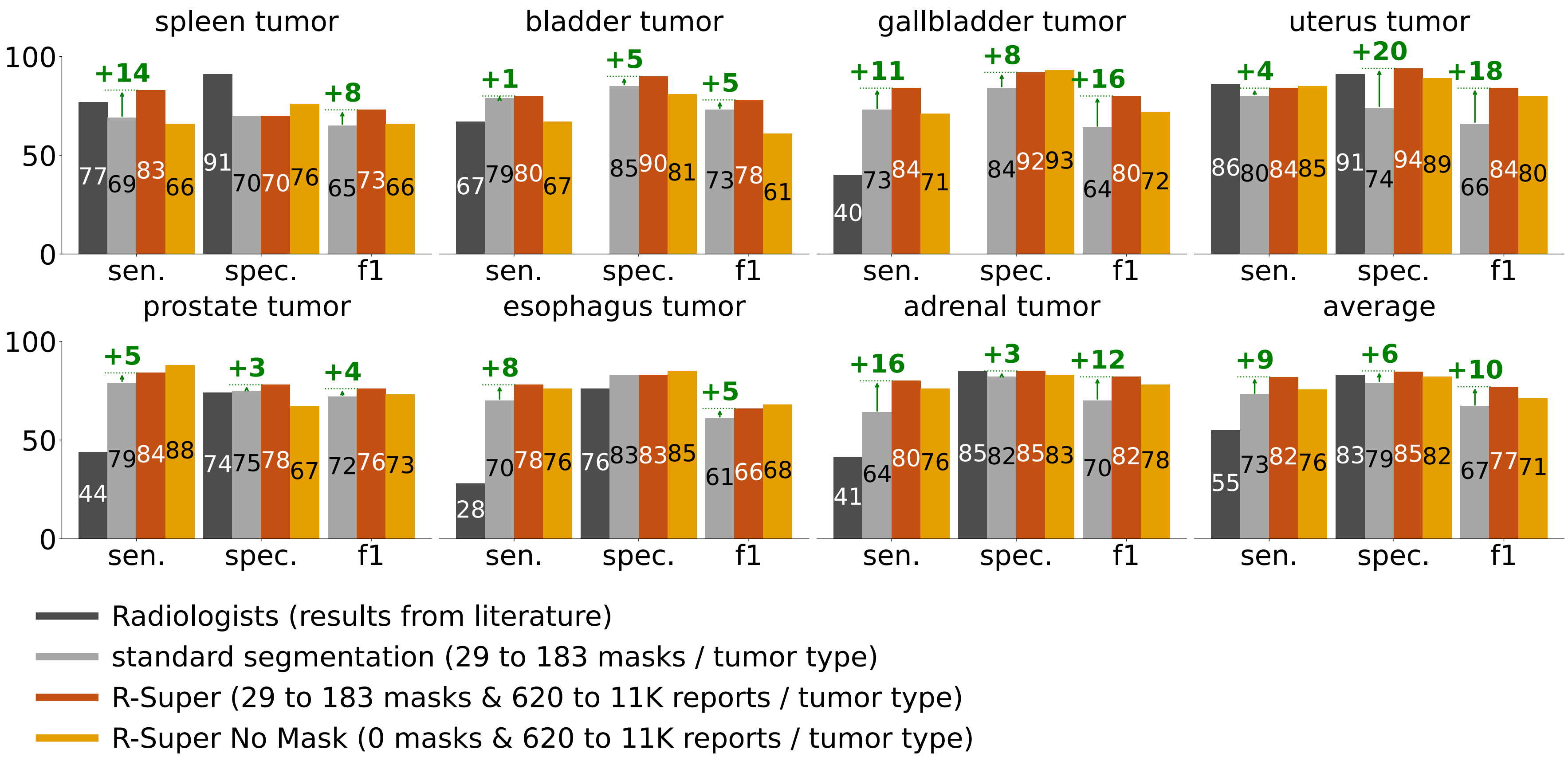


PACS (0.5M reports, 30 years) > LLM (98K tumor reports + 18K normal)

Report-based Active Learning

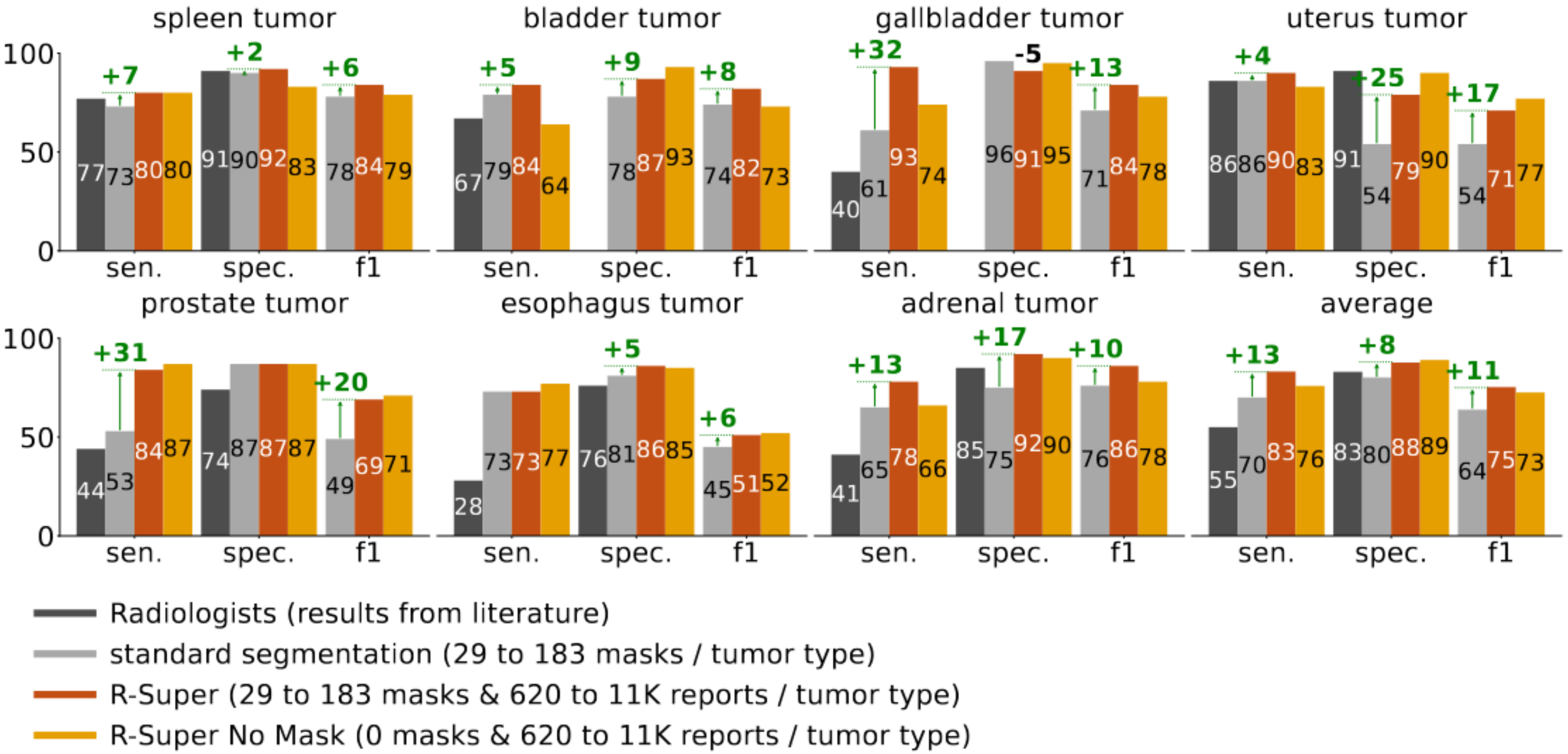


Current Work: 117K CT-Report, 7 New Tumors



- SOTA Report Generation VLMs (e.g., MedGemma, Google): <2% average sensitivity

External Validation



Conclusions

Can reports improve segmentation performance?

- Improved performance with 0, few (50), medium (344) and many (1,674) masks
- Improves results for **small tumors** and for **unseen hospitals**
- Many Image-Report pairs create better segmentation AI than few Image-Mask pairs
- Unlocks the segmentation of more tumor types
- **First** public AI that can segment these 7 tumor types
- **First** public Image-Mask pairs for these 7 tumor types

github.com/mrgiovanni/R-Super

REPORT SUPERVISION