Learning Segmentation from Radiology Reports

Pedro R. A. S. Bassi[1,2,3], Wenxuan Li[1], Jieneng Chen[1], Zheren Zhu[4,5], Tianyu Lin[1], Sergio Decherchi[3], Andrea Cavalli[2,3,6], Kang Wang[4], Yang Yang[4], Alan L. Yuille[1], and Zongwei Zhou[1,]✉

[1]John Hopkins University [2]University of Bologna
[3]Italian Institute of Technology [4]University of California, San Francisco
[5]University of California, Berkeley [6]École Polytechnique Fédérale de Lausanne

zzhou82@jh.edu; Code, Data, AI (QR Code): github.com/mrgiovanni/R-Super

MICCAI 2025 Daejeon REPUBLIC OF KOREA

Shortlisted for Best Paper and Young Scientist Awards

New Loss Functions Transform Radiology Reports into Per-Voxel Supervision for Tumor Segmentation

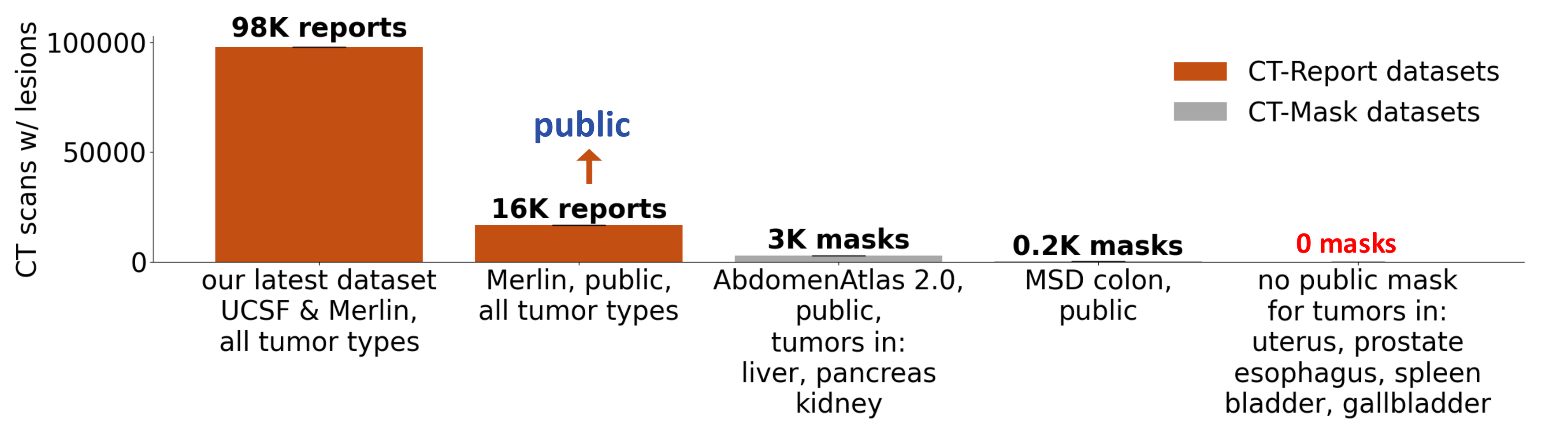
Background: Public datasets have few tumor *Image-Mask* pairs, often only tens to hundreds. By contrast, reports are written every day by radiologists—hospitals and new public datasets have more than 500K *Image-Report* pairs. Our goal is to enable AI to learn tumor segmentation directly from these reports.

Contribution: R-Super, a novel AI training method that enforce the consistency between AI segmented tumors and report descriptions such as tumor **number, size, and location**. It can train any AI architecture with *Image-Report* or *Image-Report-Mask* pairs, scaling existing *Image-Mask* datasets into much larger ones with enormous *Image-Report* pairs.

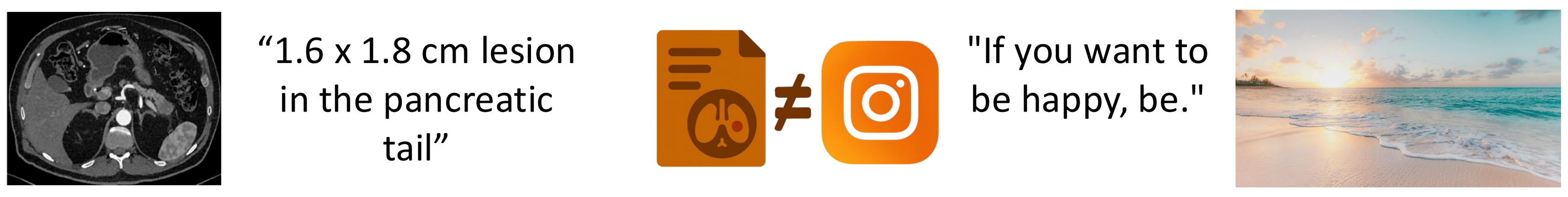
Results: Our *Image-Report-Mask* training improves **+16%/+11%** F1-Score/DSC compared with conventional *Image-Mask* training. The benefit of incorporating reports in training is significantly large both when the training dataset has as few as 50 *Images-Mask* pairs or as many as 1,700. See our GitHub (QR) for code, models, and a public-data demo.

Current & future work: We have curated a dataset of **117,000** *Image–Report* and **270** *CT-Mask* pairs for tumors in the adrenal, bladder, esophagus, gallbladder, prostate, spleen, and uterus. No publicly available *Image–Mask* pairs exist for these tumor types. With the help of R-Super, we release the **first** AI model that can segment these tumors.

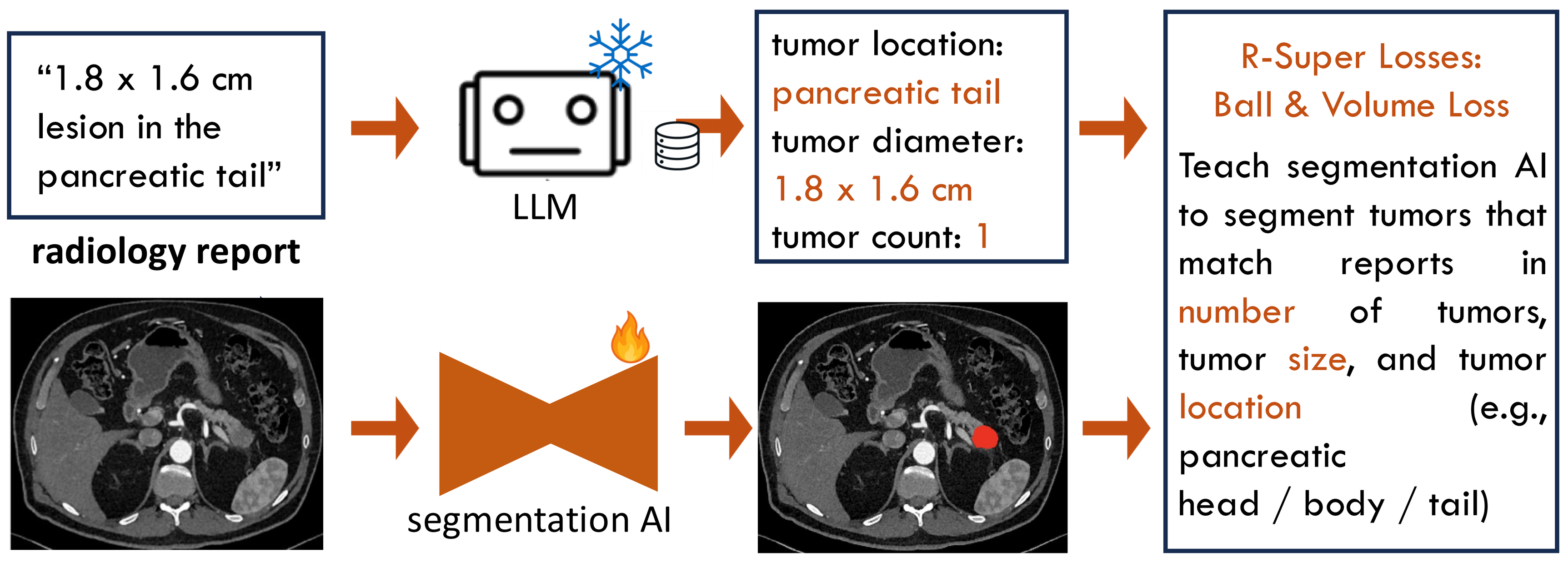
Background: Few Tumor Segmentation Masks, Many Reports



Reports ≠ Image Captions: Detailed Spatial & Lesion Information

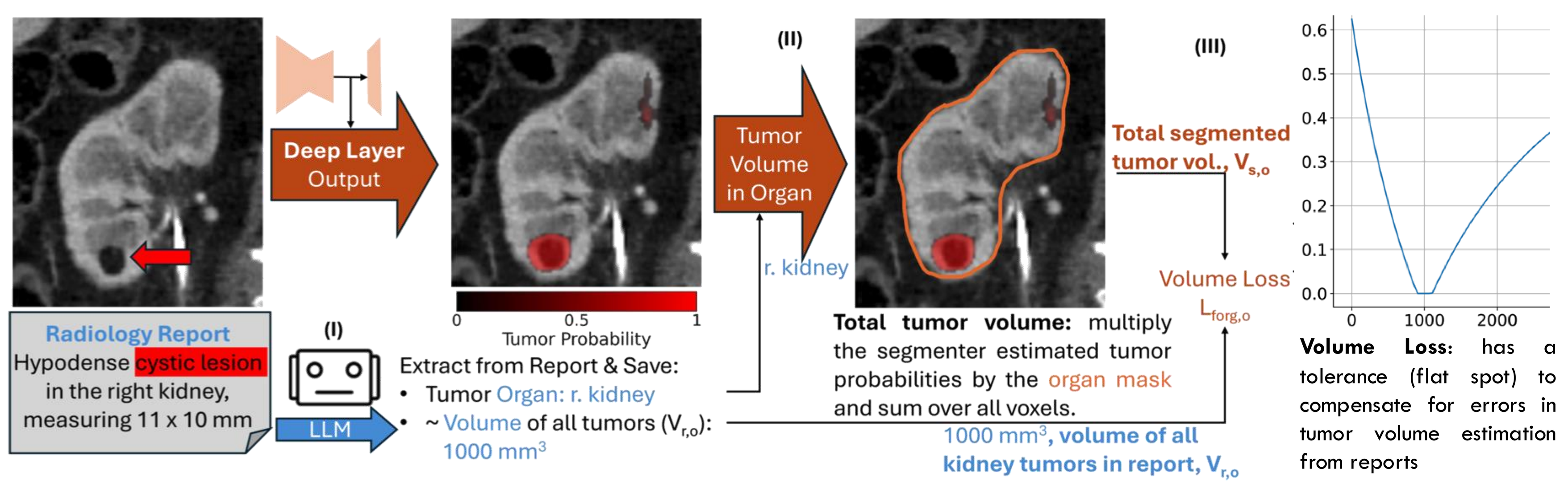


Overview: Reports Train Tumor Segmentation AI



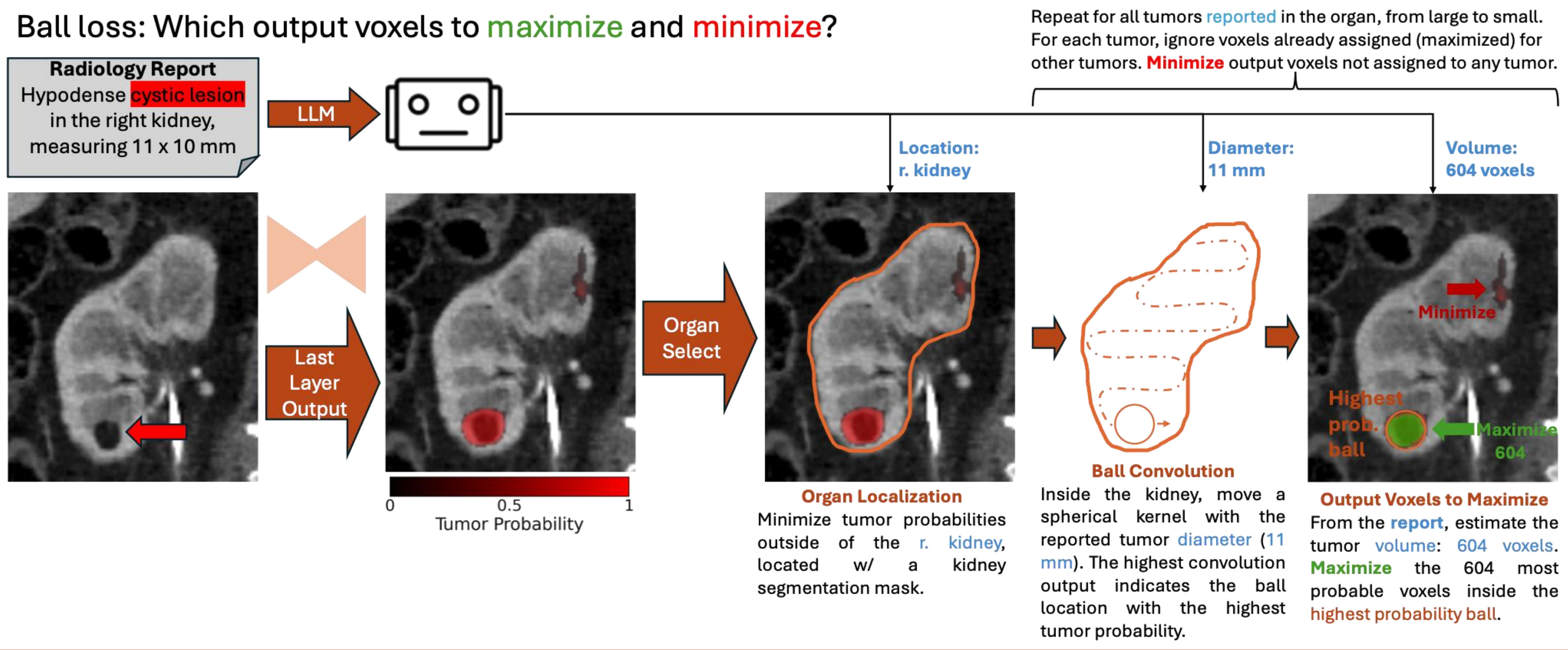
New Loss Functions — The Core of R-Super

I. The Volume Loss Matches Tumor Volume/Location in Reports

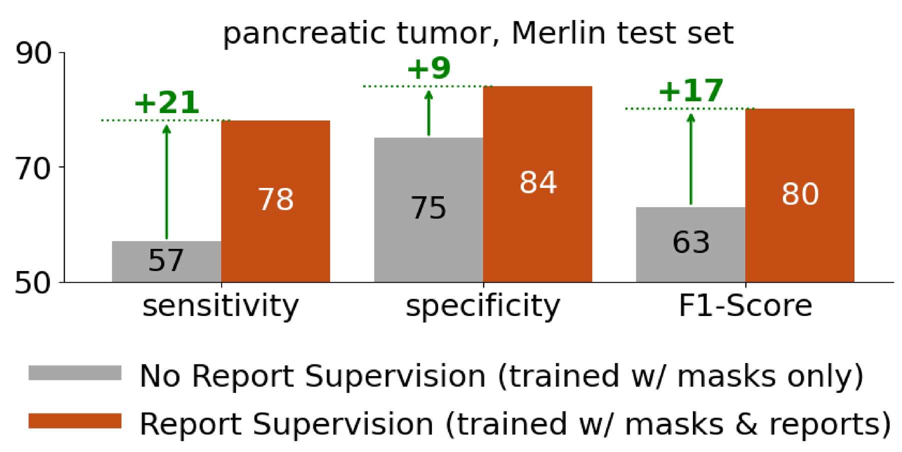


II. The Ball Loss Matches Tumor Number/Diameters/Locations in Reports

Ball loss: Which output voxels to maximize and minimize?

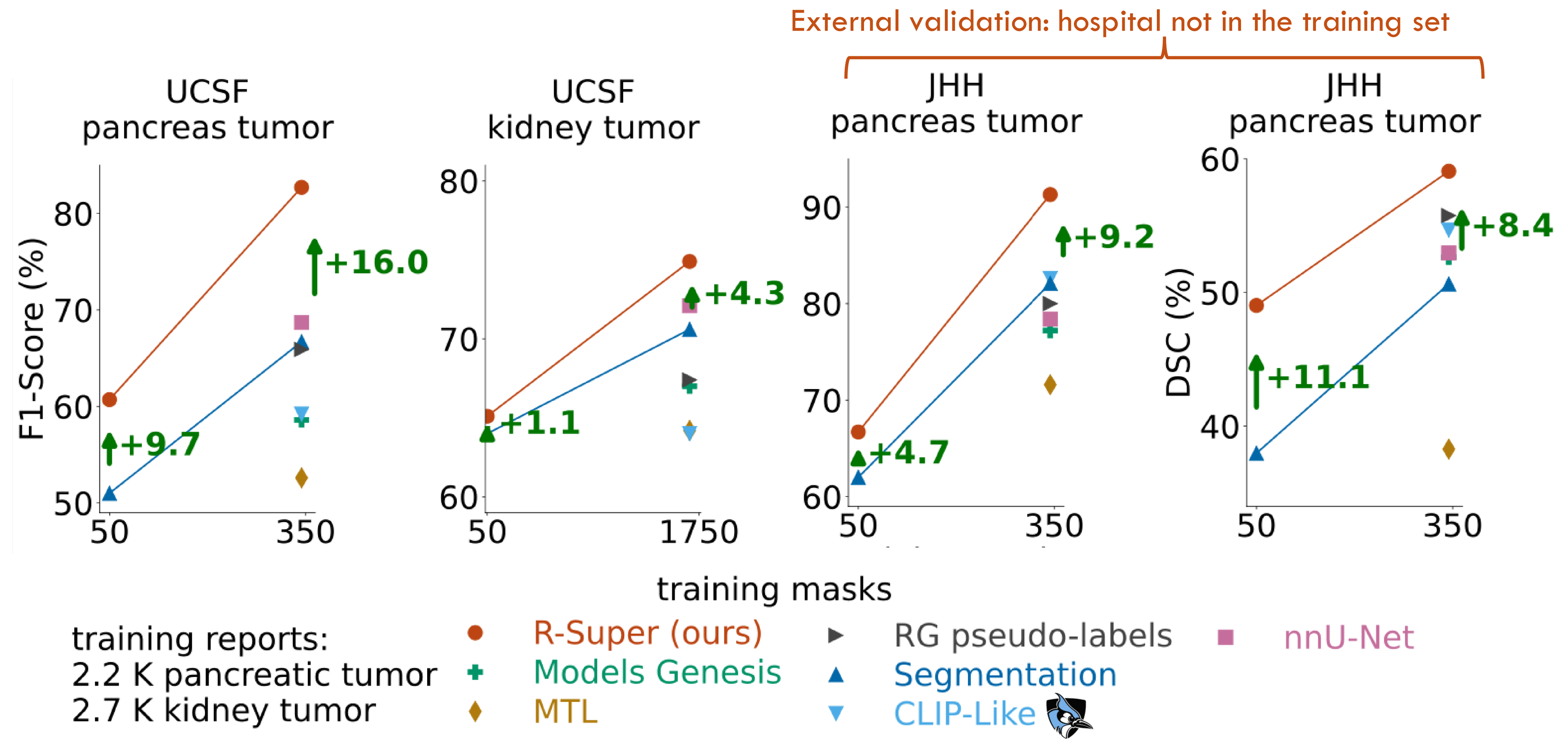


Results on Public Data: By Learning from Public Masks (*N*=344) & Reports (*N*=1.8K), R-Super Improves Pancreatic Tumor Detection



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, 1K with pancreatic lesions.

Scale: Learning from Reports & Masks, R-Super Boosts Tumor Detection & Segmentation with Few (50), Medium (344) or Many (1.7K) Masks



Early Detection: R-Super Improves the Detection of Small (<2 cm) Pancreatic Tumors by +26%/+15%/+8% in sensitivity/F1-Score/DSC — external validation

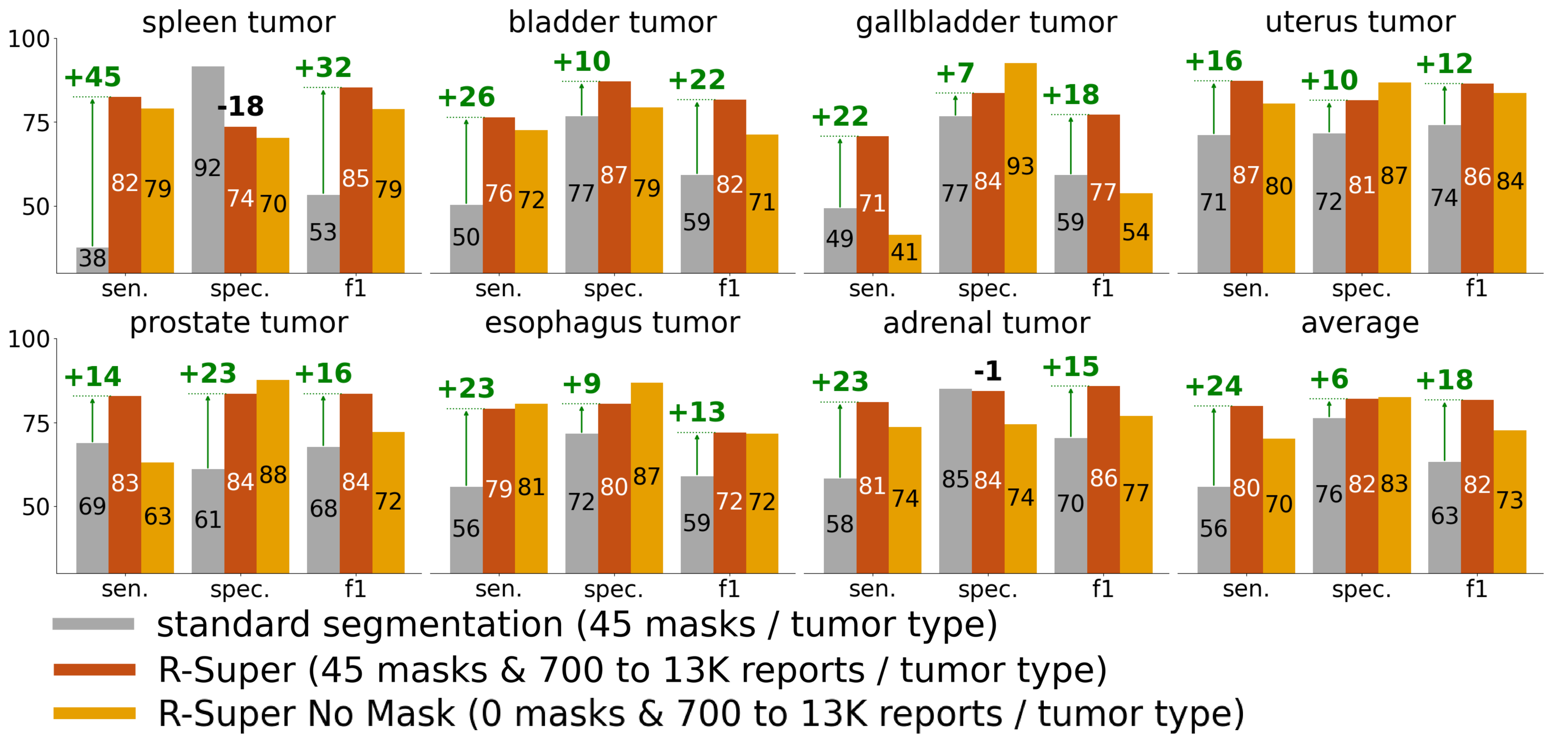
	pancreas tumor												kidney tumor					
	JHH-Test								UCSF-Test				UCSF-Test					
train paradigm	mask	rep.	dsc	nsd	F1	AUC	Se	Sp	F1	AUC	Se	Sp	mask	rep.	F1	AUC	Se	Sp
CLIP-Like [5]	344	2.2K	11	19	68	90	100	71	50	74	54	75	1.7K	2.7K	40	65	57	48
Multi-task l. [7]	344	2.2K	15	26	54	83	87	60	42	61	60	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

Key Takeaways:

- R-Super improves performance both when few and many masks are available. Thus, it can use reports to further scale datasets with **many** Image-Mask pairs, or to allow the segmentation of tumor types with **very few** Image-Mask pairs.
- R-Super improves results for **small tumors** and for **unseen hospitals**.

CLIP aligns the entire CT image with the report, but tumors occupy less than 0.0001% of the CT image, making alignment ineffective for segmentation.

Current Work: Training R-Super on 100,000 Image-Report-Mask pairs To Segment Tumor Types without Public Image-Mask pairs



Current & future work:

- Many Image-Report pairs train AI more effectively than few Image-Mask pairs.
- When Image-Mask pairs are limited or absent, Image-Report pairs provide a strong alternative, enabling segmentation of many tumor types.
- Our dataset of 117,000 Image–Report and 270 CT-Mask pairs yields an **+18%** F1-Score improvement for multi-tumor segmentation over Image-Mask training.
- We will release the **first** public AI model that can segment 7 tumor types.
- We will create Image-Mask pairs for these 7 tumor types in public Image-Report datasets (e.g., Merlin) to support broader research by the community.

This work was supported by IIT and the Lustgarten Foundation for Pancreatic Cancer Research and the National Institutes of Health (NIH) under Award Number R01EB037669.