MICCAI2025

Report Supervision

Report Supervision Loss: Scaling Tumor Segmentation with 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

Medical Image Computing and Computer Assisted Intervention (MICCAI 2025)

Problem: Few Per-voxel Tumor Labels

What are reports and per-voxel labels?

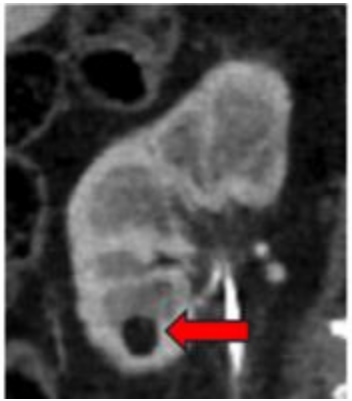
Why are reports important?

- Public CT datasets: **0 to 1.7K per-voxel tumor labels**
  - **0.9-1.7K**: liver, kidney, pancreas, lungs
  - **53-190**: Colon, Adrenal Glands
  - **0**: Uterus, Prostate, Esophagus, Spleen, Gallbladder, Bladder, Duodenum, Stomach

- One Hospital (UCSF): 0 per-voxel labels; **300K tumor reports**
  - Reports for **all tumor types**
  - Everyday work of a radiologist

Problem: Few Per-voxel Tumor Labels

- Informative: tumor size, location, and quantity

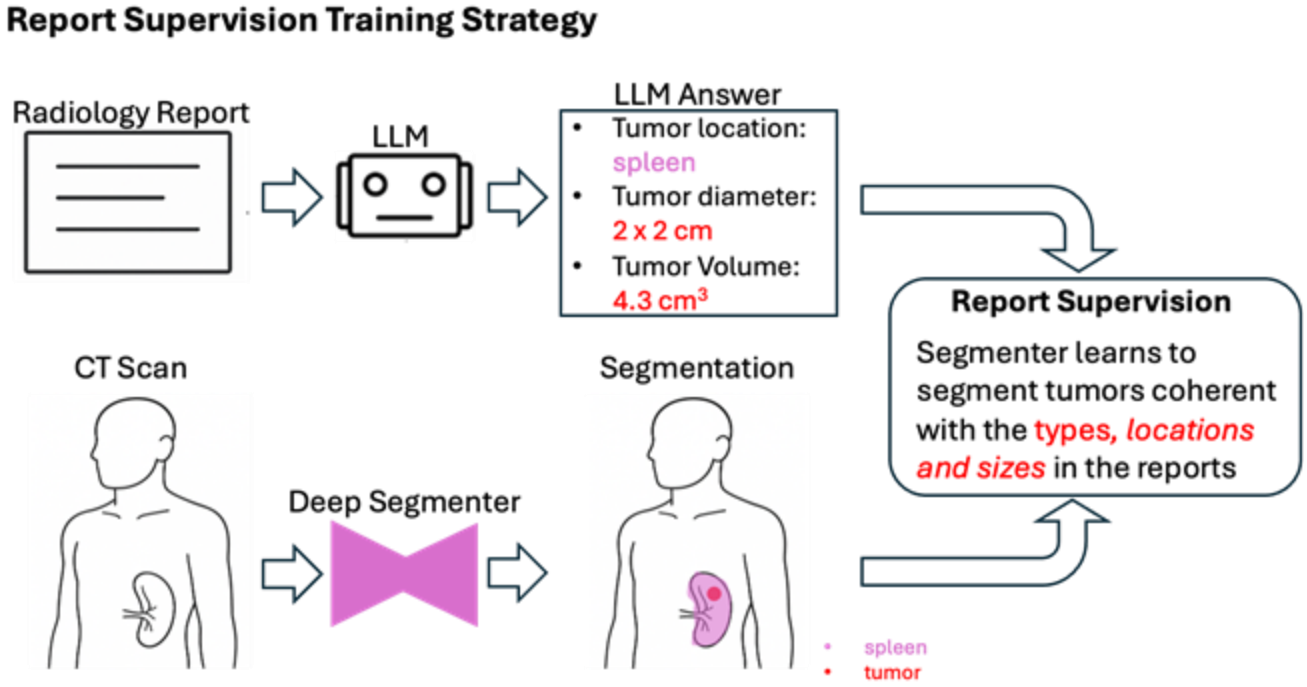


Report: [...] hypodense cystic lesion in the right kidney, measuring 11 x 10 mm [...]

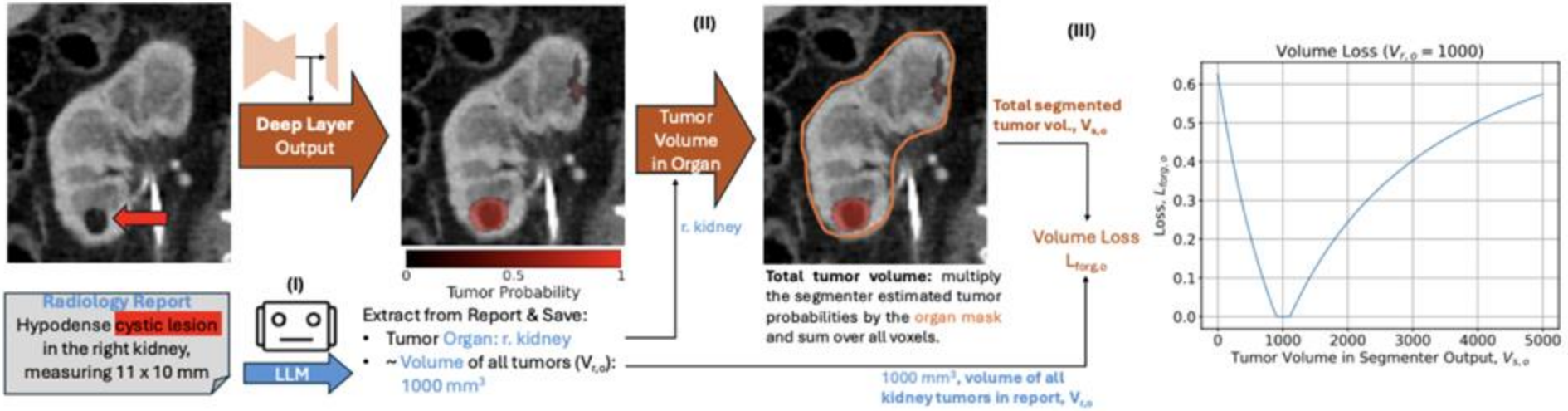
Caption: 9 misconceptions about alcohol.

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

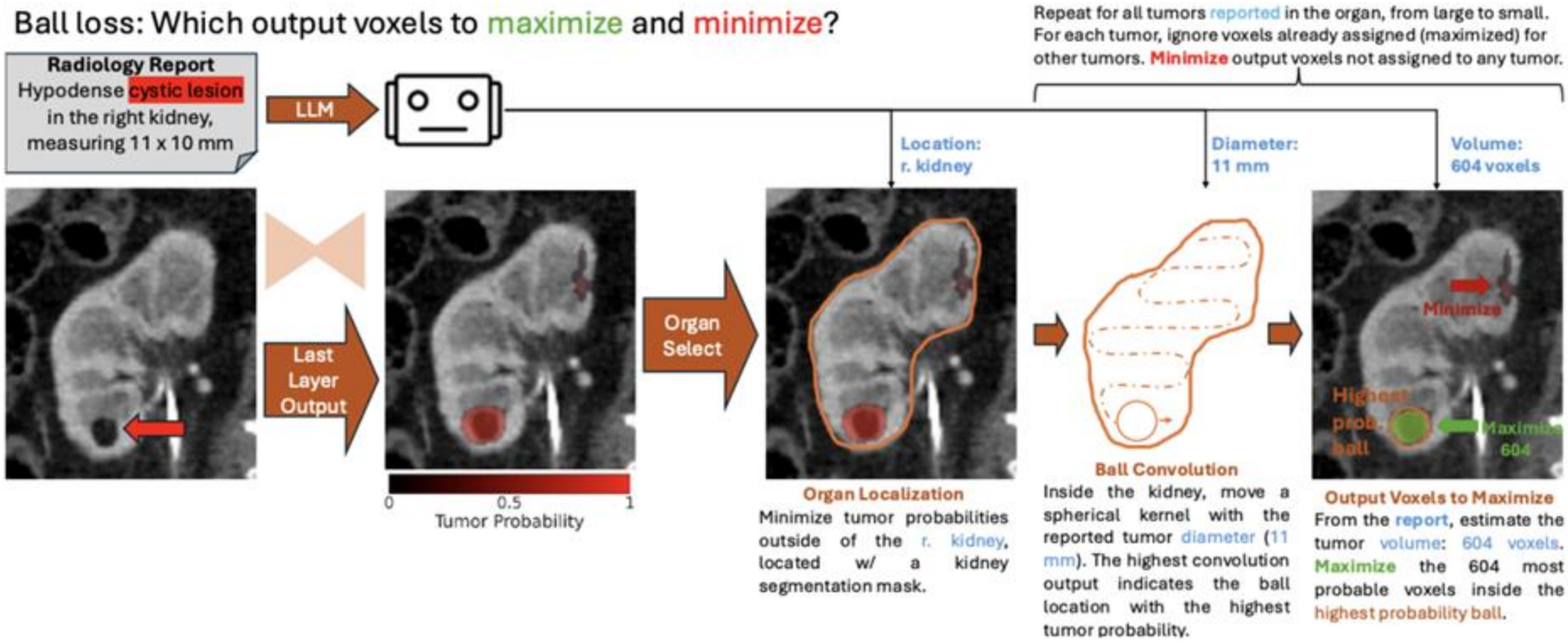
Report Supervision: Overview



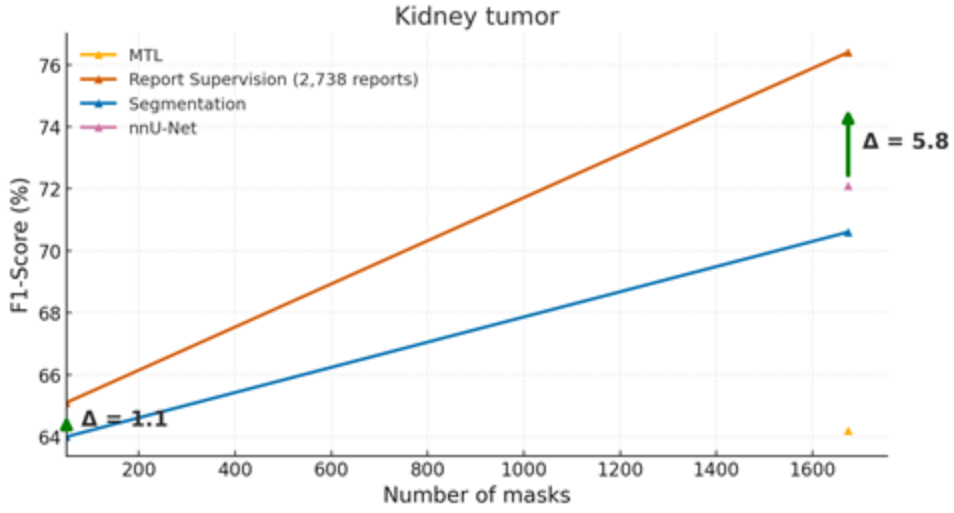
Methodology: Volume Loss

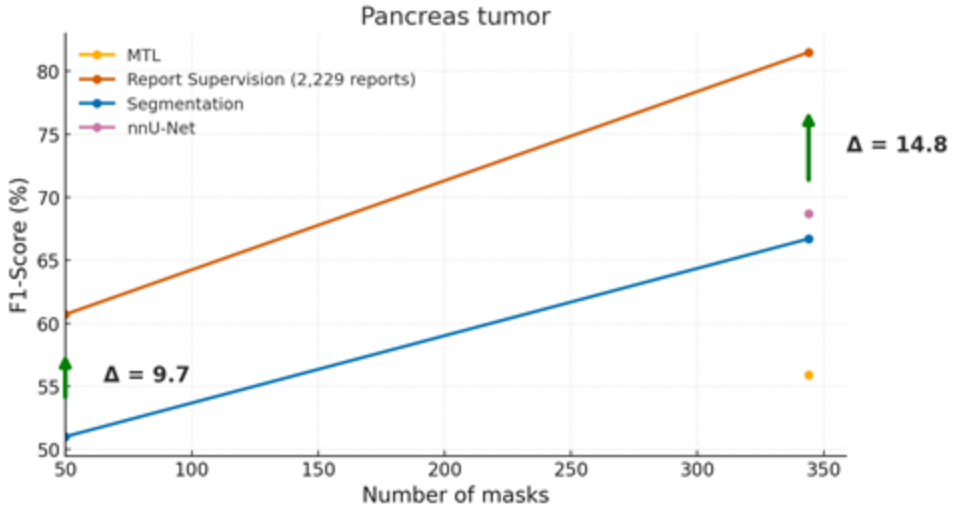


Methodology: Ball Loss



Results: Reports help with few / many masks

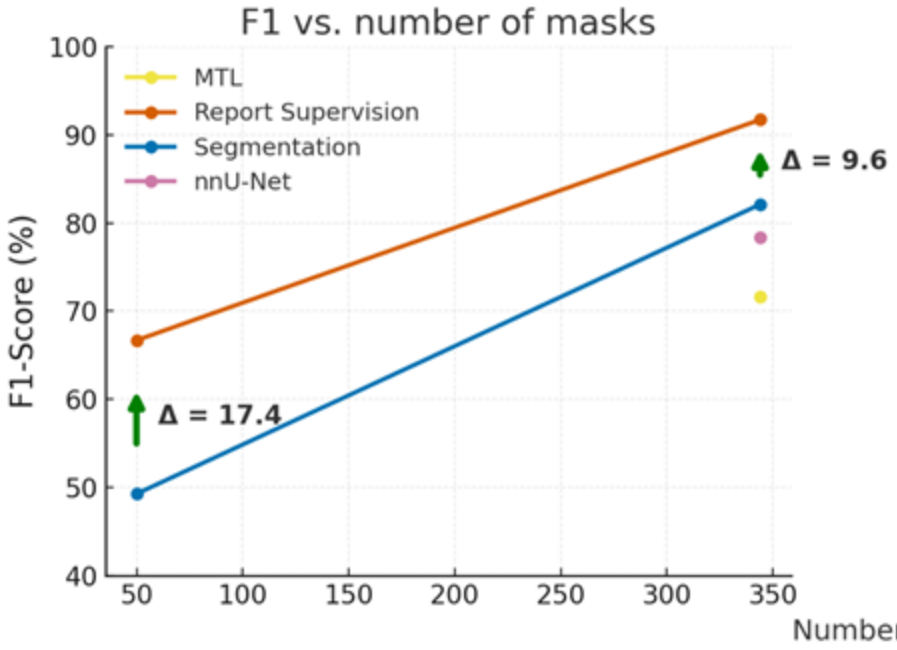


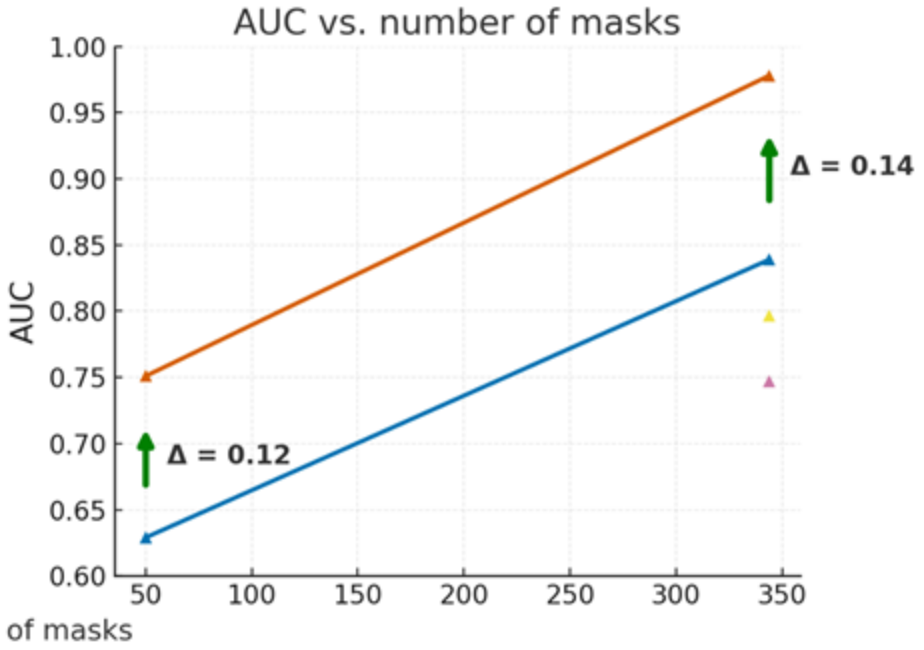


Results: IID

supervision for training		pancreas tumour				kidney tumour			
voxel-wise	report-wise	sen.	spec.	F1	AUC	sen.	spec.	F1	AUC
2 018	0	61.9 (86/139)	85.3 (191/224)	66.7	0.78	65.1 (110/169)	67.5 (131/194)	70.6	0.73
2 018	4 967	80.6 (112/139)	87.1 (195/224)	81.5	0.90	69.8 (118/169)	70.6 (137/194)	76.4	0.78
50	0	46.8 (65/139)	77.2 (173/224)	51.0	0.63	67.3 (113/168)	62.6 (122/195)	64.0	0.68
50	4 967	59.7 (83/139)	75.0 (168/224)	60.7	0.70	62.5 (105/168)	70.8 (138/195)	65.1	0.66

Results: OOD (Pancreas Tumors)





Results: OOD

supervision for training		pancreas tumor	
pancreatic tumor masks	pancreatic tumor reports	sen.	spec.
344	0	71.5% (974/1364)	88.4% (550/622)
344	2,229	84.7% (1155/1364)	92.1% (573/622)

Conclusions

Can reports improve segmentation performance?

- Improved performance in the hospital that provides the report
- Improved performance in unseen hospitals
- Improved performance with a small (50), medium (344) and large (1,674) number of masks
- Reports can: scale large segmentation datasets & create reasonable segmentation AI when very few masks are available, which can help further annotation