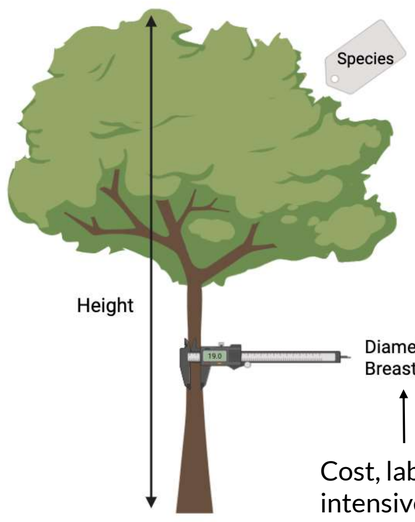
Assessing SAM for Tree Crown Instance Segmentation from Drone Imagery

Mélisande Teng, Arthur Ouaknine, Etienne Laliberté, Yoshua Bengio, David Rolnick, Hugo Larochelle

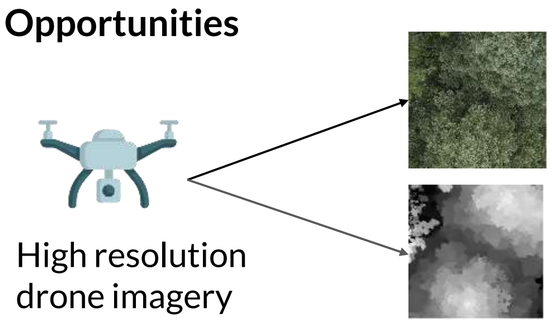
MOTIVATION

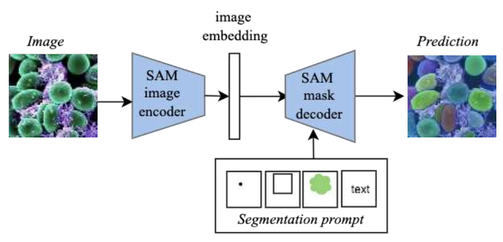
Traditional carbon estimation in tree plantations



Cost, labor and time intensive field surveys

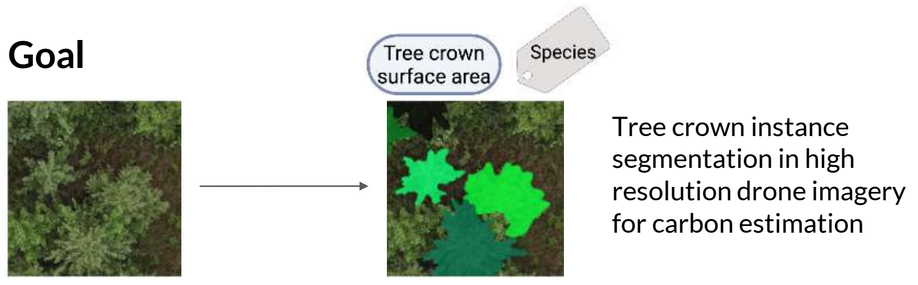
Opportunities





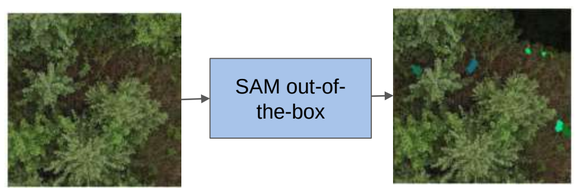
Foundation models: the Segment Anything Model (SAM)[1]

Goal



Tree crown instance segmentation in high resolution drone imagery for carbon estimation

Current limitations



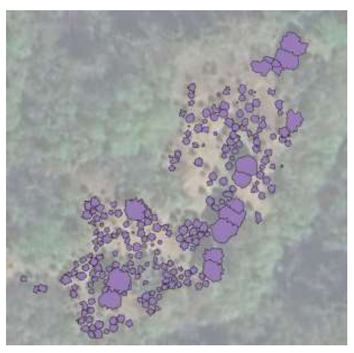
- SAM out-of-the-box is not suited to high-resolution drone imagery
- No species information

How can we leverage the Segment Anything Model for the task of tree crown instance segmentation? Can the models be improved by including task specific information in the form of the DSM?

DATASET

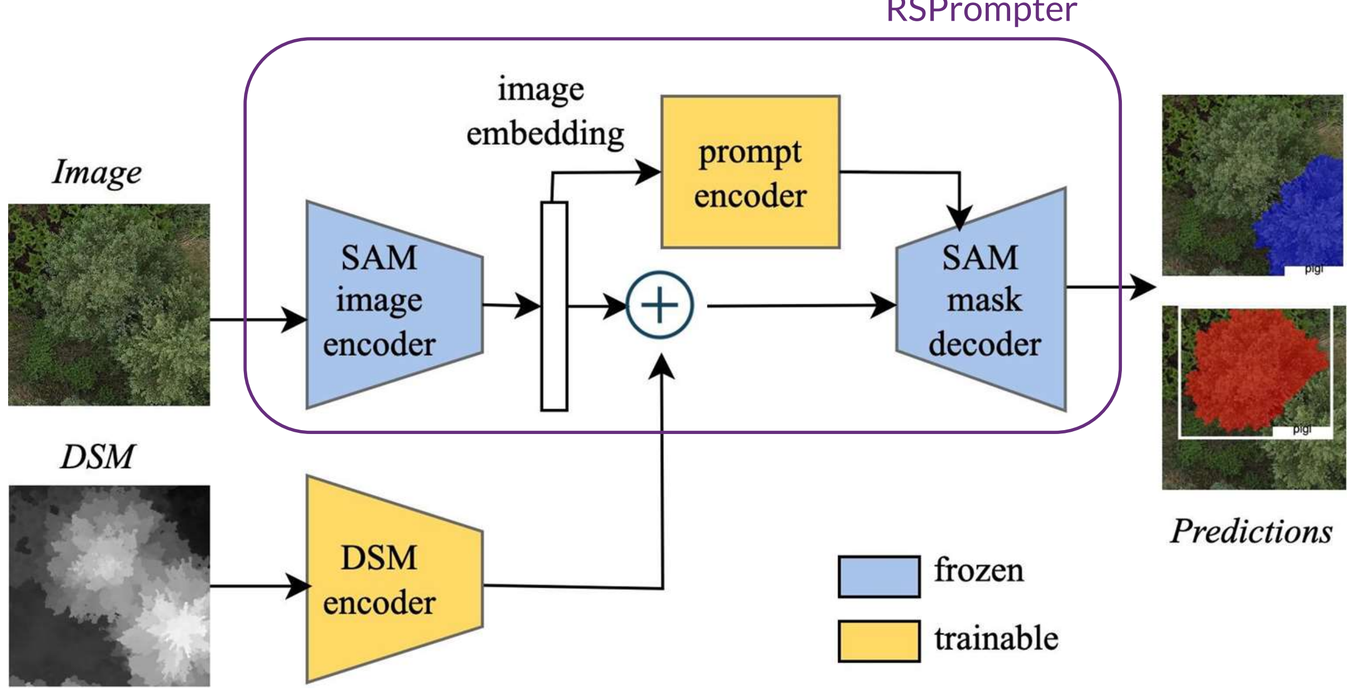
Quebec Plantations dataset

- 15 plantation sites
- 19,500 annotated individual tree crowns
- 5mm/px drone imagery
- 9 tree species of interest

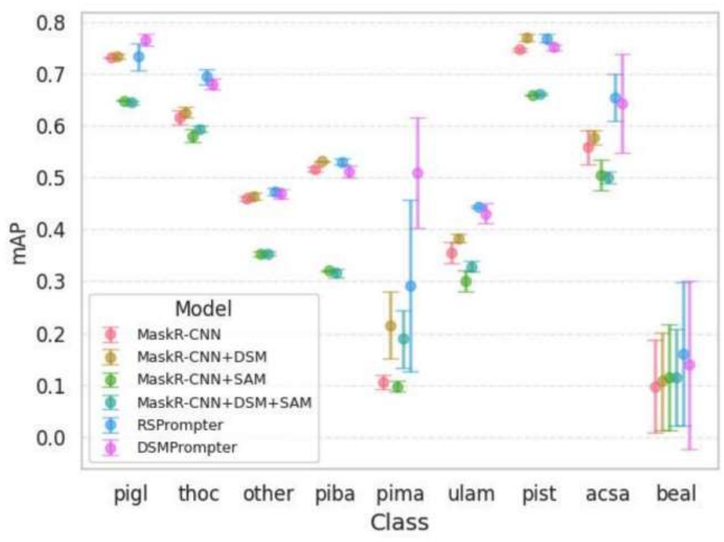


DSMPrompter

RSPrompter[2]



Model	DSM	Pre-trained	Single-class mAP	Single-class mIoU	Multi-class mAP	Multi-class wmAP
SAM (100 pps)	✗	–	8.05	35.06	–	–
SAM (10 pps)	✗	–	10.11	34.01	–	–
SAM	✓(prompts)	–	9.28	46.15	–	–
Mask R-CNN	✗	✗	59.36 (±0.21)	79.63 (±0.53)	42.69 (±2.93)	55.75 (±1.50)
	✗	✓	63.65 (±0.43)	81.82 (±0.36)	46.51 (±1.13)	58.30 (±1.23)
	✓	✓	64.64 (±0.69)	81.89 (±0.60)	48.96 (±1.06)	60.32 (±0.73)
Faster R-CNN+SAM	✗	✗	53.56 (±0.20)	76.22 (±0.20)	33.52 (±0.44)	45.79 (±0.67)
	✗	✓	57.85 (±0.66)	78.0 (±0.56)	39.79 (±1.17)	50.30 (±1.51)
	✓	✓	58.0 (±0.25)	78.27 (±0.74)	40.14 (±1.41)	52.08 (±1.73)
Mask R-CNN+SAM	✗	✓	57.6 (±0.19)	78.18 (±0.31)	39.76 (±1.19)	50.46 (±0.52)
	✓	✓	57.83 (±0.11)	77.65 (±0.5)	41.13 (±1.12)	51.33 (±0.85)
RSPrompter	✗	–	66.37 (±0.91)	82.58 (±1.63)	52.77 (±1.03)	62.37 (±2.45)
DSMPrompter	✓	–	65.03 (±1.76)	83.24 (±0.41)	54.40 (±4.00)	64.84 (±1.49)



Ground truth RSPrompter DSMPrompter

TAKEAWAYS

- Methods using SAM for inference without further tuning do not outperform a simpler Mask R-CNN model trained specifically for our task of interest.
- However when tuned properly, SAM components can be leveraged to make it a powerful tool.
- Adding the DSM improves predictions of all models.

[1] Kirillov, Alexander, et al. "Segment anything." *Proceedings of the IEEE/CVF international conference on computer vision*. 2023.
[2] Chen, Keyan, et al. "RSPrompter: Learning to prompt for remote sensing instance segmentation based on visual foundation model." *IEEE Transactions on Geoscience and Remote Sensing* 62 (2024): 1-17.

Mila Université de Montréal McGill