Aalto University
School of Science

Master’s Programme in Computer, Communication and Information Sciences

SAM2 pseudolabeling for instance segmentation

Stefan Rua

Master’s Thesis
2025

Aalto University
School of Science

Author Stefan Rua

Title SAM2 pseudolabeling for instance segmentation

Degree programme Computer, Communication and Information Sciences

Major Macadamia

Supervisor Jorma Laaksonen

Advisor Julius Pesonen (MSc)

Collaborative partner Finnish Geospatial Research Institute FGI

Date 30 September 2025 **Number of pages** 11 **Language** English

Abstract
Yappidi-yap, lorem ipsum etc.

Keywords pseudolabeling, instance segmentation, forestry

Aalto-yliopisto
Perustieteiden
korkeakoulu

Tekijä Stefan Rua

Työn nimi SAM2 pseudolabelöinti instanssisegmentaatiokoulutuksessa

Koulutusohjelma Computer, Communication and Information Sciences

Pääaine Macadamia

Työn valvoja Jorma Laaksonen

Työn ohjaaja DI Julius Pesonen

Yhteistyötaho Paikkatietokeskus FGI

Päivämäärä 30.9.2025 **Sivumäärä** 11 **Kieli** englanti

Tiivistelmä

Höpötihöp, lorem ipsum jne.

Avainsanat pseudolabelöinti, instanssisegmentaatio, metsäily

Contents

1 Introduction

Airborne remote sensing based tree mapping methods are used for forest health monitoring [1] and city planning [2] due to their efficiency, but the training process is often bottlenecked by creating high quality annotations manually. The goal of this study is to test if training results can be improved in instance segmentation tasks by refining coarse segments calculated from a canopy height model (CHM) using Segment Anything Model 2 (SAM2) [3].

2 Related Work

The previous version of SAM2, SAM has been assessed for tree crown instance segmentation on drone imagery [4]. The study uses SAM in the following ways: generating masks without any prompts, generating masks with digital surface model (DSM) maxima as point prompts, and prompting SAM with predictions from a trained Mask R-CNN model. Of these, the last approach is closest to the one examined here.

Another study examining SAM for remote sensing use proposes RSPrompter [5], a method that learns to generate prompts from the SAM's image encoder, then feeding them to the decoder.

3 Materials and Methods

3.1 Datasets

The main dataset consists of a multispectral orthophoto spanning approximately 2 km^2 and CHM-based coarse tree crown segments. The orthophoto covers both forest and urban area, has a GSD of 2.5 cm, and contains blue, green, red, near-infrared, ???, and thermal bands.

- taken by drone/helicopter?

3.2 Models

SAM2 is used for the pseudolabeling task and Mask R-CNN [6] for the supervised instance segmentation.

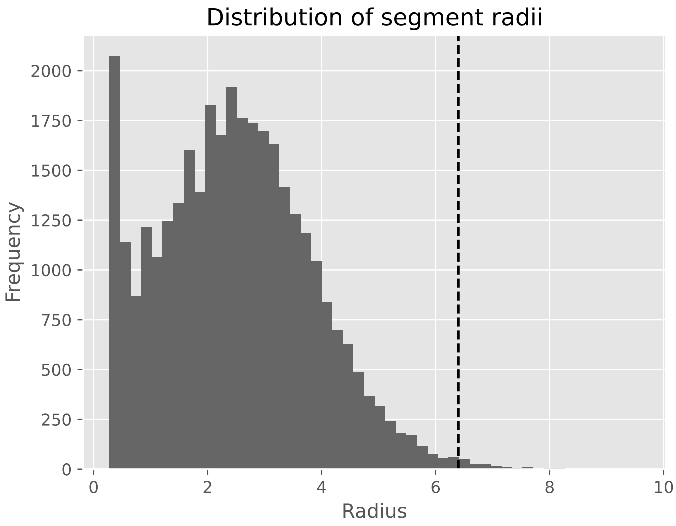
3.3 Methods

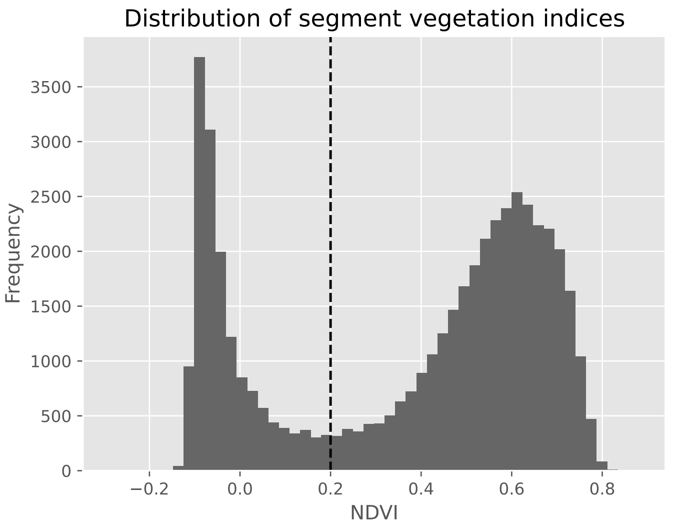
First the coarse segments were pseudolabeled using SAM2, indexing the orthphoto in a grid with a window of size 1024 and a stride of 512. For each window, only the segments whose centroids were located in the central 512x512 square of the window were selected for pseudolabeling. Then, the bounding box of each segment was passed as a prompt to SAM2, and the largest connected component of the output mask was saved as the pseudolabel. These parameters for the window size and stride were selected in order to avoid stitching artifacts and to provide SAM2 with images mathcing the native input resolution of the model.

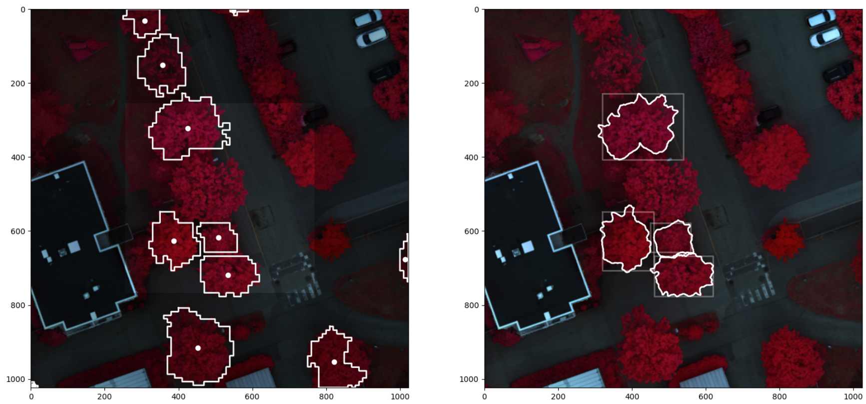
A small test area of 25 trees was segmented manually. To evaluate the quality of the coarse segments and quantify the effect of SAM2 pseudolabeling, the coarse segments and pseudolabels were compared to the manual segments using the Jaccard index [7]:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|},$$

where A and B are the segments to be compared. Then, a Mask R-CNN model with a ResNet-50 [8] backbone was trained separately on both the coarse segments and pseudolabels.


Distribution of segment radii
Frequency
Radius
0
250
500
750
1000
1250
1500
1750
2000
0
2
4
6
8
10


Distribution of segment vegetation indices
Frequency
NDVI
0
500
1000
1500
2000
2500
3000
3500
−0.2
0.0
0.2
0.4
0.6
0.8



4 Results

5 Discussion

6 Conclusions

References

[1] S. Ecke, J. Dempewolf, J. Frey, A. Schwaller, E. Endres, H.-J. Klemmt, D. Tiede, and T. Seifert, "Uav-based forest health monitoring: A systematic review," *Remote Sensing*, vol. 14, no. 13, 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/13/3205

[2] L. Velasquez-Camacho, A. Cardil, M. Mohan, M. Etxegarai, G. Anzaldi, and S. de Miguel, "Remotely sensed tree characterization in urban areas: A review," *Remote Sensing*, vol. 13, no. 23, 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/23/4889

[3] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, E. Mintun, J. Pan, K. V. Alwala, N. Carion, C.-Y. Wu, R. Girshick, P. Dollár, and C. Feichtenhofer, "Sam 2: Segment anything in images and videos," 2024. [Online]. Available: https://arxiv.org/abs/2408.00714

[4] M. Teng, A. Ouaknine, E. Laliberté, Y. Bengio, D. Rolnick, and H. Larochelle, "Assessing sam for tree crown instance segmentation from drone imagery," 2025. [Online]. Available: https://arxiv.org/abs/2503.20199

[5] K. Chen, C. Liu, H. Chen, H. Zhang, W. Li, Z. Zou, and Z. Shi, "Rsprompter: Learning to prompt for remote sensing instance segmentation based on visual foundation model," 2023. [Online]. Available: https://arxiv.org/abs/2306.16269

[6] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," 2018. [Online]. Available: https://arxiv.org/abs/1703.06870

[7] G. K. Gilbert, "Finley's tornado predictions." *American Meteorological Journal. A Monthly Review of Meteorology and Allied Branches of Study (1884-1896)*, vol. 1, no. 5, p. 166, 1884.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1512.03385