Leveraging state of the art computer vision models for tree monitoring in silvopastoral systems

Andrés Felipe Ruiz, Juan Andrés Cardoso.
International Center for Tropical Agriculture, Tropical Forages Program.
Contact: a.f.ruiz@cgiar.org

INTRODUCTION

- Tree monitoring in silvopastoral system affronts several challenges for robust identification of individual trees and tree changes over time. Recent advances in remote sensing and artificial intelligence, particularly computer vision, have potential for tree monitoring through multiresolution RGB imagery.
- Several studies have addressed this challenge before, offering good results to some extent (HighResCanopyHeight, DeepForest, TreeCrownDelineation, Treeformer) and there are also more general AI model architectures that can be used for this goal (Mask R-CNN, U-Net, Segment Anything Model).
- Some limitations related to spatial and temporal resolutions, georeferencing, visual artifacts, data handling, computational resources management, and expertise and ease to use constraint the usability of these models.

OBJECTIVE

Evaluate the performance of different AI models for tree monitoring in silvopastoral systems, identifying ease of use, feasibility and usability for RGB remote sensing imagery. Identify the basic computer vision tasks performed by the model (semantic segmentation, instance segmentation and object detection) and the possibility to derive useful insights from the model inference process.

METHODOLOGY

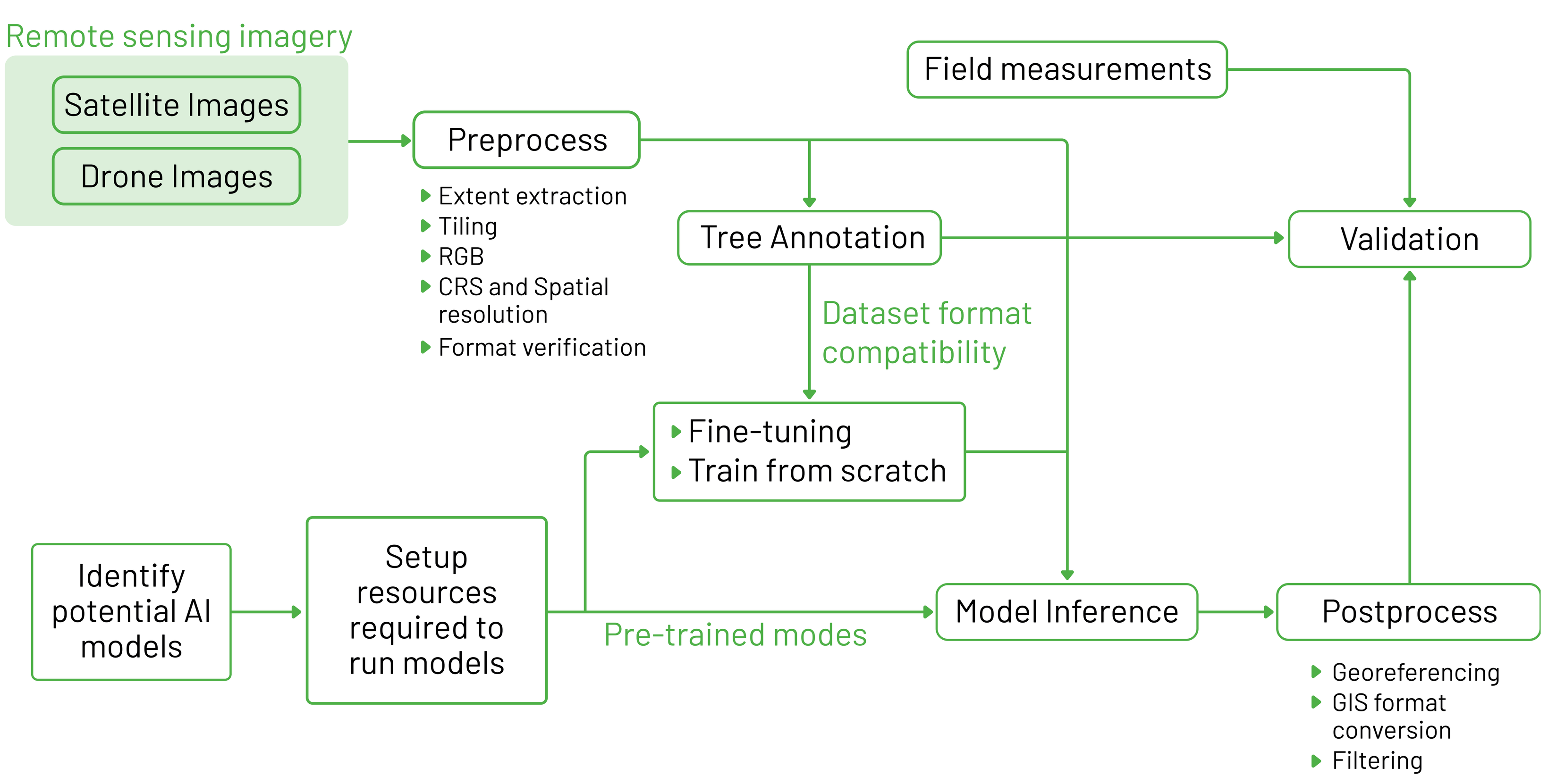

Figure 1. Methodology: potential models were selected and tested. Some models were fine-tuned or trained from scratch.

RESULTS

Table 1. Top AI models evaluated.

REPOSITORY/ MODEL	HighResCanopyHeight	Mask-RCNN	Detecttree	Deepforest	Treeformer
INTENDED SPATIAL RESOLUTION	1 m - multiresolution	4.77 m	–	0.1 m	0.2 m, 0.8 m, 0.12 m
COMPUTER VISION TASK	Semantic segmentation and height regression	Instance segmentation, object detection	Semantic segmentation	Object detection	Point mask segmentation
MODEL BASE ARCHITECTURE	Vision Transformer	Mask-RCNN	AGBoost	Retinanet	Deep network (encoder-decoder architecture)
APPROACH	Pre-trained model	Fine-tuned model	Pre-trained model, train from scratch	Pre-trained model	Train from scratch
RESULTS	Canopy height raster image	Tree instance vector image	Tree mask raster image	Tree bounding boxes	Tree centroids raster image
POST PROCESSING DERIVED RESULTS	• Tree mask raster image • Tree instance vector image • Tree bounding boxes • Tree centroids	• Tree mask raster image • Tree bounding boxes • Tree centroids	–	• Tree centroids	–

*Additional models were evaluated including DPA, however they did not fulfill the basic requirements of handling only RGB images or the setup required for execution was problematic. Segment Anything Model, for example, performed poorly for remote sensing imagery, especially for small objects like trees.

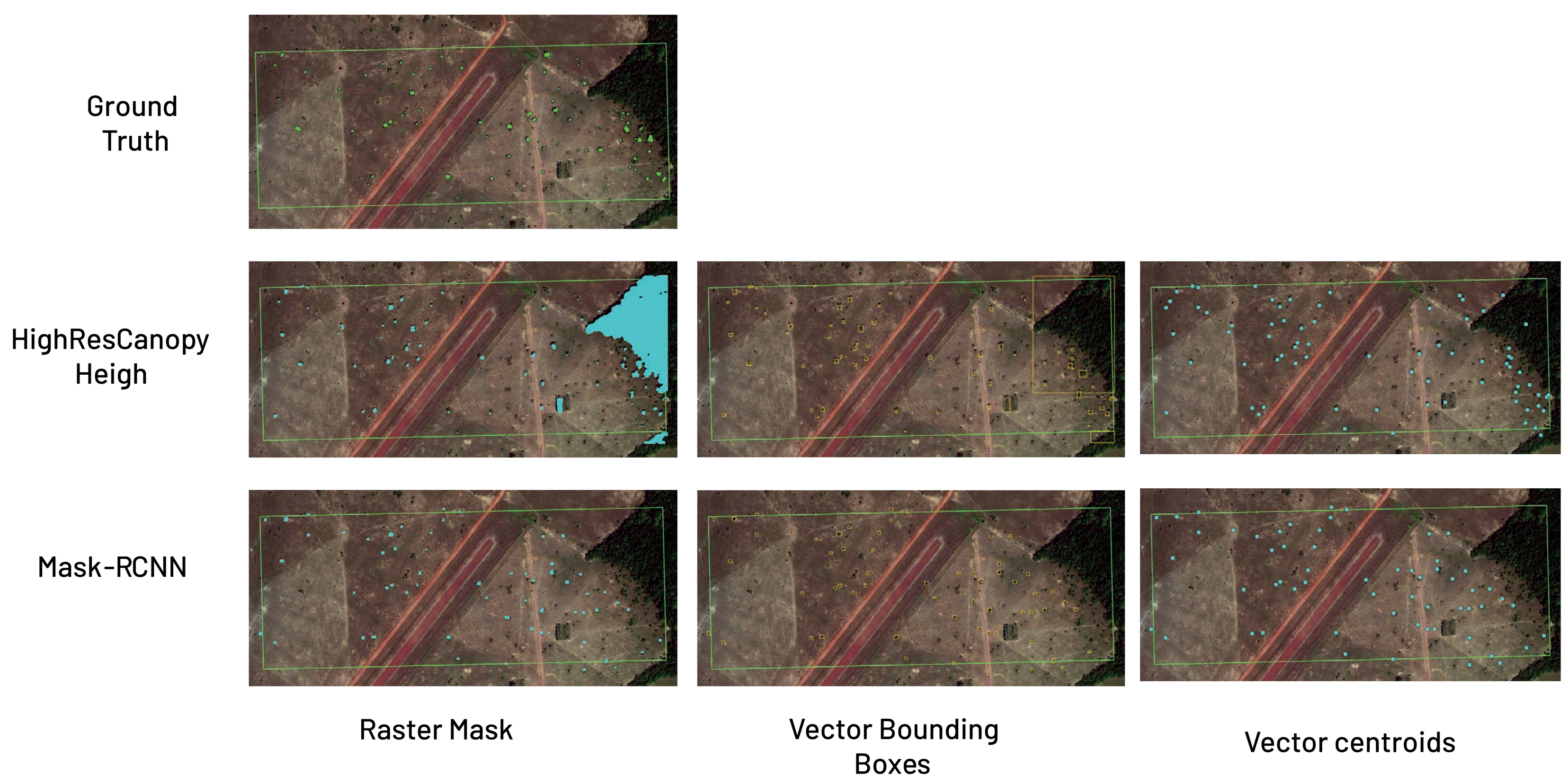

Figure 2. Visual results for the top models with more postprocessing options.

REFERENCES

Tolan, J., Yang, H.-I., Nosarzewski, B., Couairon, G., Vo, H. V., Brandt, J., Spore, J., Majumdar, S., Haziza, D., Vamaraju, J., Moutakanni, T., Bojanowski, P., Johns, T., White, B., Tiecke, T., & Couprie, C. (2024). Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. Remote Sensing of Environment, 300, 113888. https://doi.org/10.1016/j.rse.2023.113888

Weinstein, B. G., Marconi, S., Bohlman, S., Zare, A., & White, E. (2019). Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks. Remote Sensing, 11(11), 1309. https://doi.org/10.3390/rs11111309

CONCLUSIONS

- The Vision transformer model (HighResCanopyHeight) offered the best performance and versatility to handle multiple resolutions, given that this model generates a canopy height image its utility improves as it offers more information compared to other models. Mask R-CNN performs well only for the spatial resolutions it was trained on. Other models like Deepforest performed better on aerial images. The best approach is to take advantage of the results of multiple models.
- The AI models themselves are not enough for a complete tree monitoring system. Thorough preprocessing and postprocessing steps are required to have a fully usable workflow (for example tiling, georeferencing and format conversion). In practice it was noted that even though the full workflow generated usable results, an element of interactivity could be needed, demonstrating the need for a user-friendly tool to execute the models.

ACKNOWLEDGEMENTS

This work was conducted as part of the CGIAR Initiatives on Accelerated Breeding (ABI) and Sustainable Animal Productivity (SAP). We thank all donors who globally support our work through their contributions to the CGIAR System.

Alianza Bioversity & CIAT
CGIAR INITIATIVE ON Accelerated Breeding
CGIAR INITIATIVE ON Sustainable Animal Productivity

POSTER PREPARED FOR
Tropentag 2024
September 11-13, 2024
Vienna (Austria)