Check for updates

Remote Sensing in Ecology and Conservation

Zoological Society of London

METHODS AND TOOLS OPEN ACCESS

Sam-Stars: A Multimodal, Prompt-Guided Approach for Individual Tree Crown Delineation and Species Mapping

Joseph Emile Honour Percival[1] | Kanoa Lindiwe[1] | Kalena Shiroma[1] | Adrian Pascual[2] | Juan Guerra-Hernandez[3] | Rebecca Ostertag[4] | Ryan Perroy[1]

[1]Department of Geography & Environmental Science, University of Hawai'i at Hilo, Hilo, Hawai'i, USA | [2]Department of Geographical Sciences, University of Maryland, College Park, Maryland, USA | [3]Forest Research Centre, Associate Laboratory TERRA, School of Agriculture, University of Lisbon, Lisbon, Portugal | [4]Department of Biology, University of Hawai'i at Hilo, Hilo, Hawai'i, USA

Correspondence: Joseph Emile Honour Percival (jperciva@hawaii.edu)

Received: 30 January 2026 | **Revised:** 19 August 2026 | **Accepted:** 29 September 2026

Editor: Mat Disney | **Associate Editor:** Stephanie A. Bohlman

Keywords: lidar | segment anything model | stardist | WorldView-3

ABSTRACT

Tree-level information is increasingly required to reduce uncertainty in biodiversity assessment and species-aware carbon accounting, yet individual tree crown (ITC) delineation and species mapping remain challenging in compositionally heterogeneous forests at landscape scales. We present a scalable multimodal workflow for crown-resolved, species-level mapping that fuses very high-resolution (VHR) WorldView-3 imagery with airborne lidar. The approach uses lidar-guided prompting to drive a minimally fine-tuned Segment Anything Model (SAM) for high-recall canopy detection, followed by shape-aware instance separation with StarDist to delineate individual crowns. Across benchmarks on Hawai'i Island, the proposed sam-stars framework achieves strong delineation performance with limited annotation, outperforming an end-to-end Detectron2 (Mask R-CNN) baseline most strongly at small training sizes. In an object-based species-mapping workflow, crown-level multispectral and structural features are classified with gradient-boosted trees, yielding the highest species-aware semantic panoptic quality (PQ = 0.47) among evaluated methods. Applied at landscape scale using a delineation model trained on only 254 crowns and a larger independent species inventory, the workflow produced a wall-to-wall map of 333,204 crowns. When aggregated to 30 m resolution, crown-derived canopy cover was 19.0% compared with 14.9% in a publicly available land-cover product, indicating substantial under-detection in existing maps. By enabling label-efficient, crown-resolved, species-aware mapping from widely available lidar and VHR imagery, this framework supports applications ranging from invasive species detection to restoration planning and species-informed carbon accounting in structurally heterogeneous forest landscapes.

1 | Introduction

Forests are fundamental to biogeochemical cycles and biodiversity conservation, yet quantifying aboveground carbon and biodiversity patterns at high resolution remains challenging. Most global biomass products are produced at tens to hundreds of meters resolution and cannot resolve stand-level species composition or fine-scale structural heterogeneity (Avitabile et al. 2016). Crown architecture and phenology among species translate into distinct spectral and structural signatures; failure to resolve species composition can bias carbon estimates and mischaracterize biodiversity patterns. These limitations motivate mapping individual tree crowns (ITCs), where species identity, height, and crown area can be directly measured (Dalponte and Coomes 2016).

The past decade has seen an increase in high-resolution remote sensing data suitable for ITC delineation. Unmanned aerial

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2026 The Author(s). *Remote Sensing in Ecology and Conservation* published by John Wiley & Sons Ltd on behalf of Zoological Society of London.

vehicles (UAVs) can capture centimeter-scale imagery, and lidar (light detection and ranging) can provide three-dimensional point clouds that resolve crown geometry, but UAV analyses are typically limited by flight range and logistics. Satellite platforms cover entire landscapes in a single overpass, but even the finest commercial very high-resolution (VHR) systems are ~30 cm resolution. At these scales, individual crowns span few pixels, and mixed backgrounds, shadows, and atmospheric effects dilute spectral signals, particularly in closed-canopy forests (Fassnacht et al. 2016). These trade-offs motivate multimodal approaches that combine satellite imagery with airborne lidar-derived structure.

Crown-resolved species mapping has developed along several methodological directions. Object-based workflows delineate crowns and classify crown-level spectral and structural features, performing well when crown boundaries are reliable. End-to-end deep learning approaches predict crown instances, and sometimes species labels, directly from imagery, but training typically benefits from larger annotated datasets and can be sensitive to imaging conditions and domain shift across sensors or canopy structures. These limitations motivate interest in prompt-guided foundation models as a potentially more label-efficient and transferable alternative. The Segment Anything Model (SAM; Kirillov et al. 2023) is pretrained on over one billion masks and can recover object boundaries from sparse prompts, making it attractive for crown delineation when annotations are limited. Early forestry applications show promise, but in compositionally heterogeneous, closed-canopy forests performance remains sensitive to prompt placement and density, particularly under crown overlap and occlusion (Teng et al. 2025). This suggests that performance is often constrained by prompting rather than model capacity alone. We therefore use prompt-guided segmentation primarily to recover crown geometry and treat species attribution as an object-based classification problem using crown-level spectral and structural features and independent field labels.

Here, we present sam-stars, a scalable multimodal framework for crown-resolved species mapping and panoptic evaluation from VHR WorldView-3 imagery fused with airborne lidar. The central idea is to decouple high-recall canopy detection from instance separation: lidar-guided prompts drive a minimally fine-tuned SAM to generate a continuous crown-likelihood surface, which is then converted into individual separated, star-convex crown instances. We use StarDist (Schmidt et al. 2018; Tong and Zhang 2025) rather than raw SAM masks because prompt-level SAM outputs can overlap, merge neighboring crowns, or produce multiple candidate masks for a single crown. We quantify annotation efficiency by benchmarking delineation performance as a function of training set size and compare against Detectron2 instance-segmentation and panoptic baselines. We then classify delineated crowns using a gradient-boosted model trained on crown-level features and demonstrate a landscape-scale application in the Pu‘uwa‘awa‘a region of Hawai‘i Island, USA.

2 | Materials and Methods

2.1 | Sam-Stars Workflow Overview

The sam-stars workflow is organized into three stages: (i) lidar-guided crown detection using prompt-based segmentation, (ii) shape-aware instance separation, and (iii) object-based species classification with XGBoost (Figure 1). We evaluate this workflow in two settings: (1) controlled benchmarking on manually delineated crowns (Section 2.3), and (2) landscape-scale deployment across a 16,300 ha study area using independent species inventory data (Section 2.4).

2.2 | Individual Tree Crown Delineation and Classification Using Sam-Stars

For delineation, sam-stars first constructs a canopy-boundary cost surface that stabilizes edge cues by combining canopy height model (CHM) gradients, inverse NDVI (calculated as 1—NDVI, so low-greenness gaps and shadows have higher values), and fine-scale texture. The CHM was generated as the maximum normalized lidar height in each 0.31 m WorldView-3 grid cell. Candidate crown centers (“seeds”) were detected as local maxima on a Gaussian-smoothed CHM using a 3×3 maximum filter and CHM > 2 m threshold. Redundant detections within 0.8 m were merged by retaining the tallest peak, and candidates in canopy gaps (CHM < 0.5 m or high gap cost) were removed. Each retained seed generated one SAM prompt: an 8 m seed-centered box, the seed plus up to 80 foreground points from nearby tall, low-cost canopy, up to 150 background points from gaps or low-canopy, and up to 300 background points near neighboring seeds. SAM is applied to a WorldView-3 NIR-1/red/green composite (bands 7, 5, and 3; approximately 770–895, 630–690, and 510–580 nm) and minimally fine-tuned on crown masks. Per-prompt mask logits are converted to probabilities and mosaicked into a crown-likelihood surface (Figure 1).

The sam-stars method then constructs a canopy index raster by combining the SAM crown-likelihood surface with auxiliary structural cues (scaled CHM and the cost surface) and uses this as input to StarDist (Schmidt et al. 2018) to separate adjacent crowns using a star-convex shape prior (Tong and Zhang 2025). This step converts overlapping prompt-level SAM probabilities into nonoverlapping crown instances suitable for downstream feature extraction and classification.

Each delineated crown is subsequently classified into the six species classes defined in Section 2.4 using crown-level features derived from WorldView-3 and lidar, and a gradient-boosted decision tree model (XGBoost). Features summarize all eight WorldView-3 bands: Coastal (400–450 nm), Blue (450–510), Green (510–580), Yellow (585–625), Red (630–690), Red Edge (705–745), NIR-1 (770–895), and NIR-2 (860–1040). These summaries are combined with fine-scale Coastal-band texture and lidar-derived canopy structure, yielding a 59-variable feature vector per crown; details are provided in the Supplement.

2.3 | Benchmarking Experiments and Evaluation Metrics

In the controlled benchmarking experiment, we use a single, fixed, approximately 70/30 train/validation split of a reference crown polygon dataset ($n = 4141$ crowns). The reference crowns came from crown-annotation plots distributed across the ahupua‘a to capture variation in canopy structure and species

20563485, 0, Downloaded from https://zslpublications.onlinelibrary.wiley.com/doi/10.1002/rse2.70117, Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

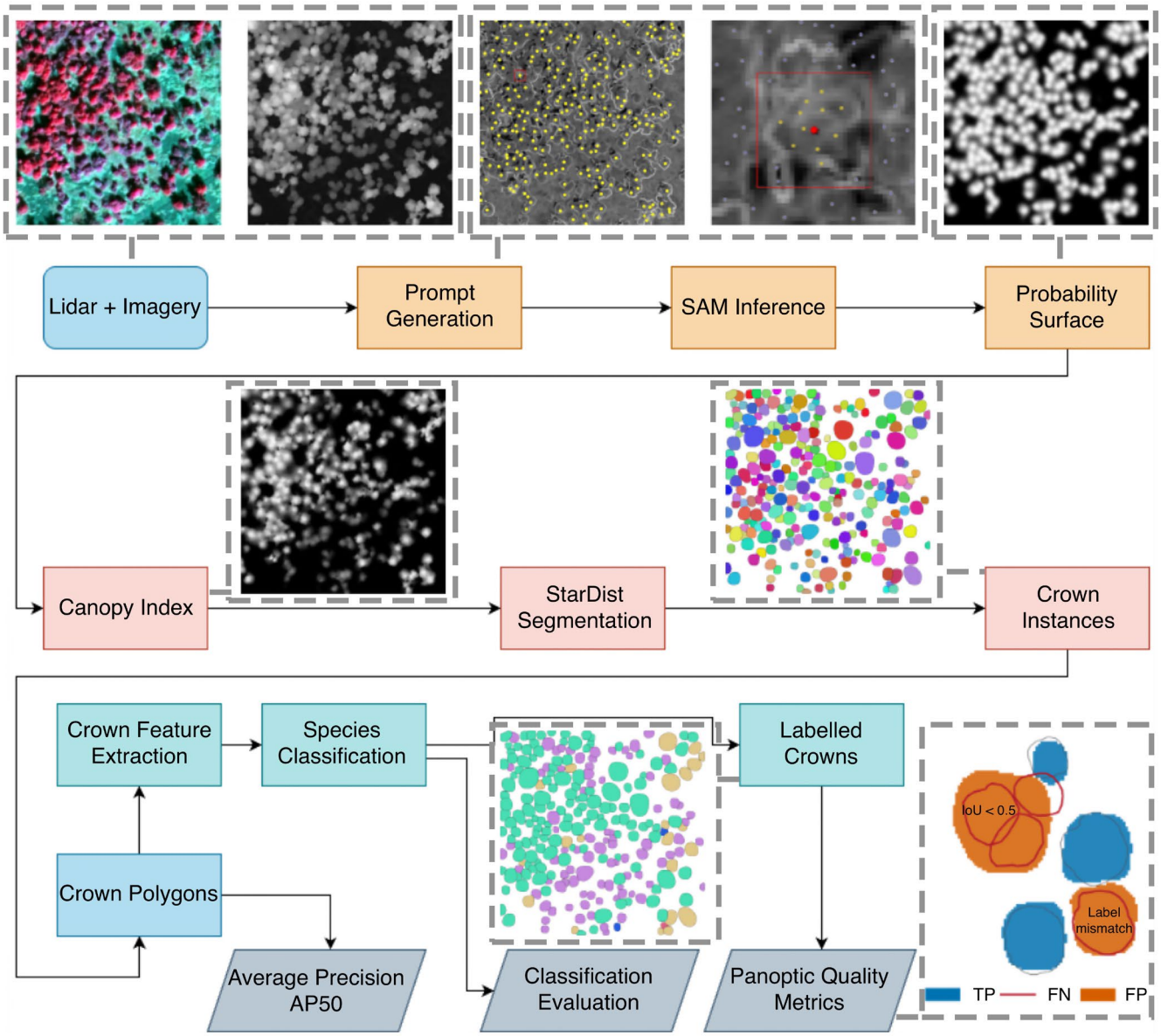


FIGURE 1 | Overview of sam-stars delineation and the downstream object-based XGBoost species classifier.

composition (Section 2.4). The split was made within this set of annotation plots before model fitting and held constant across models, with up to 2805 crowns used for training and 1241 withheld for final validation; a small separate subset was reserved for model tuning. Because training and validation crowns came from the same study landscape and sensor acquisitions, results represent held-out within-landscape performance rather than a spatially independent transfer test. We refer to the full training split as the high-label regime and use it for stratified robustness analyses and panoptic decomposition. We compare sam-stars delineation, a Detectron2 Mask R-CNN instance-segmentation baseline, and a Detectron2 panoptic baseline that jointly predicts crown instances and species labels; model details are described in the Supplement.

Delineation performance is evaluated using polygon-based mask average precision (AP) with one-to-one intersection-over-union (IoU; overlap area divided by union area) matching between predicted and reference crowns. We report AP50 as the primary delineation metric, with full metric definitions provided in the Supplement. To assess annotation efficiency, segmentation models are trained using larger subsets of the training data and evaluated on the same validation set, summarizing performance as AP50 as a function of training set size. For this label-efficiency sweep, we train models on nested subsets of the training split, increasing the number of unique crowns up to the full set.

To examine robustness to forest structure, we stratify validation crowns by canopy height and crown density and compare AP50 across strata using paired nonparametric tests. Canopy height was derived from the lidar CHM and split into ≤ 10 m versus >10 m, with 10 m chosen as a rounded value near the dataset median. Local crown density was the proportion of reference crown area within 15 m of each crown centroid; this radius captures the immediate neighborhood, roughly several crown diameters, while avoiding broader plot-scale gradients. Density values were split at the median. Species classification performance is evaluated using overall accuracy and macro-averaged

20563485, 0, Downloaded from https://zslpublications.onlinelibrary.wiley.com/doi/10.1002/rse2.70117, Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

precision, recall, and F1 on matched prediction–reference crown pairs (IoU ≥ 0.5). To isolate the effect of crown delineation on species attribution, the same trained classifier is applied to sam-stars and Detectron2 crowns.

Finally, we quantify end-to-end labeled-crown performance using species-aware panoptic quality (PQ), where a true positive requires both IoU ≥ 0.5 and a correct species label (Figure S1). We report PQ and its components—segmentation quality (SQ) and recognition quality (RQ)—along with macro-F1 for matched crowns. We also measure end-to-end delineation inference time on the fixed validation subset, reported as median seconds per km^2 and per predicted crown.

2.4 | Study Area, Data, and Landscape-Scale Application

We evaluated sam-stars in the Pu'uwa'awa'a ahupua'a on Hawai'i Island, USA (19°44′ N, 155°51′ W), a ~16,281 ha mountain-to-sea land division spanning large gradients in elevation, rainfall, and canopy structure (Figure 2). The area includes open kiawe (*Prosopis pallida*)-dominated savannas, heterogeneous mid-elevation forests, and dense upper montane koa (*Acacia koa*)–'ōhi'a (*Metrosideros polymorpha*) stands, providing diverse canopy conditions for both controlled benchmarking and landscape-scale application.

Primary data inputs consisted of pan-sharpened 0.31 m WorldView-3 multispectral imagery (two cloud-free scenes acquired in December 2020) and airborne lidar from the NOAA/USGS Hawai'i Quality Level-1 Lidar Project (≥ 8 pts. m^{-2}; acquired 2018–2019) (NOAA Office for Coastal Management 2020). All datasets were co-registered prior to analysis; preprocessing and co-registration procedures are described in the Supplement.

Reference data comprise two complementary annotation sets: (i) 4141 manually delineated tree crowns with species labels used for controlled benchmarking and (ii) 10,843 geo-referenced, photo-verified species inventory points used for landscape-scale species mapping (Figure 2). Crown boundaries were delineated from co-registered WorldView-3 imagery and lidar CHMs, with RTK-GNSS UAV orthomosaics (~1 cm GSD) used where available to resolve edges and overlaps. Species labels came from field observations and georeferenced ground or oblique drone photographs; these photos supported labeling and quality control but were not model inputs.

We restrict species classification to the five most abundant species in the study area, with all remaining taxa pooled into

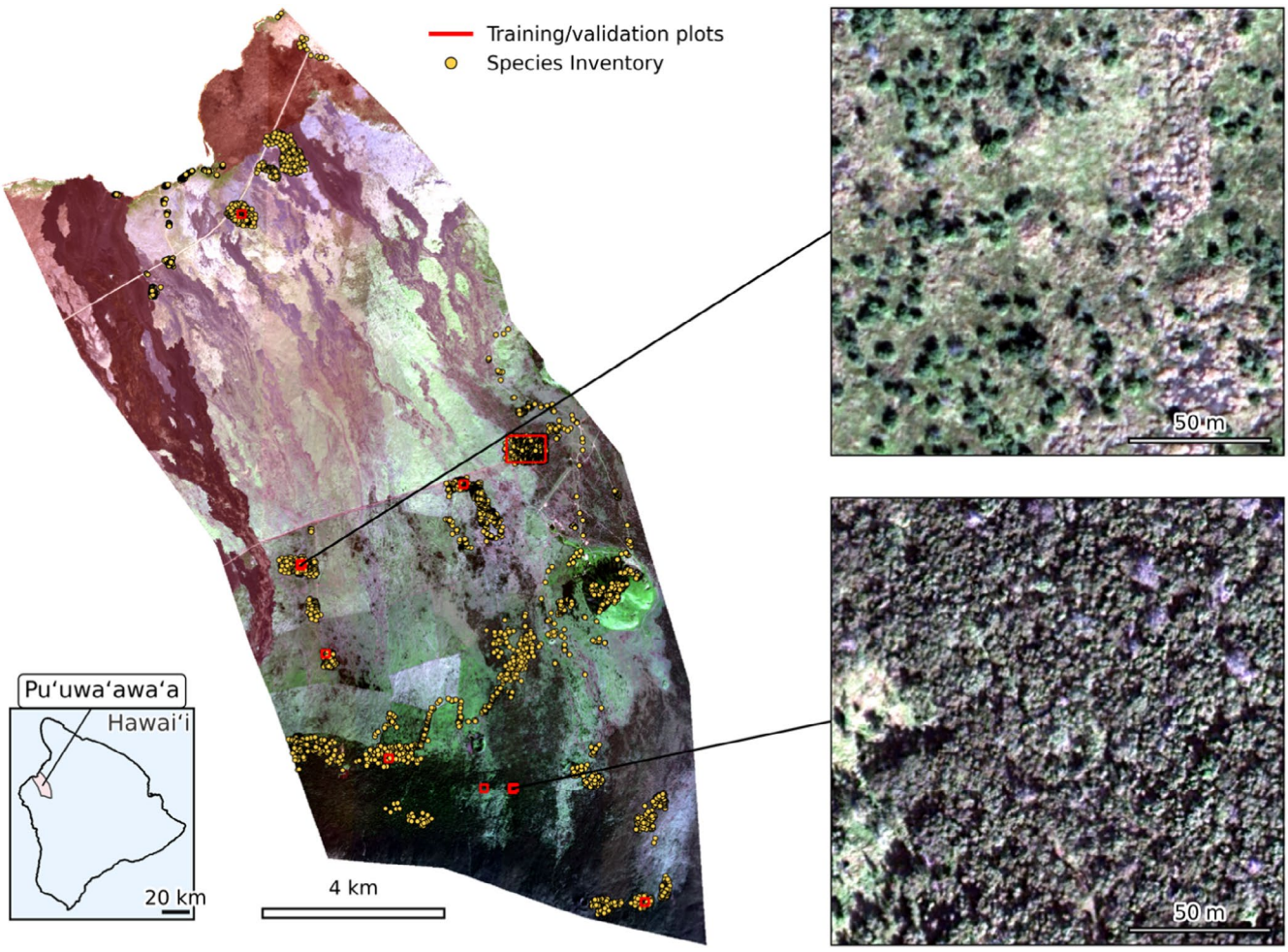


FIGURE 2 | Pu'uwa'awa'a study area on Hawai'i Island with crown-annotation plots (red) and species inventory points (yellow); insets show example canopy densities.

20563485, 0, Downloaded from https://zslpublications.onlinelibrary.wiley.com/doi/10.1002/rse2.70117, Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

a sixth "Other" class (Figure 3). These classes are METPOL (*Metrosideros polymorpha*), GREROB (*Grevillea robusta*), ACAKOA (*Acacia koa*), DIOSAN (*Diospyros sandwicensis*), PROPAL (*Prosopis pallida*), and OTHER.

Finally, to demonstrate landscape-scale deployment, we apply a sam-stars delineation model trained on 254 manually delineated crowns to the full ahupua'a. This small subset was a deliberately low-label deployment scenario; the 4141-crown dataset was retained for controlled train-size benchmarking and held-out evaluation. For species mapping, photo-verified inventory points were assigned to the delineated crown polygon that contains each point, producing point-labeled crown segments used to train the classifier. The classifier was then applied to all delineated crowns. We report performance from a species-stratified 70/30 train/test split of point-labeled segments; postprocessing is provided in the Supplement. Crown polygons are aggregated to estimate woody canopy cover and compared with the 30 m Carbon Assessment of Hawai'i land-cover map (Jacobi et al. 2017).

3 | Results

3.1 | Segmentation Performance and Label Efficiency

We evaluated crown delineation on the held-out validation set using polygon-based mask average precision at IoU ≥ 0.5 (AP50; Section 2.3). Across the training-size sweep, sam-stars achieves higher AP50 than Detectron2 at low annotation levels, with the largest gains at approximately 50–250 unique training crowns (Figure 4). Detectron2 improves more gradually, whereas sam-stars maintains strong delineation performance near the 254-crown deployment label budget (Figure 4).

Using the high-label regime model, sam-stars achieved higher AP50 for tall crowns (> 10 m) than short crowns ($\leq$ 10 m; Holm-adjusted $p < 0.001$), whereas the height effect was not significant for Detectron2. AP50 was lower in high-density than low-density canopy for both methods, although this decline was significant only for Detectron2 (Holm-adjusted $p = 0.024$). sam-stars outperformed Detectron2 in both density classes (Holm-adjusted $p < 0.001$; Figure 5).

Median inference throughput on the fixed evaluation subset was 130 s km^{-2} (0.0195 s crown^{-1}) for Detectron2 and 202 s km^{-2} (0.0303 s crown^{-1}) for sam-stars.

3.2 | Species Classification Performance on Matched Crowns

On geometrically matched crowns (IoU ≥ 0.5) in the high-label regime, sam-stars + XGBoost achieves the highest overall labeling performance (macro-F1 = 0.81), followed by Detectron2 + XGBoost (0.78) and Detectron2-full (0.71)

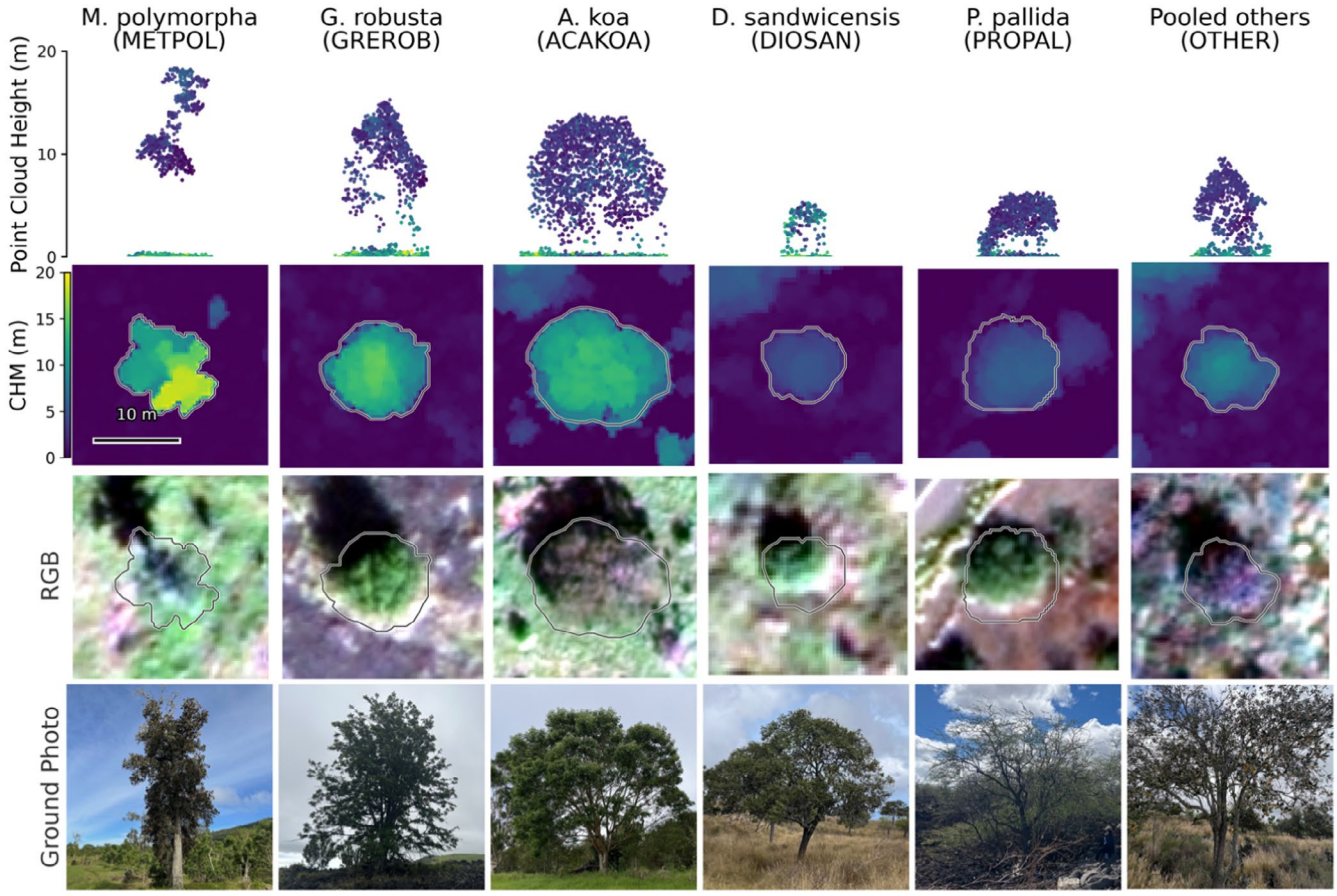


FIGURE 3 | Example crowns for the six evaluation classes with lidar point clouds, CHM, WorldView-3 RGB imagery, and ground photographs (outlines indicate reference polygons).

20563485, 0, Downloaded from https://zslpublications.onlinelibrary.wiley.com/doi/10.1002/rse2.70117, Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

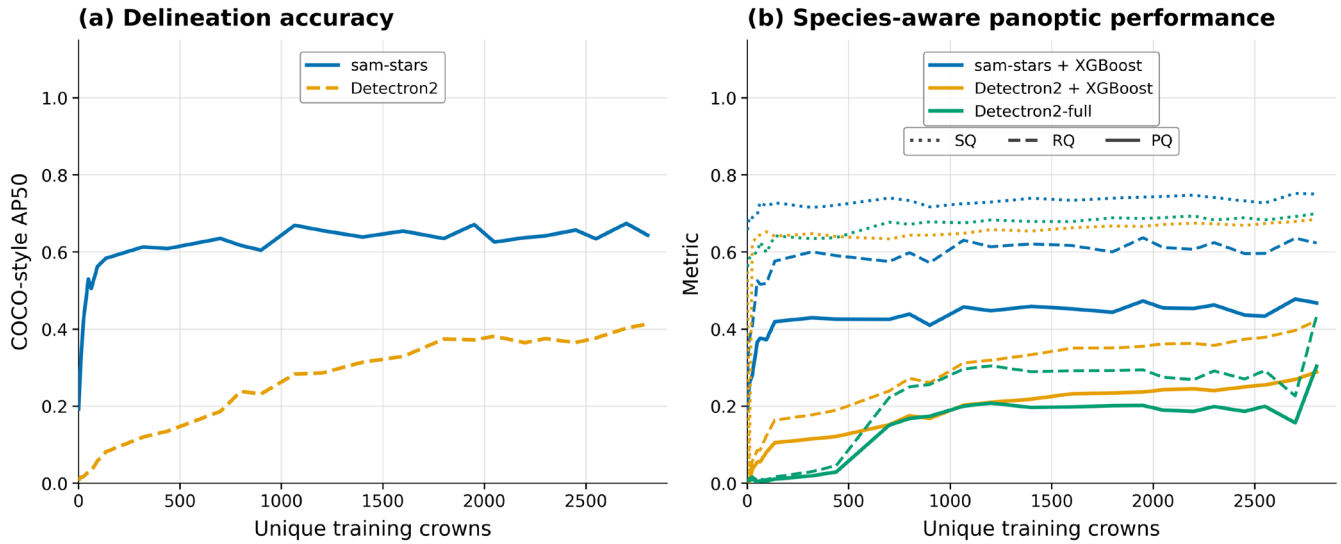


FIGURE 4 | Label-efficiency curves: AP50 versus training size (left) and species-aware SQ/RQ/PQ versus training size (right).

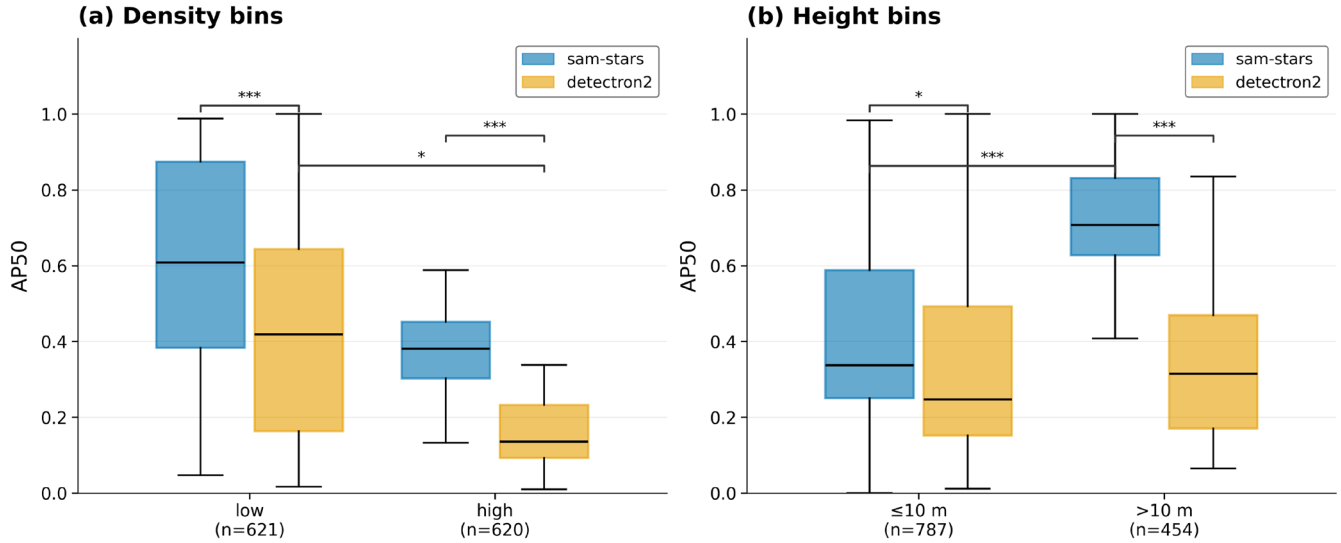


FIGURE 5 | AP50 stratified by canopy height and local crown density for the high-label regime.

(Figure 6). Across methods, the pooled OTHER class is the most difficult to classify, whereas PROPAL and ACAKOA are among the most reliably under sam-stars + XGBoost.

3.3 | End-to-End Labeled-Crown Evaluation (Semantic Panoptic Metrics)

We evaluated end-to-end labeled-crown quality using species-aware panoptic quality (PQ), where a true positive requires both IoU ≥ 0.5 and a correct species label. At $n = 2805$ training crowns, sam-stars achieved the highest SQ (0.75), compared with Detectron2 + XGBoost (0.68) and Detectron2-full (0.70). Differences in PQ reflect both segmentation and recognition quality: sam-stars achieves the highest RQ (0.62) and therefore the highest PQ (0.47), exceeding Detectron2 + XGBoost (0.29) and Detectron2-full (0.30) (Figure 7). Most false negatives arise from merged adjacent crowns in dense canopy, where a single prediction overlaps multiple references; other false positives are concentrated along crown boundaries and in shadowed or mixed pixels (Figure 8; Figure S2).

3.4 | Landscape-Scale Mapping Results and Comparison With Existing Products

At landscape scale, a delineation model trained on 254 crowns produced 333,204 crowns across the ~16,281 ha ahupua'a, covering 3091.1 ha of canopy area (19.0% of the AOI; Figure 9a). The six-class species model labeled all crowns, with METPOL dominant followed by ACAKOA and PROPAL (Figure 9b,c). On the held-out test split of the point-labeled crown dataset, the classifier achieved overall accuracy = 0.91 and macro-F1 = 0.91.

20563485, 0, Downloaded from https://zslpublications.onlinelibrary.wiley.com/doi/10.1002/rse2.70117, Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

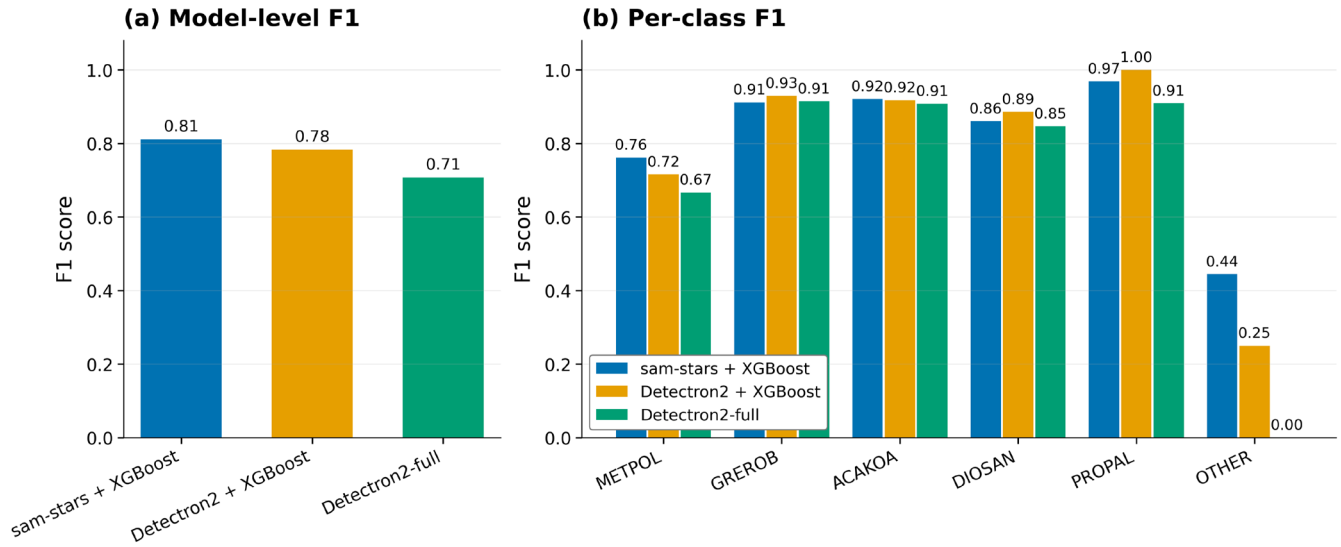


FIGURE 6 | Model-level macro-F1 and per-class species F1 on matched crowns (IoU ≥ 0.5) in the high-label regime.

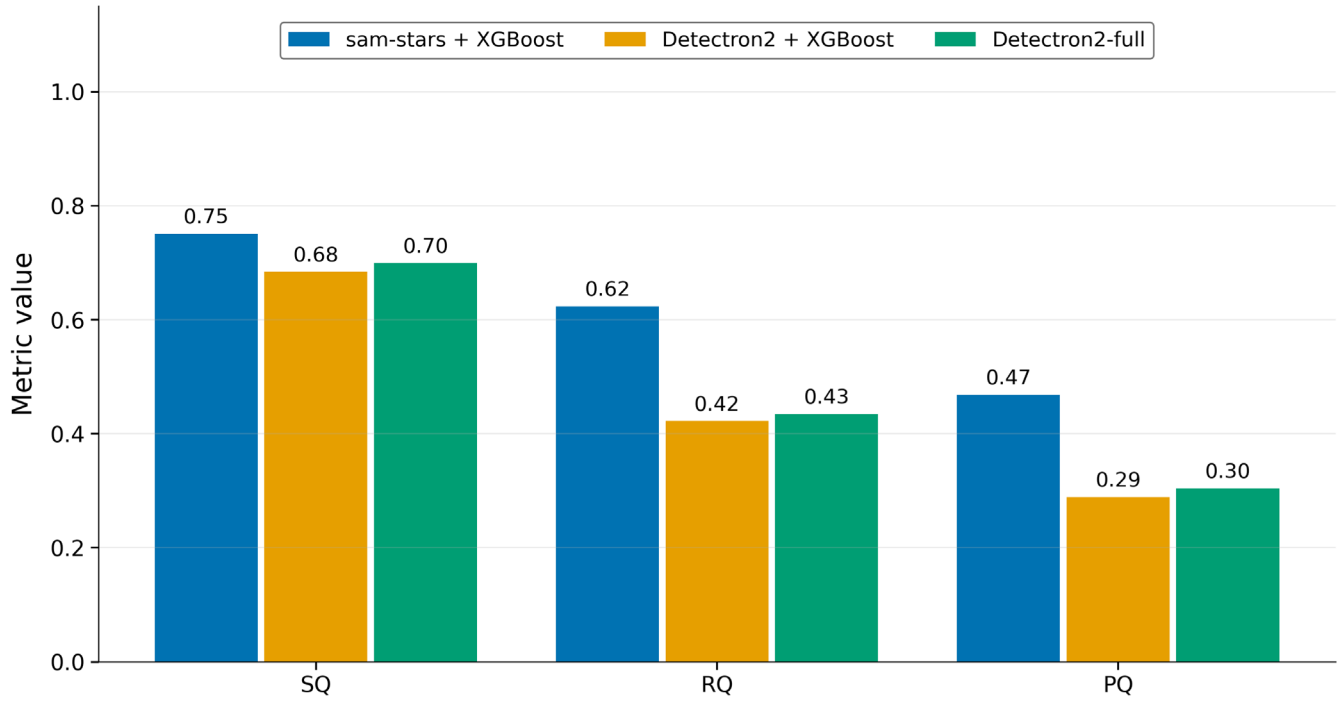


FIGURE 7 | Semantic panoptic metrics (SQ, RQ, PQ) at IoU ≥ 0.5 for the high-label regime.

When aggregated to 30 m resolution, the crown-based map estimates 19.0% canopy cover versus 14.9% in the CAH land-cover product—approximately 27% more canopy detected (Figure 10). Differences are spatially heterogeneous, with the largest underestimation by CAH in heterogeneous mid-elevation forests with many small crowns.

4 | Discussion

Across benchmarks in heterogeneous forests on Hawai'i Island, sam-stars improves label efficiency for crown delineation and yields higher species-aware labeling performance than Detectron2 baselines at small training sizes. In a landscape-scale application, a delineation model trained on 254 crowns produced over 333,000 crown instances across ~16,300 ha, and crown-derived canopy cover exceeded the CAH land-cover estimate by ~27%. This discrepancy reflects underdetection of small or fragmented canopy in coarse products, particularly in heterogeneous and open-canopy settings, and is consistent with broader evidence that scattered trees and fine-scale canopy structure are poorly captured by medium-resolution land-cover and biomass maps (Reiner et al. 2023; Liu et al. 2025).

The sam-stars workflow deliberately decouples canopy detection from instance separation and delineation from labeling. Lidar-guided prompting allows SAM to recover candidate crown geometry with limited polygon supervision, while StarDist resolves

20563485, 0, Downloaded from https://zslpublications.onlinelibrary.wiley.com/doi/10.1002/rse2.70117, Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

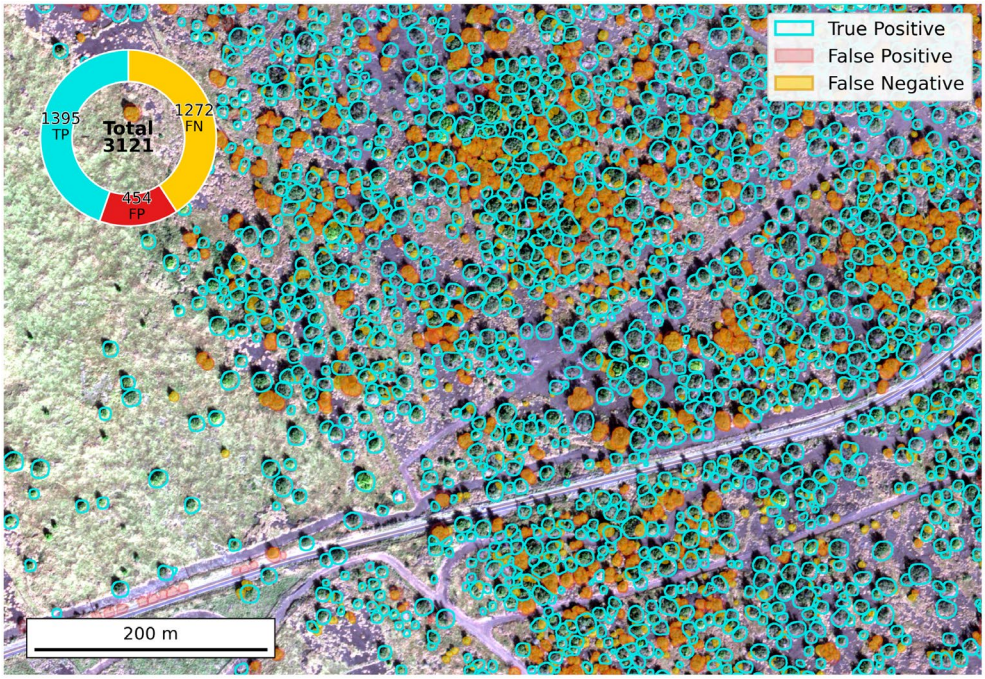


FIGURE 8 | Full-tile example showing true positive (TP), false positive (FP), and false negative (FN) outcomes. A TP requires IoU ≥0.5 and the correct species label.

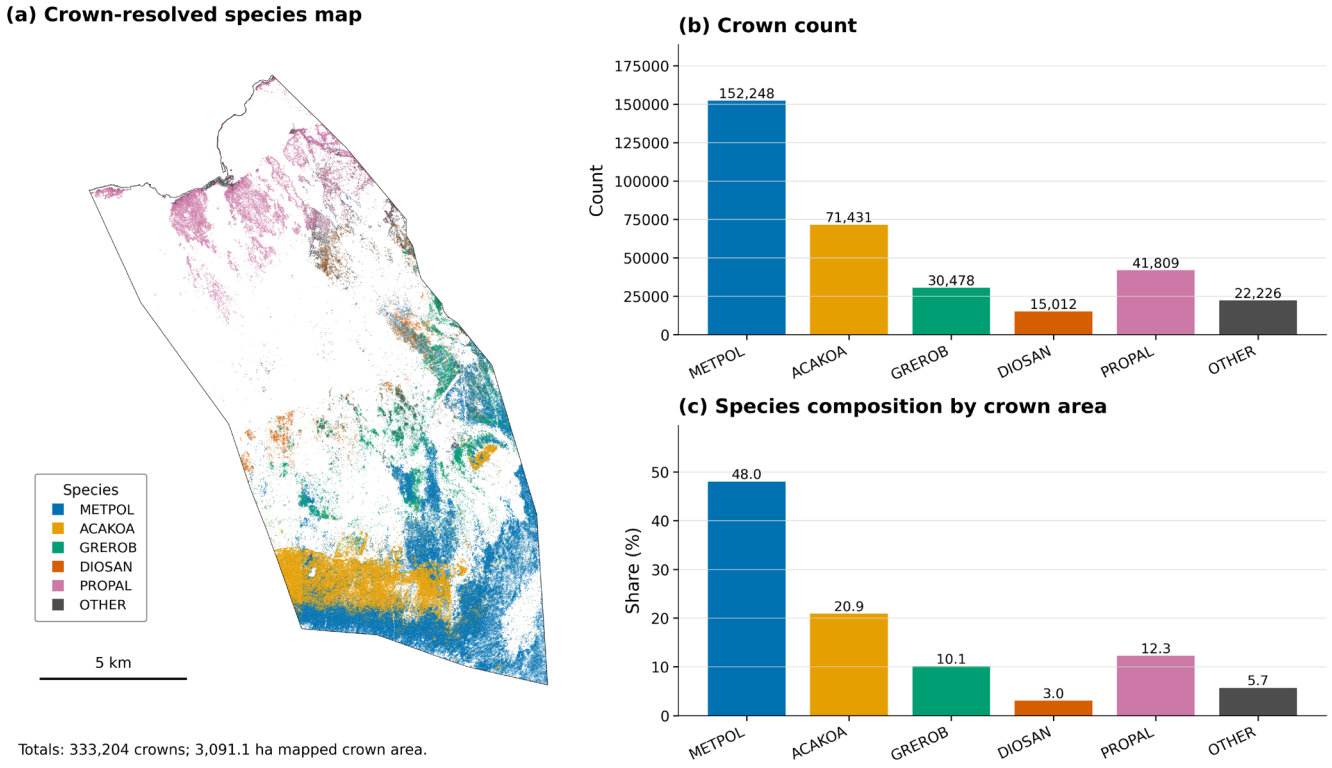


FIGURE 9 | Landscape-scale crown and species map: (a) mapped crowns, (b) crown counts by class, (c) area by class.

adjacent crowns into nonoverlapping instances. Separating species attribution allows crown geometry to be reused as inventories expand.

Performance gains are most pronounced early in the training-size sweep (≈50–250 crowns), supporting low-label deployment when annotation budgets are limited. Dense, overlapping canopy remains the dominant failure regime for both sam-stars and Detectron2, with most errors arising from merged adjacent crowns under one-to-one IoU matching rather than from missed detections. By contrast, sparse and open-canopy conditions, where individual crowns are more isolated and well separated, are well suited to the sam-stars workflow. Thus, a practical limitation is that sam-stars is best suited to crowns that are at least

20563485, 0, Downloaded from https://zslpublications.onlinelibrary.wiley.com/doi/10.1002/rse2.70117, Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

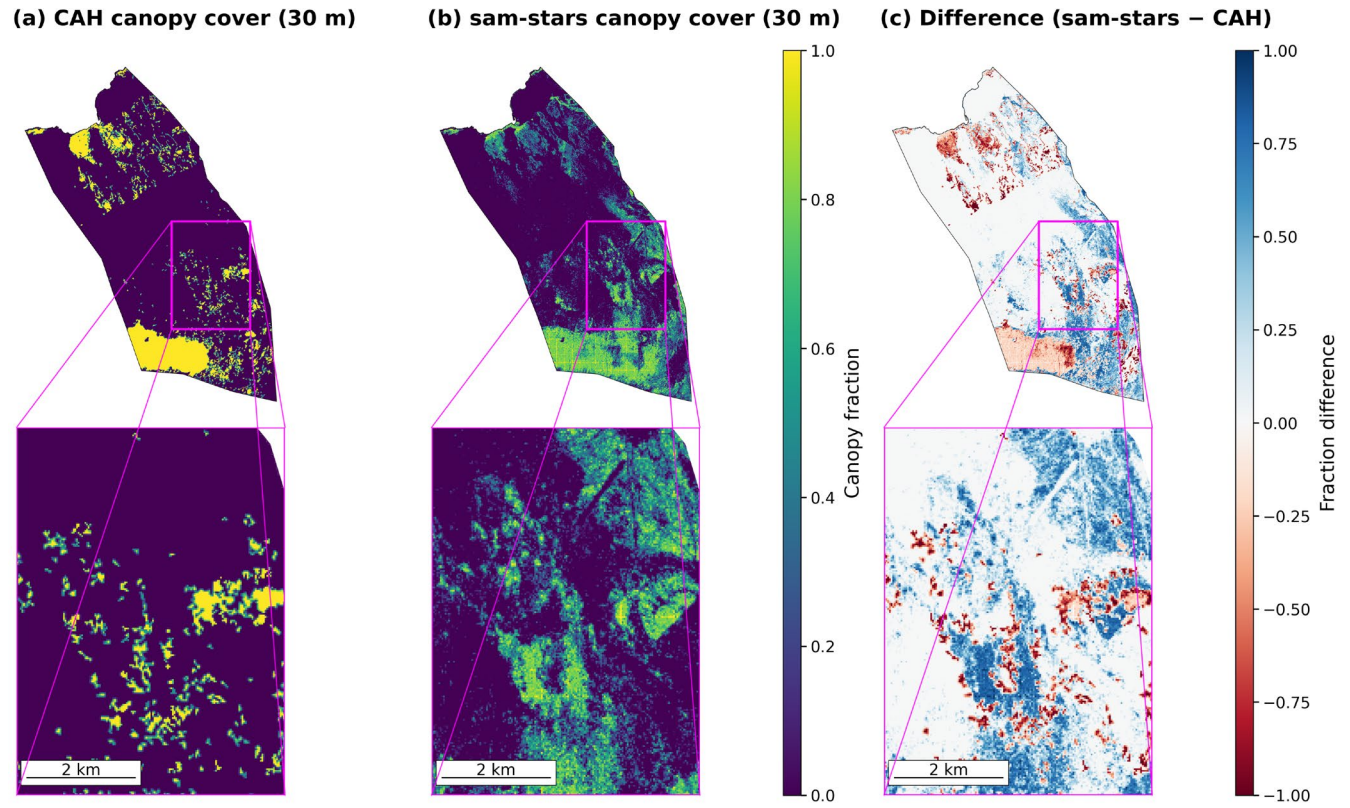


FIGURE 10 | Canopy cover comparison with CAH at 30 m resolution: CAH-derived canopy fraction, crown-derived canopy fraction, and difference (crowns—CAH).

partially separable in the CHM or optical imagery. Relatively isolated crowns and open or heterogeneous canopies provide clearer local maxima and background prompts, whereas dense, overlapping canopies can produce merged crowns, ambiguous boundaries, and lower recognition quality.

The improved species-aware panoptic quality achieved by sam-stars is driven more by recognition quality than by segmentation overlap. Applying the same species classifier to sam-stars and Detectron2 crown sets changes the macro-F1 modestly (≈0.03), while the larger difference in recognition quality indicates that the main benefit is more consistent instance recovery. This gain comes with slower inference relative to Detectron2 because of prompt-guided SAM inference and StarDist separation. In many deployments, however, manual crown digitizing and field support dominate total effort, so reduced delineation labels can outweigh slower runtime. Applying the sam-stars workflow in other landscapes requires accurately co-registered VHR imagery and lidar-derived canopy-height data. A modest local crown reference set is needed primarily to validate delineation performance and support site-specific fine-tuning where necessary; species mapping additionally requires local species observations to train the classifier. Because image geometry, shadowing, canopy structure, and species composition vary among landscapes, performance should be evaluated locally. With appropriate local validation, sam-stars provides a practical, label-efficient approach to crown-resolved species mapping across heterogeneous forest landscapes.

Author Contributions

Joseph Emile Honour Percival: conceptualization, methodology, software, validation, visualization, writing – original draft, formal analysis, data curation, investigation. **Kanoa Lindiwe:** data curation, software, writing – review and editing. **Kalena Shiroma:** data curation, writing – review and editing. **Adrian Pascual:** data curation, writing – review and editing. **Juan Guerra-Hernandez:** data curation, writing – review and editing. **Rebecca Ostertag:** writing – review and editing, conceptualization, supervision, funding acquisition. **Ryan Perroy:** conceptualization, writing – review and editing, supervision, funding acquisition, resources, project administration. Intellectual Contributions: Regional stakeholders and domain experts reviewed interim and final mapping products; their feedback helped refine labels, check outputs, and interpret the classified map.

Acknowledgments

We thank Lucas Fortini for EagleView imagery access. We are grateful to Edith Atkins (Pu‘uwa‘awa‘a) for site knowledge and map feedback, and Jonathan Price for Hawai‘i vegetation expertise and training-data/model-output checks. This work was enabled in part by funding from the National Science Foundation award 2149133. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Funding

This work was supported by the National Science Foundation (Grant No. 2149133).

20563485, 0, Downloaded from https://zslpublications.onlinelibrary.wiley.com/doi/10.1002/rse2.70117, Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Conflicts of Interest

One of the authors is an Associate Editor of Remote Sensing in Ecology and Conservation. This author had no involvement in the editorial handling of this manuscript, including editor assignment, peer review, or decision-making. The remaining authors declare no conflicts of interest.

Data Availability Statement

The lidar data used in this study are publicly available from NOAA Digital Coast (NOAA Office for Coastal Management 2020). WorldView-3 imagery is commercially licensed and cannot be redistributed. Derived training/evaluation datasets and outputs are provided at https://doi.org/10.5281/zenodo.17193938, and the sam-stars library code is hosted at https://github.com/iosefa/samstars.

References

Avitabile, V., M. Herold, G. B. M. Heuvelink, et al. 2016. "An Integrated Pan-Tropical Biomass Map Using Multiple Reference Datasets." *Global Change Biology* 22: 1406–1420. https://doi.org/10.1111/gcb.13139.

Dalponte, M., and D. A. Coomes. 2016. "Tree-Centric Mapping of Forest Carbon Density From Airborne Laser Scanning and Hyperspectral Data." *Methods in Ecology and Evolution* 7: 1236–1245. https://doi.org/10.1111/2041-210X.12575.

Fassnacht, F. E., H. Latifi, K. Stereńczak, et al. 2016. "Review of Studies on Tree Species Classification From Remotely Sensed Data." *Remote Sensing of Environment* 186: 64–87. https://doi.org/10.1016/j.rse.2016.08.013.

Jacobi, J. D., J. P. Price, L. B. Fortini, S. M. Gon, and P. Berkowitz. 2017. "Baseline Land Cover, Baseline and Projected Future Carbon Storage and Carbon Fluxes in Ecosystems of Hawaii." In *Baseline and Projected Future Carbon Storage and Carbon Fluxes in Ecosystems of Hawai'i*, edited by P. C. Selmants, C. P. Giardina, J. D. Jacobi, and Z. Zhu, 9. U.S. Geological Survey Professional Paper.

Kirillov, A., E. Mintun, N. Ravi, et al. 2023. "Segment Anything." In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, 3992–4003. https://doi.org/10.1109/ICCV51070.2023.00371.

Liu, S., J. Zhang, L. Wang, et al. 2025. "Mapping Previously Undetected Trees Reveals Overlooked Changes in Pan-Tropical Tree Cover." *Nature Communications* 16: 5561. https://doi.org/10.1038/s41467-025-60662-z.

NOAA Office for Coastal Management. 2020. *2018–2020 NOAA USGS Lidar: Hawai'i, HI*. National Oceanic and Atmospheric Administration, Office for Coastal Management. https://www.fisheries.noaa.gov/inport/item/68082.

Reiner, F., M. Brandt, X. Tong, et al. 2023. "More Than One Quarter of Africa's Tree Cover Is Found Outside Areas Previously Classified as Forest." *Nature Communications* 14: 2258. https://doi.org/10.1038/s41467-023-37880-4.

Schmidt, U., M. Weigert, C. Broaddus, and G. Myers. 2018. "Cell Detection With Star-Convex Polygons." In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018. Lecture Notes in Computer Science*, edited by A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, 265–273. Springer International Publishing. https://doi.org/10.1007/978-3-030-00934-2_30.

Teng, M., A. Ouaknine, E. Laliberté, Y. Bengio, D. Rolnick, and H. Larochelle. 2025. *Assessing SAM for Tree Crown Instance Segmentation from Drone Imagery*. https://doi.org/10.48550/arXiv.2503.20199.

Tong, F., and Y. Zhang. 2025. "Individual Tree Crown Delineation in High Resolution Aerial RGB Imagery Using StarDist-Based Model." *Remote Sensing of Environment* 319: 114618. https://doi.org/10.1016/j.rse.2025.114618.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** Species-aware panoptic evaluation schematic. A true positive (TP) requires IoU ≥ 0.5 and the correct species label; false positives (FP) are unmatched or incorrectly labeled predictions, and false negatives (FN) are unmatched reference crowns. **Figure S2:** Common error modes under crown overlap and occlusion. Examples show merged adjacent crowns, ambiguous crown boundaries, and false positives concentrated along boundaries or shadowed/mixed pixels.

20563485, 0, Downloaded from https://zslpublications.onlinelibrary.wiley.com/doi/10.1002/rse2.70117, Wiley Online Library on [16/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License