Article

Seeing the Forest for the Trees: Mapping Cover and Counting Trees from Aerial Images of a Mangrove Forest Using Artificial Intelligence

Daniel Schürholz [1,2,*], Gustavo Adolfo Castellanos-Galindo [3,4,5], Elisa Casella [2,6], Juan Carlos Mejía-Rentería [7] and Arjun Chennu [2]

1. Max Planck Institute for Marine Microbiology, 28359 Bremen, Germany
2. Leibniz Centre for Tropical Marine Research (ZMT), 28359 Bremen, Germany; arjun.chennu@leibniz-zmt.de (A.C.)
3. Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), 12587 Berlin, Germany
4. Institute of Biology, Freie Universität Berlin, 14195 Berlin, Germany
5. Smithsonian Tropical Research Institute, Balboa 0843, Panama
6. Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, 30172 Venice, Italy
7. Grupo de Investigación en Ecología de Estuarios y Manglares, Departamento de Biología, Universidad del Valle, Cali 25360, Colombia
* Correspondence: daniel.schuerholz@leibniz-zmt.de

Abstract: Mangrove forests provide valuable ecosystem services to coastal communities across tropical and subtropical regions. Current anthropogenic stressors threaten these ecosystems and urge researchers to create improved monitoring methods for better environmental management. Recent efforts that have focused on automatically quantifying the above-ground biomass using image analysis have found some success on high resolution imagery of mangrove forests that have sparse vegetation. In this study, we focus on stands of mangrove forests with dense vegetation consisting of the endemic *Pelliciera rhizophorae* and the more widespread *Rhizophora mangle* mangrove species located in the remote Utría National Park in the Colombian Pacific coast. Our developed workflow used consumer-grade Unoccupied Aerial System (UAS) imagery of the mangrove forests, from which large orthophoto mosaics and digital surface models are built. We apply convolutional neural networks (CNNs) for instance segmentation to accurately delineate (33% instance average precision) individual tree canopies for the *Pelliciera rhizophorae* species. We also apply CNNs for semantic segmentation to accurately identify (97% precision and 87% recall) the area coverage of the *Rhizophora mangle* mangrove tree species as well as the area coverage of surrounding mud and water land-cover classes. We provide a novel algorithm for merging predicted instance segmentation tiles of trees to recover tree shapes and sizes in overlapping border regions of tiles. Using the automatically segmented ground areas we interpolate their height from the digital surface model to generate a digital elevation model, significantly reducing the effort for ground pixel selection. Finally, we calculate a canopy height model from the digital surface and elevation models and combine it with the inventory of *Pelliciera rhizophorae* trees to derive the height of each individual mangrove tree. The resulting inventory of a mangrove forest, with individual *P. rhizophorae* tree height information, as well as crown shape and size descriptions, enables the use of allometric equations to calculate important monitoring metrics, such as above-ground biomass and carbon stocks.

Keywords: mangrove forests; forest inventory; monitoring; habitat mapping; UAV; UAS; artificial intelligence; machine learning; instance segmentation; semantic segmentation; above ground biomass; carbon stock

1. Introduction

For the past decades, the global area covered by mangrove forests has receded because of direct and indirect anthropogenic causes such as land use changes, deforestation, pollution and climate change [1]. The potential impacts of the disappearance of mangrove forests to local communities and adjacent ecosystems are manifold due to the critical services that these forests provide (coastal protection [2], fish nurseries [3], feeding grounds [4], carbon sequestration [5], etc.). The urgency of the current state of affairs has lead to the launch of many protection, rehabilitation and reforestation efforts of mangrove forests worldwide [6,7]. For these efforts to succeed, careful observation and detailed analysis of forest conditions are required to identify problems, calibrate predictive models and enact mitigatory management actions [8].

The condition of most forests can be assessed on different scales: individual trees, the collection of trees in a forest stand or the complete forest ecosystem (considering biotic and abiotic factors) [9]. An individual tree can be assessed in the field through many indicators such as nutritional status, presence of parasites/pathogens, crown transparency, diameter at breast height (DBH), crown length and crown width (m), to provide a few examples. Then, these indicators are collected for trees in several plots, aggregating the measurements in inventories and extrapolating for trees onto the forest stand. Creating inventories of a forest enables certain ecosystem indicators to be derived, which can be its biomass (above- and below-ground), canopy structure, tree species composition and community structure [10,11]. For example, to calculate the above ground biomass (AGB) for a forest using allometric equations, the following variables must be collected for each individual tree: its species, height, DBH [12] and, to calculate the canopy structure, the crown size and shape must be acquired.

The manual in situ measurement of these variables is a labor-intensive task when a forest of several hectares is surveyed, even with advances in on-ground sensing technologies [13,14]. Thus, a limited number of small plots are surveyed depending on the aim, the sampling costs, the extent of the forest, the tree sizes and species diversity found in a patch of forest (e.g., 35×35 m plots for trees over 50 cm DBH) [15]. There is a trade-off between the sampling cost and the accepted uncertainties that appear when extrapolating the measurements to the complete forest area [16]. Recent studies suggest that field surveys entail significant errors in measurement and plot positions [16,17]. As in other intertidal systems, in-situ plot measurements in mangrove forests can be difficult to execute, given that tidal regimens, muddy terrain, pneumatophores and stilt roots, remote locations and other factors severely reduce the accessibility. Furthermore, DBH can be difficult to measure for some mangrove species (i.e., *Rhizophora mangle*), due to their complex trunk-growing structure [18], and correct crown size and shape is difficult to measure visually, given the irregular shape and clumpiness of the canopies [19].

In recent decades, researchers have used fly-over strategies to capture plane-view images of forests to use for inventory creation. This has been fueled by the advancements in remote sensing, image analysis and machine learning. These advancements have enabled analyses of mangrove forests and their dynamics across vast scales [20–22]. In these studies, spectral indices, such as normalized difference vegetation index, are calculated for each pixel to describe and classify mangrove forests, being able to label the tree species and tree density within a pixel, as well as canopy width and forest fragmentation [22]. The benefits of Earth-observation technologies are the large spatial coverage and frequent acquisition of images. Paired with machine learning automation, studies of long time-series of images can be carried out. Recent improvements in satellite image resolutions (i.e., 0.031 m for the World-View 3 satellite) have allowed for more resolved classification of trees using semantic segmentation neural networks [23,24], detection of individual trees using instance segmentation networks [25–28] and detection of mangrove forest clearings [29] on high-resolution RGB images. Nonetheless, the calculation of certain variables, such as the height of trees extracted from canopy height models (CHMs) is error-prone at the current

resolution of satellite imagery and should be paired with low-flying platforms, such as planes or UASs [28] for better validation and performance.

Several recent studies have pointed out and demonstrated the value offered by UASs for monitoring coastal environments, such as mangrove forests [30–33]. The imagery taken with UASs can be processed with structure from motion (SfM) software to produce geo-referenced orthorectified photo-mosaics (orthomosaics) and digital surface models (DSMs). Paired with novel image segmentation techniques, precise area coverage of individual tree species in a forest are determined and other surrounding land cover classified (i.e., grass, shrubs, water, sand, mud, etc.) [34,35]. Certain terrain classes such as mud and sand are used to calculate the height of forest canopies or of individual trees by subtracting their elevation from the elevation of trees in the DSM [36,37]. Furthermore, using hyperspectral and multispectral cameras yielding high-dimensional input data, the area covered by multiple tree species in a forest can be accurately segmented [38]. Individual tree crown segmentation, delineation and classification can be facilitated by the advancement of machine learning algorithms on the high resolution RGB and LiDAR images of low-flying platforms [39]. Recent studies segmented mangrove trees in forest plots using images from RGB or LiDAR sensors mounted on a consumer-grade UASs together with object-based image analysis (OBIA) algorithms, and compare the predicted segments to on-ground measurements [19,36,40]. Despite the success of OBIA algorithms on UAS images to detect mangrove trees, they rely upon tree crowns that are visually well separated and detailed elevation maps. The potential benefit of state-of-the-art instance segmentation techniques is to handle dense canopies and rely only on imaging data. A recent review [41] of deep learning applications for tree crown segmentation noted the potential of instance segmentation applications, hindered mainly due to the insufficient training data. The development of instance segmentation workflows of high resolution RGB images acquired from consumer-grade UASs is critical to be used as validation for global Earth-observation efforts and as preparation for improved resolution in future satellite sensors.

In this work, we develop and present a complete workflow to delineate individual trees of the *Pelliciera rhizophorae* mangrove species and calculate inventory measurements (i.e., tree height, crown shape and size, geo-location, etc.), as well as map the land cover for other classes: *Rhizophora mangle*, water and mud (see Figure 1). The input data were a set of orthomosaics and DSMs created from images captured with consumer-grade UASs in three mangrove forest stands located in the Utría National Park on the Colombian Pacific coast (Figure 2). We implement two separate deep learning networks: (i) a semantic segmentation neural network to identify area coverage of the two mangrove species, mud and water classes and (ii) an instance segmentation neural network to delineate individual *Pelliciera rhizophorae* mangrove trees. We present a novel tiling/untiling algorithm (from here onwards, we refer to stitching or merging tiles together as "untiling") for the correct preservation of predicted tree instances located at the edges of tiles of large orthomosaics. We also provide a comparison of three different semantic segmentation untiling techniques to resolve the overlapping borders of tiles. We automate the calculation of a CHM, created from a digital elevation model (DEM) using the classified ground pixels and compare it to a DEM created from manually selected ground areas. Finally, using the delineated trees and the CHM, we provide an inventory of the trees in the mangrove forest with their specific height, crown size and crown shape as well as area cover and height distribution values for the other tree classes.
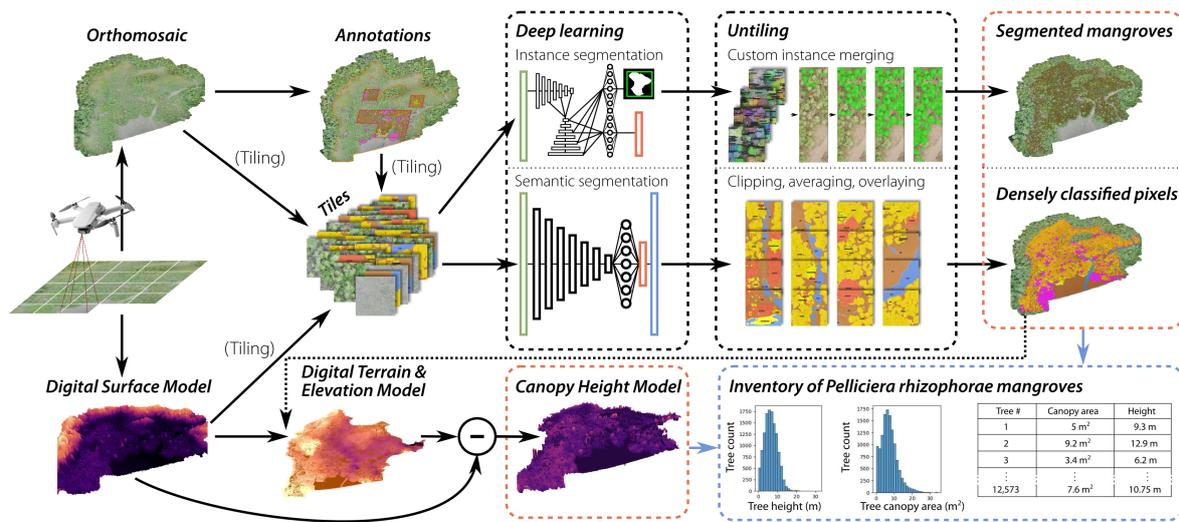
Figure 1. From airborne images to a detailed tree inventory: we present a workflow that creates a tree inventory for 35 hectares of a mangrove forest. The workflow starts with data acquisition using UASs flown over the mangrove forests. We then build top-down orthomosaic and DSM with SfM software. We implement a tiling and a novel untiling process of the orthomosaics and DEM images. Instance segmentation neural networks were used for detecting individual trees and semantic segmentation networks were used to map land cover. Using the classified ground regions, we created and interpolated the digital terrain model (DTM) into a DEM. Subtracting the DEM from the DSM yielded a comprehensive canopy height model. For each automatically segmented tree instance, the tree height was derived from the CHM. This creates an inventory of trees with their heights and crown areas, even in a dense forest canopy, and enables the calculation of above ground biomass, an important measure for monitoring and carbon stock assessments.
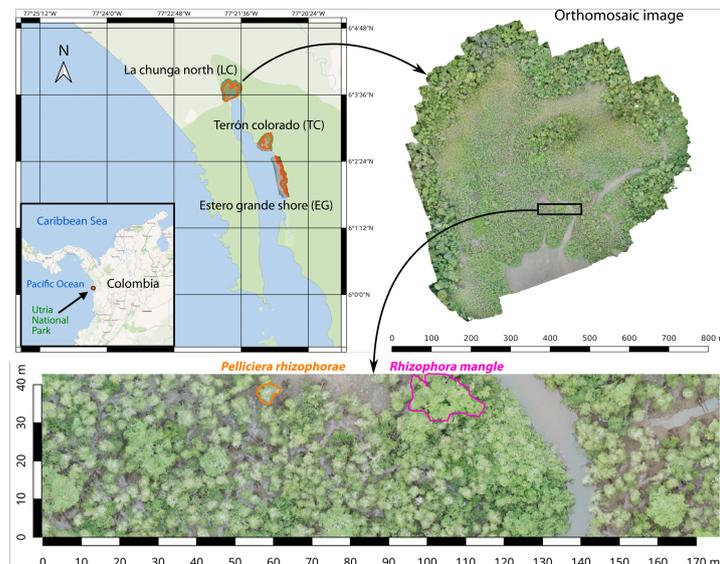
Figure 2. Surveying a dense canopy in a remote forest area: we used a consumer-grade UAS to survey a mangrove forest located in the Utría national park in the Colombian Pacific coast. Three surveyed plots of mangrove forests were used. For each plot large orthomosaic images were created, with fine spatial resolution (e.g., 3.64 cm/pixel) of the underlying mangrove trees. The two dominant species of mangrove trees are the *Pelliciera rhizophorae* species and the *Rhizophora mangle* species. Each of the three plots provide unique challenges for canopy segmentation, given that their conditions differ in ground composition, exposure, tidal level during the survey and lighting/blurring in the images.

2. Materials and Methods

The complete workflow, from data tiling to tree inventory, was developed in the Python programming language, using Snakemake [42] to manage the analytical workflow.

2.1. Study Site and Input Data Structure

We focused on three mangrove forest sites of the Utría National Park: La Chunga North (LCN), Terron Colorado (TC) and Estero Grande Shore (EGS) (see Table 1 for area sizes). These mangrove forests are mainly comprised of two mangrove species: *Pelliciera rhizophorae* and *Rhizophora mangle*. *P. rhizophorae* is endemic to the East Pacific and Caribbean regions and is listed as vulnerable in the International Union for Conservation of Nature (IUCN) Red List for endangered species [43]. It lives in intermediate to upstream estuarine environments with medium to high tidal ranges. The *R. mangle* species is more widespread across the Atlantic/East Pacific bio-geographic region and is listed as of "least concern" in the IUCN Red List for endangered species. It is found in downstream to intermediate estuarine environments with low to medium intertidal shifts.

Table 1. Mangrove forest study sites and digital products details.

	La Chunga North (LCN)	Terron Colorado (TC)	Estero Grande Shore (EGS)
UAS images			
Quantity	289	346	106
Areas			
Surveyed	367,806 m^2	241,752 m^2	425,851 m^2
Mangrove forest	223,456 m^2	120,726 m^2	110,960 m^2
Annotated	50,347 m^2	28,410 m^2	—
Resolutions			
Ortho. image	$19,855 \times 21,068$ px	$16,375 \times 18,923$ px	$10,478 \times 24,485$ px
Ortho. pixel	3.64 cm/px	3.27 cm/px	5.83 cm/px
DSM image	$15,145 \times 15,377$ px	$13,148 \times 13,454$ px	$6759 \times 15,468$ px
DSM pixel	7.29 cm/px	6.55 cm/px	11.7 cm/px
GCPs			
Quantity	3	2	4
RMSE *	0.011 m	0.0097 m	1.13 m
Tiles **			
Total	3304	2438	2070
Annotated	196	168	—

* Root-mean-square error (RMSE) for ground control points (GCP) over all (X,Y,Z) coordinates. ** Tiles of size 512×512 pixels with 30% overlap.

The aerial footage of the sites was captured in 2019 (19–22 February) using two consumer-grade UASs the DJI Phantom 4 and DJI Mavic Pro (SZ DJI Technology Co., Ltd—Shenzhen, China). The DJI Phantom 4 has an integrated photo camera, the DJI FC330, which has a 1/2.3″ CMOS sensor with 12.4 M effective pixels, a focal length of 4 mm, a pixel size of 1.56×1.56 μm and a resolution of 4000×3000 pixels (px). The DJI Mavic Pro was equipped with the integrated DJI FC220 camera with 4000×3000 px resolution, 12.35 M effective pixels and 26 mm wide-angle lens. The flights were programmed to follow the trajectories in an automated mode by means of the commercial app "DroneDeploy". Ground control points (GCPs) were positioned in the field, and their geographic location was acquired. We used two single-band global navigation satellite system (GNSS) receivers: an Emlid Reach RS+ single-band real-time kinematics (RTK) GNSS receiver (Emlid Tech Kft.—Budapest, Hungary) as a base station, and a Bad Elf GNSS Surveyor handheld GPS (Bad Elf, LLC—West Hartford, AZ, USA). RINEX static data from the base station was processed with

the Precise Point Positioning Service (PPP) of the Natural Resources of Canada (https://webapp.csrs-scrs.nrcan-rncan.gc.ca/geod/tools-outils/ppp.php, accessed on 26 June 2023), while rover position was processed using the RTKLib software (https://rtklib.com/, accessed on 26 June 2023) through a post processed kinematics (PPK) workflow. The final absolute positional accuracy of the products is below one meter because the results of the PPP workflow has a positional accuracy between 0.2 m and 1 m. The acquired images and GCPs were analyzed and used as inputs in the software Agisoft Metashape Professional 1.6.2 (https://www.agisoft.com/, accessed on 26 June 2023). With this SfM-MVS (structure from motion-multi-view stereo reconstruction) method we created an orthomosaic and a digital surface model for each site, similar to a previous study in the same geographic region [32]. Table 1 shows more details about the photogrammetric products.

2.2. Annotations

The preparation of the image data for machine learning started with the annotation of classes of interest. The LCN and TC sites were used for training and testing the deep neural networks; the EGS site was used as an out-of-distribution dataset. In the created orthomosaics it was easy to visually distinguish the regions of mangrove forest from the surrounding terrestrial forest. We delimited the area of the mangrove forest to only use this region during the prediction by the machine learning process (see orange outline in Figure 3 and Table 1 for area sizes). In LCN, 61% of the area is covered by mangrove forest, in TC 50% is covered by mangrove forest and in EGS 26% of the area is covered in mangrove forest. Inside the mangrove forest stands of LCN and TC, we selected three subplots per site to annotate the classes manually, specifically for the machine learning training process (see red outline in Figure 3; see Table 1 for the area sizes). In LCN, 22% of the mangrove forest area was annotated and in TC 24% was annotated.

Inside these subplots, different types of annotations were made for training semantic segmentation and instance segmentation CNNs (Figure 3). For semantic segmentation networks, pixel annotations were required. We selected *P. rhizophorae*, *R. mangle*, short-sized *R. mangle*, water and mud as our target classes (see Table 2A for annotation numbers). It was possible to visually differentiate between *P. rhizophorae* and *R. mangle* species in most cases. In some areas, distinct short-sized and shrub-like tree patches were visible. After comparing to on-ground images it was clear that these patches were comprised of short-sized *R. mangle*. Water pixels were also manually annotated. After these annotations were finished, the remaining non-annotated pixels were labeled as mud.

Table 2. (**A**) Pixel-wise and (**B**) tree-wise annotation details per site.

(A) Pixel-wise annotations for semantic segmentation			
Label	**La Chunga North**	**Terron Colorado**	**Total**
Pelliciera rhizophorae	24,304 m^2 17,753,335 px (48%)	10,268 m^2 9,429,282 px (35%)	34,572 m^2 27,182,617 px (42%)
Rhizophora mangle	7998 m^2 5,842,042 px (16%)	6637 m^2 6,094,429 px (22%)	14,635 m^2 11,936,471 px (19%)
Short-sized *Rhizophora mangle*	3716 m^2 2,714,167 px (8%)	629 m^2 577,334 px (2%)	4345 m^2 3,291,501 px (5%)
Water	2214 m^2 1,617,468 px (4%)	1239 m^2 1,137,770 px (4%)	3453 m^2 2,755,238 px (5%)
Mud	12,115 m^2 8,849,536 px (24%)	9637 m^2 10,020,035 px (37%)	21,752 m^2 18,869,571 px (29%)

Table 2. *Cont.*

(B) Tree-wise annotations for instance segmentation			
Label	**La Chunga North**	**Terron Colorado**	**Total**
Pelliciera rhizophorae	2855 trees	1756 trees	4611 trees

* Pixel-wise annotation percentages are relative to the total annotated area in each plot.
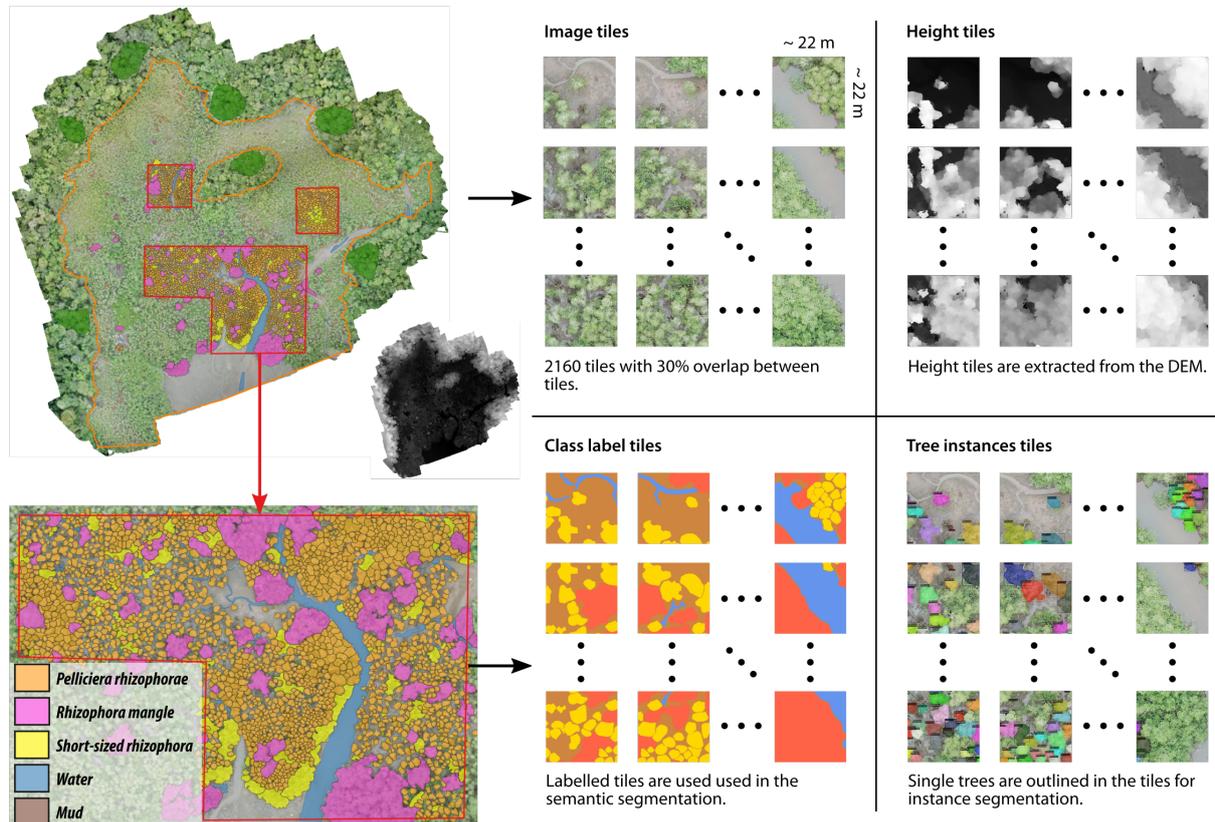
Figure 3. Annotating and tiling for AI: within the orthomosaics, working regions were marked, inside which the mangrove forests were considered for further classification (see orange lines). Inside sub-regions (red polygons), annotations were created for 5 classes (*Pelliciera rhizophorae*, *Rhizophora mangle*, short-sized *R. mangle*, water and mud). The areal annotations were used for semantic segmentation and the individual *P. rhizophorae* tree annotations were used in instance segmentation. The large orthomosaic images and their corresponding annotations were tiled using different strategies and allowed to downsize the classification problem to fit within the constraints of our computational resources. Different combinations of input signals from the plots were used by merging color pixels and the height information from the DSM.

Tree instances were only marked for the *P. rhizophorae* species. Each tree was visually identified on the orthomosaic images and delineated using shapes in QGIS v3.12 (https://www.qgis.org, accessed on 26 June 2023). In total, 4611 *P. rhizophorae* trees were annotated, 2855 in LCN and 1756 in TC (Table 2B). Individual *R. mangle* trees were difficult to visually delineate, and therefore areas of contiguous canopy of this species were annotated.

2.3. Data Tiling

The large sizes of the orthomosaic files (i.e., $21,068 \times 19,855$ pixels for LCN, 1.3 GB) are not directly suited for supervised learning with neural networks due to computational restrictions. In machine learning pipelines, the large orthomosaics are processed by taking smaller tiles as the processing unit. We implemented tiling with windows of a fixed size

of 512 × 512 pixels (around 17 × 17 m), which allows for an average of 30 trees of the *P. rhizophorae* species inside each tile. The tiling can be done with or without overlap between adjacent tiles to reduce uncertainties of predictions around tile borders by the CNNs. Using overlap also requires us to merge tree instances that are split between the borders of 2 or more tiles. We selected 30% overlap between tiles (154 × 512 pixels), allowing *P. rhizophorae* tree masks to maintain their complete shape in at least one tile. Identical tiling procedures were applied to all four linked layers of each study site: the orthomosaic, the elevation image (DSM), the class annotation regions and the tree annotations (Figure 3).

2.4. Deep Learning: Semantic and Instance Segmentation Networks

We used two separate CNNs: a semantic segmentation network for dense pixel-wise predictions and an instance segmentation for delineation of *P. rhizophorae* trees (Figure 4). As input for both networks, we used the RGB tiles extracted from the orthomosaic images and the elevation tiles extracted from the DSM. We also ran the process with RGB + height tiles but a preliminary analysis showed no real benefit to considering the height information for the deep learning process. Thus, for the data experiments and final predictions, we only considered RGB tiles.
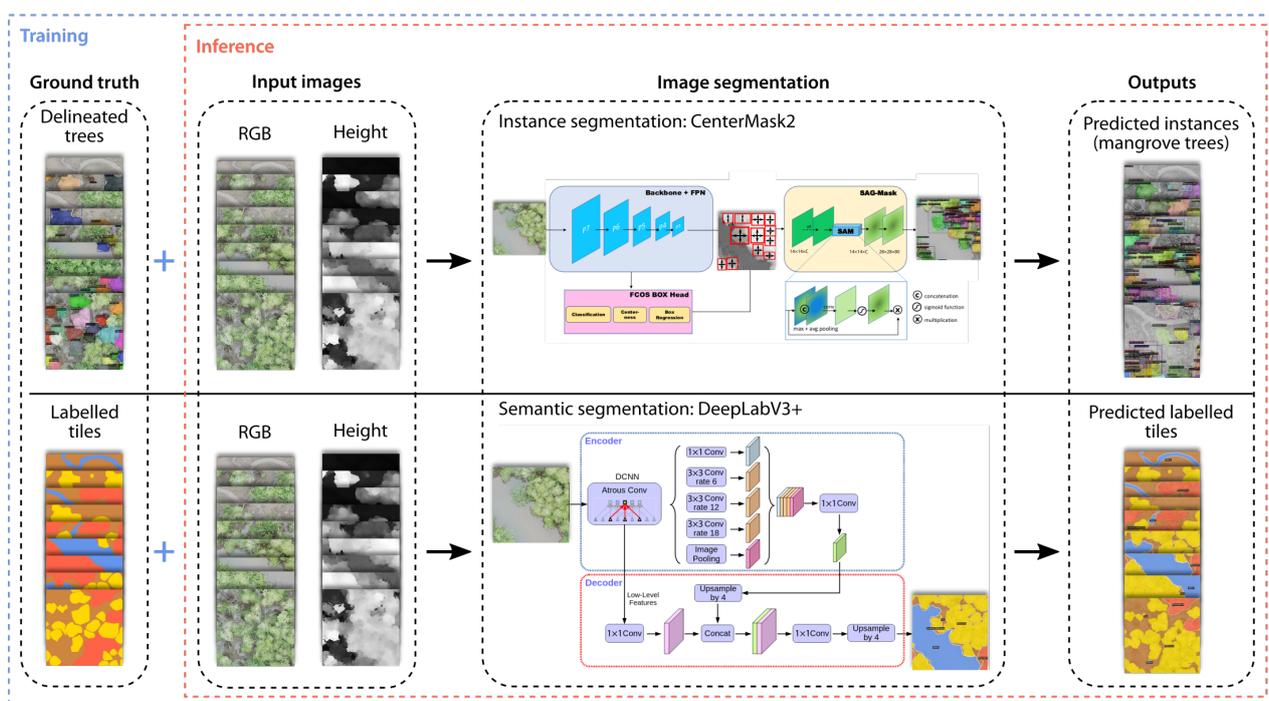
Figure 4. Two networks to rule them all: our workflow uses AI to convert orthomosaics of the mangrove forests into habitat maps and a tree inventory. The input of an RGB, height or RGB+height tile goes through a series of convolutional filters to extract deep features. The instance segmentation network CenterMask2 uses a spatial attention module to suggest prediction masks inside bounding boxes, which potentially delineate the canopy of individual *P. rhizophorae* trees. The semantic segmentation network uses an encoder and decoder framework to assign one of five semantic labels (see Figure 3) to each pixel. The network architecture illustrations are adapted from [44] for CenterMask2 and from [45] for DeepLabV3+.

We implemented the DeepLabV3+ [45] semantic segmentation network with the Detectron2 Python library [46], which is build on the PyTorch machine learning library [47]. This algorithm has been successfully applied towards pixel-wise segmentation of natural habitats in top-down images [48,49]. A recent study [38] used a modified version of DeepLab for semantic segmentation of hyperspectral images in Brazilian forests. We selected the ResNet-101 backbone for the DeepLabV3+ architecture, which also uses separate

atrous convolutional layers to ensure higher-resolution outputs and reduce execution time. Starting from network weights from training with the ImageNet dataset, we retrained the whole network parameters with our image data. For training, we used 300 tiles in batches of 4, and employed 15,000 iterations in total. For the optimizer, we used a learning rate scheduler with polynomial decay (weight decay of 0.001) and warm-up period of 1000 iterations, developed for the DeepLab network. We use an initial learning rate of 0.01, a "hard pixel mining" loss function, and a loss weight of 1. The DeepLab network was trained on two NVIDIA RTX 2080 Ti GPUs (NVIDIA, Inc.—Santa Clara, CA, USA) with 12 GB of memory each. The annotation input for the training of the network were densely annotated tiles (see Figure 3). The outputs of the semantic segmentation network were vectors of five class probabilities for each pixel in a tile. The highest probability value was selected as the class prediction in each pixel.

For instance segmentation, we implemented the CenterMask2 network on the Detectron2 framework, an improved version of the CenterMask instance segmentation network [44]. The authors show that CenterMask2 outperforms the more commonly used MaskRCNN (mask region-based convolutional neural network), which has been recently used in tree segmentation studies [27,50,51]. CenterMask2 is an anchor-free one-stage instance segmentation network that implements a spatial attention-guided mask. The pre-trained backbone (on the ImageNet dataset) we used was the VoVNetV2-99 network [52], and its stem and first residual module parameters were frozen. The network ran for 15,000 iterations with batches of 16 images. It used a warm-up multi-step learning rate scheduler, with 0.001 weight decay, 1000 warm-up iterations and steps at 10,000 and 13,000 iterations. The CenterMask2 network ran on two NVIDIA RTX 3090 Ti GPUs with 24 GB memory each. The annotation input for the training of the network were common objects in context (COCO)-style JSON files with tree shape descriptions and locations on the annotated tiles (see Figure 3). The output of the instance and segmentation networks were *P. rhizophorae* tree instance descriptions with bounding boxes, locations, masks and mask prediction scores (prediction confidence). On average, the training of the network took 3 h and 20 min for each experiment.

Given the low number of total training tiles (364) across sites, we used augmentations for both networks, with random flips of the images, cropping and rotations with the Detectron2 training pipeline. We analyzed the amount of data (before augmentation) needed for a better performance of the instance segmentation network. After separating 10% of the tiles as a testing dataset, we created several training datasets using 50%, 60%, 70%, 80% and 90% of the remaining tiles, thus ensuring a consistent testing dataset with no overlap with the training datasets (Figure 5a). We also compared the performance when considering "empty" tiles in the training set, in which no *P. rhizophorae* instance was present, to not over-fit the network. As a measure of performance for instance segmentation we used the mean average precision (AP) as defined by the COCO dataset (https://cocodataset.org/#detection-eval, accessed on 26 June 2023). This index measures the percentage of predicted instance masks for which the IoU (intersection over union) with the ground-truth annotation is larger than a list of 10 different thresholds. The thresholds go from 50% to 95% in steps of 5%, and then the percentages of masks with an IoU larger than the threshold at each step are averaged to get the final AP.

We trained the semantic segmentation network on 90% of the tiles and 10% testing tiles. We measured the performance of the network (Figure 5c) with precision (user's accuracy) and recall (producer's accuracy) confusion matrices and with the Cohen's Kappa score, overall accuracy, overall recall, overall precision and the F1-score (the harmonic mean of overall recall and precision values).

Additionally, we measured the agreement between *P. rhizophorae* and *R. mangle* predictions between the instance and semantic segmentation networks (Figure 5b). For this we calculated the area fraction inside instance predictions that is predicted as *P. rhizophorae* or *R. mangle* by the semantic segmentation network.

(**a**)

(**b**)

(**c**)
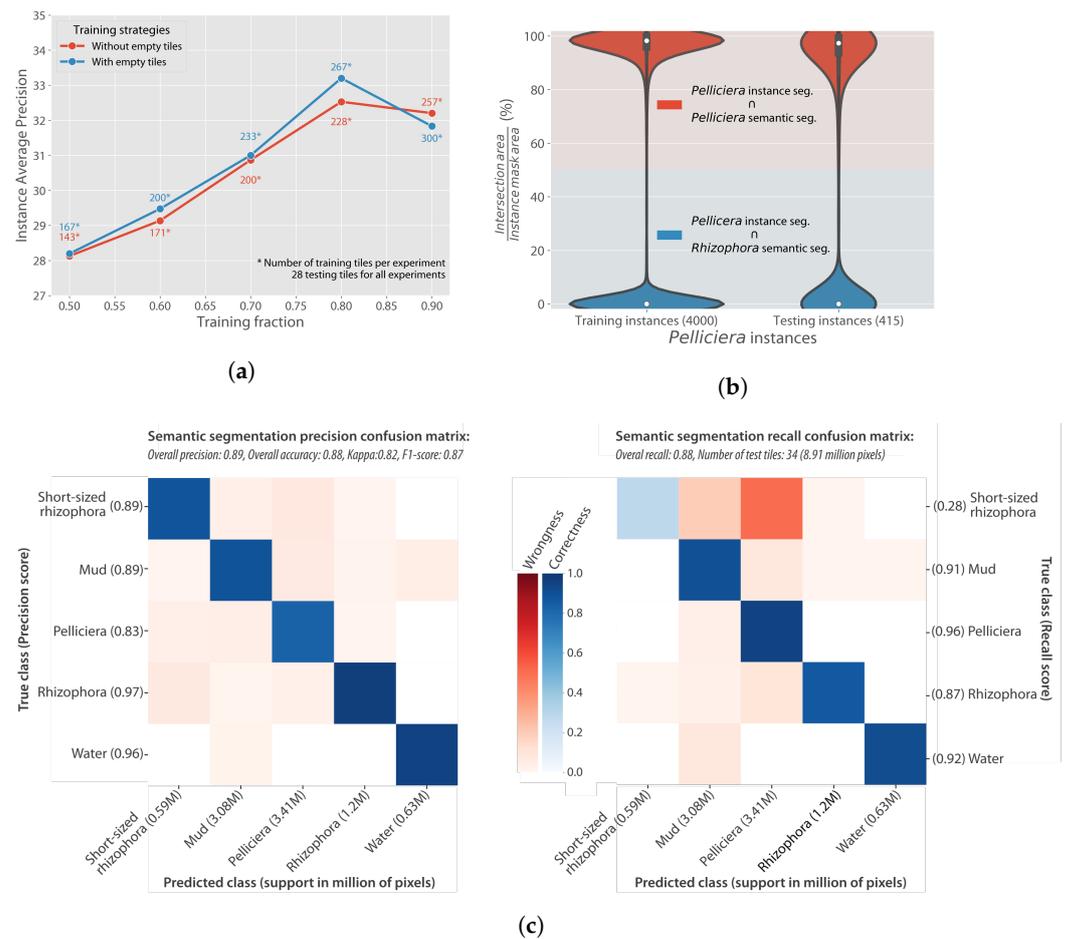
Figure 5. Evaluation of training modality: we trained the networks with 6 different classification regions annotated on 2 separate plots, looking for an optimal mix of annotation effort and generalization performance from the networks. The higher the number of tiles used in training the network the better the performance of the prediction (**a**). The best performance (33.1% instance average precision) was achieved with 80% of the training tiles (267), using datasets with empty tiles (tiles were no *P. rhizophorae* instance is found). In (**b**), we compared the agreement (or error) between predictions of instance and semantic segmentation networks. Agreement of *P. rhizophorae* predictions for both training and testing instances in both sites had a median of 97%. The error between *P. rhizophorae* instances and *R. mangle* areas was very low, with a mean of 2.6% overlap for training instances and 4.5% for testing instances. In (**c**), we show the semantic segmentation performance. All classes had a precision of over 80% and all classes except the short-sized *R. mangle* class had high recall scores (>87%). The low recall score of short-sized *R. mangle* (28%) shows a large confusion with the *P. rhizophorae* class.

2.5. Untiling Strategies

The predictions of the network on individual tiles had to be untiled back together to recover a consistent prediction over the complete mangrove forest area. Given that the tiling process was done with overlap between the tiles, different strategies had to be applied to accurately recover and resolve the predictions in overlapping regions. The untiling process had to be implemented independently for instance segmentation and semantic segmentation predictions.

2.5.1. Untiling Instance Segmentation Tiles

Untiling the predicted instance tiles was done with a novel developed algorithm (see Algorithm 1 for the pseudo-code) to control the preservation of tree instances in border regions across tiles. The algorithm is controlled by two thresholds: one for the minimum

predicted mask score and one for the overlap between two or more predicted instances, which intersect in the prediction. A schematic of the untiling steps is shown in Figure 6.

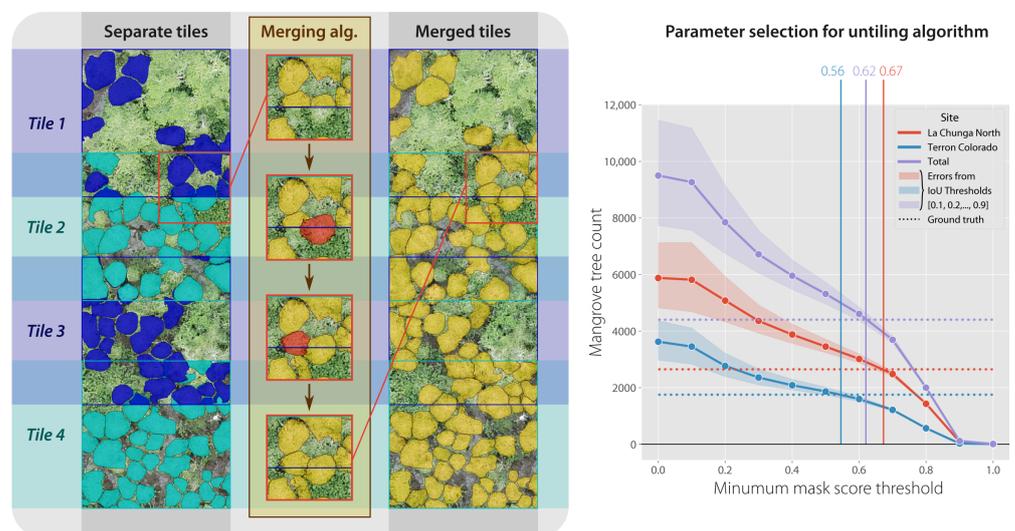
Algorithm 1 Tree instances untiling algorithm

1: $tiles \leftarrow M(tile\_width \times tile\_height \times num\_tiles \times instances\_per\_tile)$ ▷ M is a Matrix
2: $mask\_minimum\_score \leftarrow \alpha$
3: $overlap\_threshold \leftarrow \beta$
4: $tiles \leftarrow RemoveTilesWithoutInstances(tiles)$
5: $tiles \leftarrow RemoveInstancesWithLowScores(tiles, mask\_minimum\_score)$
6: $untiled\_map \leftarrow M0(orthomosaic\_width \times orthomosaic\_height)$ ▷ A matrix filled with zeroes
7: $new\_instance\_id \leftarrow 0$
8: **for** *tile* **in** *tiles* **do**
9:    **for** *instance* **in** *tile.instances* **do**
10:       $new\_instance\_id \leftarrow new\_instance\_id + 1$
11:       $temp\_tile \leftarrow Crop(untiled\_map, tile.coordinates)$
12:       $intersected\_instances \leftarrow temp\_tile \cap instance.mask$
13:       $merge\_to\_instance \leftarrow NULL$
14:       $intersected\_instance \leftarrow NULL$
15:       **for** *intersected_instance* **in** *intersected_instances* **do**
16:          $intersection \leftarrow intersected\_instance.mask \cap instance.mask$
17:          **if** $intersection.size > (instance.size \times overlap\_threshold)$ **then**
18:             **if** $!merge\_to\_instance \parallel intersected\_instance > merge\_to\_instance$ **then**
19:                $merge\_to\_instance \leftarrow intersected\_instance$
20:             **end if**
21:             $temp\_tile[intersection] \leftarrow intersected\_instance.id$
22:             $instance.mask[intersection] \leftarrow False$
23:          **else**
24:             **if** $intersection.size > (intersected\_instance.size \times overlap\_threshold)$ **then**
25:                $intersection \leftarrow intersected\_instance$
26:                $temp\_tile[intersection.mask] \leftarrow new\_instance\_id$
27:                $instance.mask[intersection] \leftarrow True$
28:             **end if**
29:          **end if**
30:       **end for**
31:       **if** $merge\_to\_instance \neq NULL \,\&\, intersected\_instance \neq NULL$ **then**
32:          $temp\_tile[instance.mask] \leftarrow merge\_to\_instance.id$
33:          $intersected\_instance.size + = intersection.size$
34:          $Delete(instance)$
35:       **else**
36:          $temp\_tile[instance.mask] \leftarrow new\_instance\_id$
37:       **end if**
38:       $untiled\_map[tile.coordinates] \leftarrow temp\_tile$
39:    **end for**
40: **end for**

We first filter the tiles that do not have instances predicted in them. Then, we filter instances that have a prediction score (confidence) under a given threshold *mask_minimum_score* in the range $[0.0 - 1.0]$. We create an empty matrix the same size as the original orthomosaic image (*untiled_map*). We iterate over all remaining instances in all remaining tiles, creating a unique ID for any new instance that we keep. We crop the region corresponding to the tile in the large orthomosaic image and save it to *temp_tile*. We then calculate the overlap between the new *instance* and every *intersected_instance*. We iterate over the overlapping instances and calculate the intersection size with the current *instance*. We compare this overlap with mask size of the current *instance* times a given *overlap_threshold*

in the range [0.0–1.0] (Algorithm 1 line 17–23). If the overlap size is larger than this value, we assign the current instance pixels to one of the overlapping instances in *temp_tile*. To decide into which instance to merge, we first check that no *merge_to_instance* variable was set or that the *intersected_instance* size is larger than the previously saved instance in *merge_to_instance* (Algorithm 1 line 18–20). We then replace the intersection location in *temp_tile* with the ID of the current *intersected_instance*. We also remove the intersected area from the current *instance*. Otherwise, in case the *intersection.size* is larger than ($intersected\_instance.size \times overlap\_threshold$), we assign the intersection to the current *instance* in *temp_tile* (Algorithm 1 line 24–28). Afterwards, if *merging_to_instance* is set, we assign all pixels in *temp_tile* of the current *instance* to that instance in *temp_tile* and delete the current *instance* (Algorithm 1 line 31–35), or else we just add the (remaining) parts of the current *instance* to its location in *temp_tile*. Finally, we merge the updated *temp_tile* back to the larger *untiled_map*, which after all iterations will contain tree instances without any overlap and clear crown boundaries. The algorithm's execution time is bound to the number of tiles (tile size and overlap) and number of instances predicted in each tile.


Figure 6. Instanceuntiling algorithm and parameters. We provide a heuristic algorithm for untiling the predicted instance segmentation within overlapping tiles. The algorithm works by filtering low-scoring predicted tree masks and handling overlapping tile sections with an overlap threshold to merge overlapping instances. In the illustration we show the process of merging two or more instances into one or more instance, such that a coherent shape and tree count is preserved. We calculated the ideal minimal mask score threshold and overlap threshold to preserve the original count of trees in annotated areas. For the overall scene reconstruction we found that a 0.62 minimum mask confidence threshold, together with a 0.5 overlap threshold predicted the same tree count as the original annotation count. Any change in minimum mask confidence needs an adjustment in the overlap threshold (error shown in shaded regions).

We measured the effects of the predicted mask score and overlap threshold variables by looking at which values make the count of trees closest to the original annotations in the annotation regions (Figure 6).

2.5.2. Untiling Semantic Segmentation Tiles

The predicted semantic tiles were untiled following three different strategies: overlaying, clipping and averaging (schematic in Figure 7). Overlaying simply places each new tile in its original position without considering any overlapped tile in that region. We overlaid tiles starting in the top left corner of the orthomosaic image, going from top to

bottom, and moving to the subsequent column until the last tile is reached in the bottom right corner. This gives preference to predictions in tiles that are further down the list, where only the last tile to be untiled maintains its complete area and all other tiles maintain 49% of it (given a 30% overlap example). Clipping means that the half of the overlap region is clipped off the border of tiles and then placed in its original location on the orthomosaic. In a 30% overlap example, corner tiles retain 72% of their central area, tiles at the edge of the orthomosaic retain 60% and every other tile retains 49%. Averaging means taking the mean of network softmax values in the overlapping regions before the argmax function is used to select the predict class. In a 30% overlap example, corner tiles will have 28% of its area averaged, border tiles 40% and all other tiles 51%.
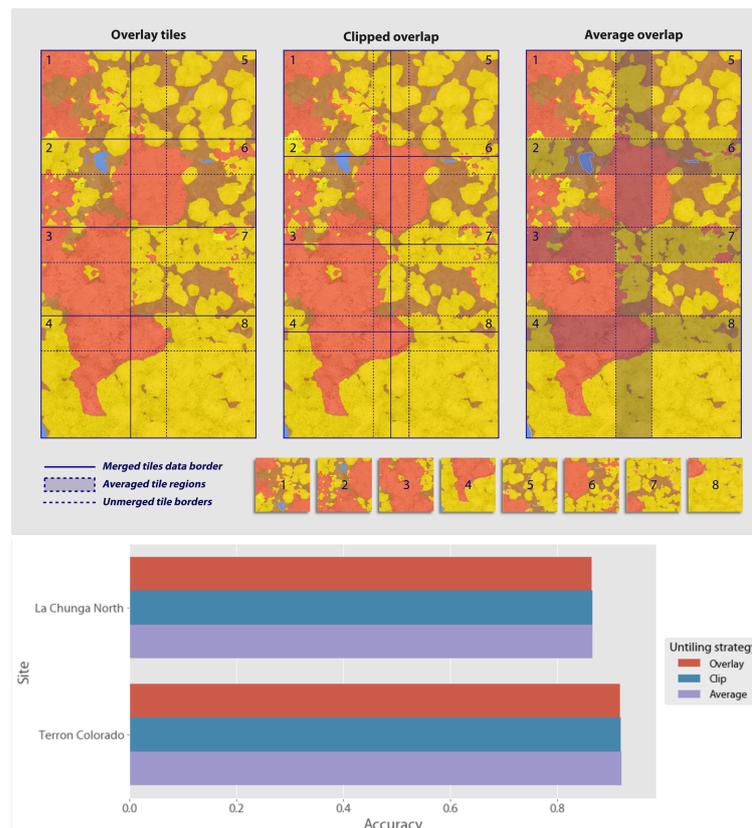
Figure 7. Semantic tile-merging strategies. For the semantic segmentation tiles we tried 3 different untiling strategies to recover the predicted habitat map: overlaying, clipping and averaging. We compared them to the ground truth densely annotated tiles and calculated their accuracies. No clear advantage was detected for any of the untiling strategies, which hints at the good prediction confidence of state-of-the-art semantic segmentation networks, even around borders of images.

We measured the accuracy for each untiling strategies by dividing the total number of predicted pixels of every class (inside the annotation regions in each site) by the total number of pixels for that class in the manual annotation (Figure 7).

2.6. Digital Terrain Model, Digital Elevation Model and Canopy Height Model

After creating the untiled orthomosaics of semantic and instance segmentation predictions we created a digital terrain model, digital elevation model and a canopy height model. In this study we reference DTM as a model only showing terrain features (i.e., mud and water pixels), selected from the DSM, which is the raw elevation model that considers all natural and artificial features on the map. The DEM is the result of interpolating the DTM to describe the elevation of the terrain below natural and built/artificial features. A CHM is the subtraction of a DEM from the DSM. In this study, we selected ground points in

the orthomosaics to create a DTM and then interpolated the empty areas with smoothing, to generate a DEM [36,37].

We compared 2 strategies to select ground points and generate the DTMs. The first strategy was manually selecting ground points (in QGIS) that visually looked like mud or water region close to the mangrove trees. We corroborated that the selected region did not contain any higher elevation pixels in the DSM (corresponding to the surrounding trees), given that the initial resolutions of the orthomosaic and DSM were not identical. The manual selection of points took around 2 h for the TC site and 3 h for LCN.

The second strategy used our semantic segmentation predictions as they also contain ground pixels (mud and water classes). We use those regions to select the relevant points to interpolate into a DEM. Given that the predictions might contain errors, we used a threshold of 95% network confidence of the ground predictions to select pixels. This yields a very small number of ground predicted regions (under 0.5% of pixels). Finally, to remove residual pixels that may contain high elevation values in the DSM, we convolve a window of 2000×2000 pixels across the entire DSM and select pixels with elevation under a parameterized percentile value. The pixels that passed through this filtering were very likely to be only the ground level regions and were used as ground points for the DTM interpolation.

For both strategies, we use the Geo-spatial Data Abstraction Library's (GDAL) *fill_no_data* function to interpolate and smooth out the DTM into a DEM. This function uses the inverse distance weighting (IDW) algorithm to interpolate missing values in a raster, followed by 3 smoothing passes with a 3×3 kernel. We then subtract the DSM elevation from the DEM elevation to obtain a CHM. We calculated the height of a tree by selecting the maximum elevation inside its contoured shape from the CHM.

We illustrate the complete process in Figure 8. We compared the resulting elevation of the trees using both strategies by plotting them against each other, and by comparing the bias of the mean and the 95% limit of agreement using Bland–Altman (or mean-difference) plots (Figure 8). We use the first "manual" ground pixel selection strategy as control for the second "automatic" ground pixel detection strategy.
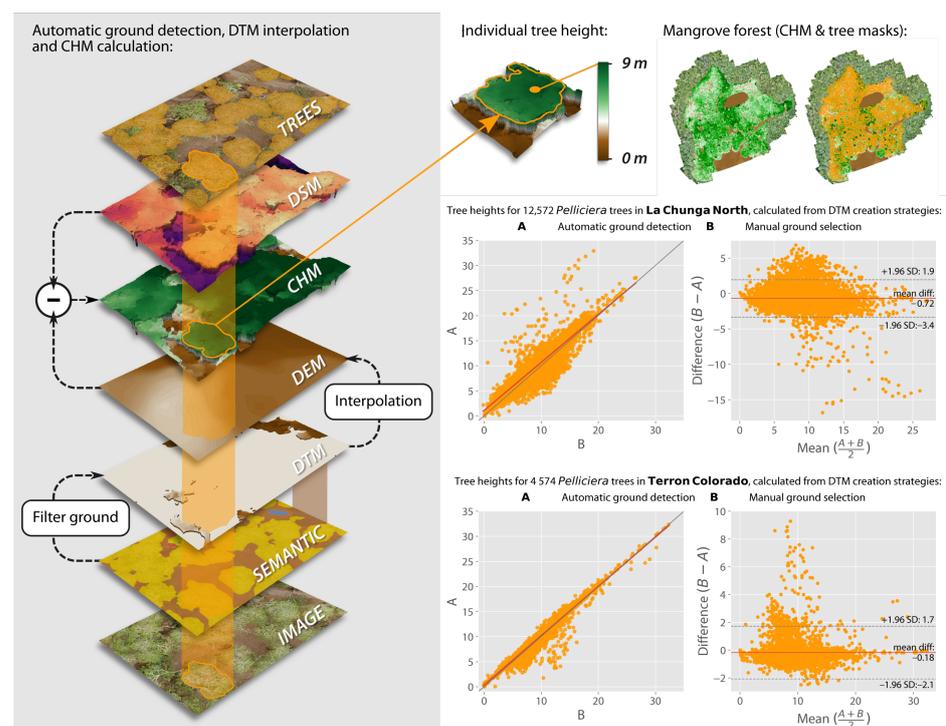

Figure 8. Digital elevation model (DEM) and canopy height model (CHM) strategy comparison: we illustrate our automatic ground selection and interpolation process. From the semantic segmentation

predictions we select and filter high confidence ground pixels, which we then use to interpolate the DSM values in the corresponding ground locations. We subtract the DSM values from the DEM to generate a CHM. Finally we "cookie-cut" the predicted tree instances on the CHM to calculate height statistics of the tree crown. We check if the automatically extracted DEM is correlated to a DEM generated from manually selected ground regions in the plot. The tree heights from both methods did not show a significant bias for either technique as shown in the regression plots and mean-difference plots for both LCN and TC sites. Outliers can be caused by imperfections in the original DSM.

2.7. Forest Inventory

We summarize the attributes of the automatically delineated trees, such as crown shapes and heights, into an inventory of the forest (Figure 9). We calculate mean and maximum pixel heights inside predicted tree crown shapes for both DEM creation strategies. We also calculate and plot the tree crown diameter from the major axis of the ellipsis with the same second moment as the crown polygon. Other metrics calculated from the instance contour are the tree crown eccentricity, which is the ratio of the focal distance (distance between focal points on the ellipsis covering the tree crown shape) over the major axis length (a value of 0 means the shape is a perfect circle), and tree crown area in square meters. We also plot the tree height in meters against the canopy area in square meters using a linear regression plot. These measurements were extracted with the "regionprops" function of the "scikit-image" Python library [53].
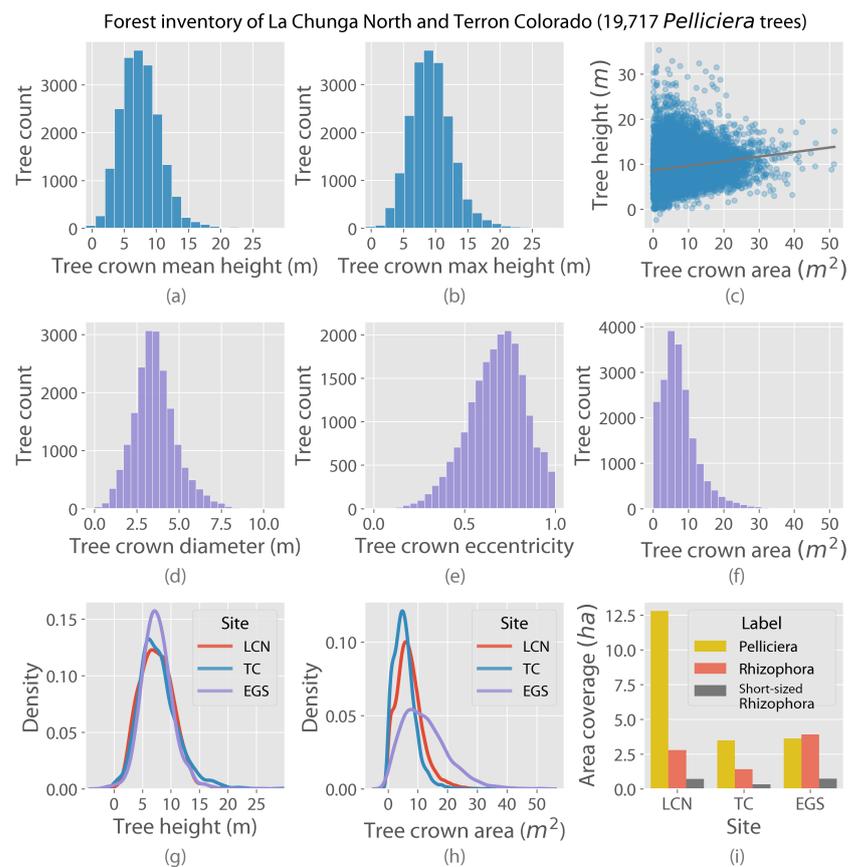
Figure 9. Combined inventory of the La Chunga North and Terron Colorado mangrove forest plots. We measured the (**a**) mean height and (**b**) maximum height in predicted *P. rhizophorae* instances as well as (**d**) tree crown diameter, (**e**) eccentricity and (**f**) area. We plotted (**c**) tree height against tree crown area. We compare (**g**) the tree heights and (**h**) tree crown areas of *P. rhizophorae* from the training sites LCN and TC with the out-of-distribution site Estero Grande Shore. We also compare the (**i**) area coverage of *P. rhizophorae*, *R. mangle* and short-sized *R. mangle* across the three sites.

Finally, having the trained pipeline, we tile, predict the semantic and instance segmentation outputs and untile the out-of-distribution EGS site. In order to measure the scalability of the method, we then compare *P. rhizophorae* tree heights and tree crown areas for all three sites. We also compare the area cover of the *R. mangle* and the *P. rhizophorae* species as well as that of the short-sized *R. mangle* class from the semantic segmentation predictions. Finally, we calculate the pixel-wise height distributions in the CHMs for area-wise predictions of the three tree classes.

3. Results

The presented workflow allows for automatic delineation of individual *P. rhizophorae* trees and the segmentation of *R. mangle* canopy areas, as well as other land cover classes (mud and water). We review the accuracy of both instance and semantic segmentation networks, as well as of the untiling of the predicted tiles, and finally of the automatic calculation of tree measurements, such as height from the generated CHM.

3.1. Deep Learning Performance

We measured the performance of both instance and semantic segmentation networks separately but also compared their agreement on predictions for the *P. rhizophorae* class and overlap with the *R. mangle* class.

In Figure 5a, we show the performance of the CenterMask2 network when both tiles with *P. rhizophorae* instances and tiles without *P. rhizophorae* instances were considered in the training procedure. For both cases, the performance peaked with 80% of the training tiles (228 tiles without and 267 with empty tiles). When considering empty tiles, the AP was 33.2% and without the empty tiles it was 32.6%. With the 90% training fraction, the performance reduced by 1.2% when considering empty tiles and only by 0.3% when not. The best performing network was used for the final tile predictions.

The performance metrics for the semantic segmentation network are shown in Figure 5c. The overall precision for the network was 89%, the overall recall 88%, the F1-score was 87%, the overall accuracy was 88%, and the Kappa score was 82%. The precision confusion matrix also shows the per-class performance, where *R. mangle* has the highest score (97%), followed by water (96%), mud (89%) and short-sized *R. mangle* (89%) and finally *P. rhizophorae* (83%). In the recall matrix, the highest value was for *P. rhizophorae* with 96%, while by far the lowest was the short-sized *R. mangle*, with 28%. The major confusion that affected the recall values was between *P. rhizophorae*, mud and short-sized *R. mangle*. Other minor confusions occurred between water and mud and between short-sized *R. mangle* and *R. mangle*.

The two networks showed good overlap between their *P. rhizophorae* predictions, with median values of 98% for training instances and 97% for testing instances. Nonetheless, some *P. rhizophorae* tree crown instances in the testing tiles had fewer pixels predicted as *P. rhizophorae* by the semantic segmentation network inside their area (lower 25% quartile of 85% overlap). Similarly, there seemed to be little confusion between predictions of the two mangrove species. We found a median of 0.05% of all training and testing instances and a mean of 2.6% for training instances and 4.5% for testing instances. The instances in testing tiles showed higher overlap with up to 12% overlap for the upper 75% quartile.

3.2. Untiling Accuracy: Tree Instances

Our novel instance untiling algorithm (Algorithm 1) for tree crown masks can be modulated by two parameters: the mask prediction score and the overlap (IoU) threshold. To understand the interplay between the two parameters, we plot the mask score threshold value against the *P. rhizophorae* tree count after the untiling algorithm has been applied (Figure 6). The forest area used in this experiment is the sum of all the annotated regions in each site; hence, the dotted "ground truth" lines show the total number of manually annotated *P. rhizophorae* trees. The error, shown in shaded areas, corresponds to the different values obtained from changing the overlap threshold (from 10% to 90% overlap). For the

LCN site, the ideal minimum mask score threshold was at 67% and an overlap threshold of 50%. For TC, the mask threshold was at 56% confidence and the overlap threshold at 50%. When combining both sites, the ideal mask score was 62% and an overlap threshold of 50%. The minimum mask score changed between 59% and 65% when the overlap threshold was changed from 10% to 90%, respectively. We used the ideal value of a 62% mask score threshold and 50% overlap threshold for the final predictions of the complete mangrove forest sites.

3.3. Untiling Accuracy: Semantic Labeling

Similar to the instance segmentation network, we measured the accuracy of untiling the results of semantic segmentation prediction on tiles with overlap while employing three different merging strategies (Figure 7). For each strategy and site, we calculate the accuracy by comparing the labeled pixels of each annotated regions against the labels in the untiled prediction. The accuracy variability was negligible for all strategies. In LCN the accuracy was 86.4% for the overlay and clip strategy and 86.6% for average, while in TC, it was 91.5%, 91.6% and 91.7%, respectively. These accuracy values for the final untiled areas correlate with the accuracy reported for the testing tiles in the confusion matrices (Figure 5c). This portrays the great generalization capabilities of the semantic segmentation network, even in image borders.

3.4. Automatic Creation of Digital Elevation Model and Canopy Height Model

After untiling as described, we compared two ways to generate the needed DEM to accurately calculate the CHM: manually selecting ground pixels versus machine-predicted (semantic segmentation network) mud and water pixels. In Figure 8, we show that for a vast majority of the *P. rhizophorae* trees, the heights calculated from the CHMs from both DEMs correspond by staying close to the one-to-one line in the regression plots (Figure 8). We predicted and compared 12,572 *P. rhizophorae* trees in the LCN site and 4574 *P. rhizophorae* trees in TC. The Bland–Altman (mean-difference) plots show little bias in tree height predictions both in LCN (-0.72 m of mean difference) as in TC (-0.18 m of mean difference) from the automatic ground detection against the manual ground selection technique. In LCN, a small number of outliers were found outside of the -3.4 lower 95% limit of agreement (-1.96 SD line) standard deviation, where some trees were predicted as taller when using the automatic ground detection. Inversely, in TC, some trees were predicted as taller when using the manual ground selection strategy DEM, pushing the upper 95% limit of agreement (the $+1.96$ SD line) to 1.7, but the lower 95% limit was higher at 2.1.

3.5. Tree Inventory and Area Coverage

In Figure 9, we summarize the tree-level description of the forest stands created by our workflow. This includes the *P. rhizophorae* tree inventory and the area coverage of the *R. mangle* mangrove species and short-sized *R. mangle*. For the automatic ground detection CHM, the mean pixel height in *P. rhizophorae* predicted masks had a mean value of 7.58 m and the mean of maximum height values was 9.33 m (Figure 9a). The height values in the 25% and 75% quantile range were 5.35 m to 9.5 m for the automatic CHM, and 20.48% of trees had a maximum height over 10 m (Figure 9b).

We also calculated the tree crown diameter (major axis of ellipse), eccentricity and areas in square meters (Figure 9d–f). The mean of the crown diameters was 3.9 m. The distribution of eccentricity of the tree crowns tended towards 1.0 with a mean of 0.67, meaning that their shapes were more elongated and less circle shaped. The mean of tree crown areas was 6.77 m^2. The largest crowns measured up to 20 m^2. For the *P. rhizophorae* trees, we checked the correlation of tree height with the canopy areas (Figure 9c). We noticed that shorter trees did not have larger crown areas (Figure 9c). The opposite was not the case, since we find small canopy areas with large heights.

We compared tree heights and tree crown areas of the two in-distribution sites (LCN and TC) with the out-of-distribution EGS site (Figure 9g,h). The calculated heights show an almost identical distribution, with very similar means and with 50% of the trees in the 5–10 m range. The tree crown areas present similar distributions between LCN and TC, with means around 7 m^2 and most trees having an area under 10 m^2. Trees in the EGS site show a wider distribution with a similar mean than the other two sites but with 40% of trees in the 10–20 m range.

Finally, we calculated the area coverage for *P. rhizophorae*, *R. mangle* and short-sized *R. mangle* from the semantic segmentation predictions. In LCN, the *P. rhizophorae* species was the most common class with 12.79 ha, followed by *R. mangle* with 2.8 ha and short-sized *R. mangle* with 0.6 ha. In TC, the difference was not as pronounced, with *P. rhizophorae* covering 3.49 ha and *R. mangle* covering 1.41 ha and short-sized *R. mangle* with 0.34 ha. In the out-of-distribution site, EGS, *P. rhizophorae* covered 3.63 ha and *R. mangle* covered 4.1 ha, and short-sized *R. mangle* covered 1.1 ha. The average height of *R. mangle* areas over the three sites had a range of 6–12 m with a mean of 10 m. The heights of short-sized *R. mangle* areas was lower, mostly in the 3.3–5.4 m range.

4. Discussion

In this study, we propose a novel method for creating an inventory of mangrove forests and their surroundings. We also provide a technique for the automatic creation of a DEM and CHM, to calculate heights of individual trees and tree areas. We show that machine learning with deep neural networks has the potential to greatly increase the throughput and precision of surveys of hard-to-access forest areas. Furthermore, by detecting the contour of individual tree crowns and their respective heights, valuable information is obtained for allometric analysis. We show that the workflow can be scaled to handle large mangrove forest regions and generalizes well to new survey data that were not in the training dataset.

4.1. Effort Reduction of On-Ground Work and Annotation

Mangrove forests present difficult conditions for on-ground field surveys, given their complex root systems, tidal regimens and remote locations. The use of airborne imaging systems can alleviate the effort by covering large distances in a short time and not being hindered by the complex setting of the forest floor. UASs, in particular, provide a controllable platform for high resolution imaging of target areas from above. In this study, we used the photogrammetric products (orthomosaic and DSM) constructed from aerial imagery captured with consumer-grade UASs in a remote and inaccessible area of Utría National Park on the Colombian Pacific coast. We used UASs with their default RGB cameras because this technology is easily accessible for local park authorities. Other studies, in contrast, have used more expensive sensors, such as multispectral or hyperspectral cameras, as well as LiDAR sensors [19,38].

We set out to establish that state-of-the-art deep learning techniques can enable even consumer-grade imagery to deliver information-rich survey output at the scale of entire mangrove forests. Given the large extent (103 hectares; Table 1) of the forests captured in the orthomosaics, we annotated subplots that would approximately represent 20% of the total mangrove area (Figure 3). To capture the variability in the sites, we used the following criteria when selecting annotation subplots: presence of both mangrove species, mud and water presence, location in the plot and height differences in the DSM. To train the semantic segmentation network, we densely annotated large areas such that no pixel was left un-annotated. To measure the performance of the untiling algorithms, we also selected rather larger regions to annotate (three per site) instead of directly annotating smaller-sized tiles that would fit in the network. The contouring of individual *P. rhizophorae* trees in QGIS was the most time consuming part of the process, but this time can be reduced by using novel annotation software designed for supervised learning with large orthomosaic images, such as TagLab [54].

The decision to not include the *R. mangle* species in the instance segmentation process was made due to the difficulty for the human annotators to visually identify individual tree crowns from each other. This could be overcome by using more specialized sensors that capture higher spatial and spectral resolutions and UASs with steadier flight control, considering the cost trade-off. Even so, the uneven growth patterns of mangrove crowns can be a limiting factor in comparison to other types of forests, where individual trees are easily distinguishable or where forest canopies have more spaced patterns [34].

We also included the short-sized *R. mangle* class, given that some parts of the forest had a shrub-like aspect that differed from surrounding trees. Most of these areas were exposed to incoming tide, and a smaller fraction were found in-between patches of *P. rhizophorae* trees. After comparing with on-ground images, we determined that those areas were covered in short-sized *R. mangle* trees. Given that it was not possible to visually identify individual tree crowns in the aerial images, we annotated area patches that covered one or more trees.

4.2. Instance and Semantic Segmentation

Using two deep neural networks that produce different outputs helped us achieve three distinct goals. First, the instance segmentation network CenterMask2 was trained to identify individual tree crowns for the *P. rhizophorae* mangrove species. Instance segmentation networks were developed for detecting everyday objects in urban settings but have been successfully transferred to a variety of other fields, such as natural environments [55,56]. Our implementation achieved an AP of 33% using over 80% of our annotated regions for training. This a good performance considering some quality artifacts in the orthomosaics of the images, such as blurring and the reduced training samples. Another source of error was the contour of annotations, given that mangrove canopies were not always 100% distinguishable between species and between trees of the same species. Furthermore, AP is a very stringent metric of performance as it heavily penalizes small errors in the mask overlap.

The second goal that our automation pipeline achieved was to annotate *R. mangle* areas with recall of 87% and precision of 97% (Figure 5c). We were not able to annotate individual trees for this species but were able to describe the area cover. In such cases, where individual trees cannot be detected, area cover and its height distribution can be used to monitor the species AGB [57]. By using the detected trees from instance segmentation and the areas from the semantic segmentation, we can account for every species in the mangrove forest. In Figure 5b, we show that *P. rhizophorae* and *R. mangle* have little to no overlap between the semantic and instance segmentation predictions, indicating a robust separation of these two classes.

The third goal of our workflow was to retrieve ground pixels (i.e., mud and water) to produce a DTM and a subsequent interpolated DEM. The semantic segmentation network predicted areas of the mud and water classes with high precision (89% and 96%, respectively), allowing for accurate detection of ground areas surrounding the mangrove trees.

4.3. Automating the Canopy Height Model

The creation of a DEM from accurately detected ground areas allowed us to extract a consistent CHM, from where individual tree heights could be estimated. The automatization reduces the time effort of manually selecting ground pixels by 3 h per plot. In the created DEM, nonetheless, we found small imperfections noticeable in the outliers of the mean-difference comparison in Figure 8. This was the result of artefacts from the difference in resolution of the DSM and orthomosaic. For example, some pixels in the bordering regions of mangrove trees and ground pixels were predicted as ground but had an elevation value in the DSM that corresponded to the trees. We reduced these errors by selecting only predicted ground pixels with high confidence (>95%) and further filtering pixels under a certain elevation in tiles along the scene (see Methods). After this filtering, the error in the heights of *P. rhizophorae* trees between the two methods was not significant. The outliers

can be further corrected by checking and correcting small imperfections in the automatically generated DEM, which still takes only a couple of minutes compared to hours of selecting ground pixels for a manual DEM. Furthermore, in long-time monitoring settings, the time gain of automating CHM creation is additive. Finer CHM calculations with closer-to-ground sensing techniques can be used for global-scale canopy height estimation studies [58].

4.4. From Pixels to Tiles to Trees

In our workflow, we propose a novel instance untiling algorithm that minimizes errors on tile borders (Algorithm 1; Figure 6). By tiling the forest plots with overlap, we enhance the probability that trees in border regions will be recovered correctly. Nonetheless, it also complicates the untiling process since the decision has to be made if two or more overlapping masks represent the same or different trees. The two settable parameters in our algorithm allow for adjusting the untiling process to match available on ground data (count of trees). The mask prediction score threshold reduces the number of trees considered for the final prediction by discarding low-confidence predictions such that less overlap occurs in the borders. Then, the overlap threshold parameter handles the case when two or more instances do overlap, and depending on the sizes of their masks and their intersection, we consider merging or dividing the masks. The algorithm gives preference for the already existing tiles in the final prediction because it checks first the existing instances for their size versus the intersection size. The algorithm also works if multiple instances are overlapping with the incoming instance, and each is merged into, merged together or split accordingly. In our study case, we utilize an overlap of 30% between tiles, but this algorithm works on any overlap sizes.

Similarly, for the semantic segmentation predictions, we combine the tiles using different strategies (Figure 7). In contrast to the large size of the mangrove forest plots, the benefits of different strategies seem negligible, but it can be relevant if the overlap is larger. We found that averaging was the best way to reconstruct the underlying scene more accurately, similar to what is recommended in [59]. If the overlap is larger (over 50%) and tile sizes smaller, this strategy is also better suited to combine tiles [38]. Nonetheless, with newer state-of-the-art semantic segmentation CNNs, tiling with overlap might no longer be required, given their high confidence predictions, even in border-adjacent pixels.

4.5. Seeing the Forest for the Trees: An Inven(s)tory

The final output of our workflow was an inventory of individual *P. rhizophorae* trees and area cover and height distribution for *R. mangle* and short-sized *R. mangle* (Figure 9). The distribution of heights of *P. rhizophorae* trees in our automated inventory fell within the range found in the literature, with most trees in the 5–10 m range and 15–20% larger trees in the 10–20 m range [60]. The regions classified as *R. mangle* trees had slightly taller values (6–12 m), with a larger Section (38%) of trees surpassing 10 m, which also correlates to literature descriptions of the species' height [61]. Our decision to separate the short-sized *R. mangle* regions to another category was confirmed to be helpful for the class predictions, given the lower height (3.3–5.4 m) for regions of this category. As mentioned previously, these regions hold shorter trees of the *R. mangle* species, which grow like shrubs compared to taller *R. mangle* trees in more protected areas. Separating these two growth forms of the *R. mangle* species could help tailor the allometric equations for calculating AGB to be more precise.

Describing tree crown shapes and sizes from aerial imagery is a complicated task that has been tried with different methods [62]. By using instance segmentation networks on well-defined training data, the task can be seemingly simplified [41]. Our workflow allows for individual tree crown predictions, and the possible descriptions go beyond tree crown diameters. We calculate tree crown areas and eccentricity, which are parameters that can be used for further understanding growth patterns of mangrove tree species in response to environmental factors (e.g, tide shifts, terrain rugosity, wind direction and speed, etc.).

The semantic segmentation prediction also enables us to study the gaps between trees or those separating forest stands. This helps to understand the growth patterns of the whole mangrove forest and the species distributions, depending on environmental variables, such as distance to shore, tidal locations, forest cover loss and water channel formation [29]. It can also aid in detecting deforestation incidents or other disturbances in the environment.

4.6. Scaling Up: Limitations and Future Work

Our dual-network workflow was able to create a detailed inventory of large mangrove area plots. We show that it can scale and be applied onto new large mangrove forest plots (see height comparison plots in Figure 9), with the only condition being that the potential mangrove forest area in the new plot is delineated. In future work, our workflow will be applied onto seven large mangrove plots in the Utría National Park to analyze patterns in the forests. We extract critical information from medium-quality data and show that with consumer-grade technology (UAS and RGB images), complex analyses of forests can be supported for short-term studies or long-term monitoring.

Nonetheless, with better spatial and spectral resolution in the orthomosaics and better spatial and height precision in the DSM, the errors in the predictions could be improved. For example, the use of multi/hyper-spectral cameras mounted on low-flying platforms can improve class separability [38], and the use of LiDAR sensors can improve the CHM precision [19,40,63]. This richer data improves predictions in natural environments, even when more complex communities are targeted [38,64]. Additionally, advancements in earth-observation technologies are allowing us to apply instance segmentation networks on satellite imagery [28]. Research on imagery from low-flying platforms can, in the short-term, be used as detailed monitoring tools and validation information for global studies and, in the long-term, prepare the data-pipelines for enhanced satellite imagery.

The exponential improvement in machine learning platforms also promises to improve the performance of automated monitoring workflows. Both instance and semantic segmentation networks are constantly improving, and as more computational resources are made available, larger and more capable models will be used routinely. Furthermore, the current development of panoptic segmentation networks will allow us to simplify workflows such as ours by classifying foreground and background objects/classes at the same time, removing the need for inter-network comparisons [65].

We use two networks to describe parts of a mangrove forest scene in different ways: pixel-wise and object-wise. We did not include ground measured data in this study, both due to the inaccessibility of the location and to establish the possibility for a quick aerial survey to support rich survey output. Additionally, the scale of the forest area predicted compared to the area that could be manually measured was very large. By comparing the two networks' predictions to each other, we can assure that the underlying scene was consistently described. For the application on new sites, the community composition of the forest must be assessed, and the prediction classes must be adapted accordingly. This constitutes a known drawback of mutli-class supervised learning. Nonetheless, the backbone weights of the networks can be reused for training given that top-down forest features do not change significantly between mangrove trees, providing a starting point for new forest surveys using aerial data.

Our workflow provides a blueprint for automatic forest inventory creation, facilitating rapid automated assessments of large areas of mangrove forests with consumer-grade technology. It benefits from the advancements in UAS technology and artificial intelligence, enabling unprecedented detail in forest-wide inventories, especially in inaccessible areas such as remote mangrove forests.

Author Contributions: Conceptualization: D.S., A.C., G.A.C.-G.; Data Curation: D.S., J.C.M.-R.; Investigation: D.S., E.C., G.A.C.-G., J.C.M.-R.; Methodology: D.S., A.C.; Formal analysis: D.S., A.C.; Software: D.S.; Supervision: A.C.; Validation: D.S., A.C.; Visualization: D.S.; Project Administration: D.S.; Resources: A.C.; Funding acquisition: A.C., G.A.C.-G., J.C.M.-R.; Writing—original draft: D.S.,

References

1. Goldberg, L.; Lagomasino, D.; Thomas, N.; Fatoyinbo, T. Global declines in human-driven mangrove loss. *Glob. Chang. Biol.* **2020**, *26*, 5844–5855. [CrossRef] [PubMed]
2. Menéndez, P.; Losada, I.J.; Torres-Ortega, S.; Narayan, S.; Beck, M.W. The Global Flood Protection Benefits of Mangroves. *Sci. Rep.* **2020**, *10*, 4404. [CrossRef] [PubMed]
3. Castellanos-Galindo, G.A.; Krumme, U.; Rubio, E.A.; Saint-Paul, U. Spatial variability of mangrove fish assemblage composition in the tropical eastern Pacific Ocean. *Rev. Fish Biol. Fish.* **2013**, *23*, 69–86. [CrossRef]
4. Carugati, L.; Gatto, B.; Rastelli, E.; Lo Martire, M.; Coral, C.; Greco, S.; Danovaro, R. Impact of mangrove forests degradation on biodiversity and ecosystem functioning. *Sci. Rep.* **2018**, *8*, 13298. [CrossRef] [PubMed]
5. Alongi, D.M. Carbon sequestration in mangrove forests. *Carbon Manag.* **2012**, *3*, 313–322. [CrossRef]
6. Ellison, A.M.; Felson, A.J.; Friess, D.A. Mangrove Rehabilitation and Restoration as Experimental Adaptive Management. *Front. Mar. Sci.* **2020**, *7*, 327. [CrossRef]
7. Friess, D.A.; Yando, E.S.; Abuchahla, G.M.O.; Adams, J.B.; Cannicci, S.; Canty, S.W.J.; Cavanaugh, K.C.; Connolly, R.M.; Cormier, N.; Dahdouh-Guebas, F.; et al. Mangroves give cause for conservation optimism, for now. *Curr. Biol.* **2020**, *30*, R153–R154. [CrossRef]
8. Innes, J.L. Design of an intensive monitoring system for swiss forests. In *Mountain Environments in Changing Climates*; Routledge: London, UK; 1994; p. 15.
9. Ferretti, M. Forest Health Assessment and Monitoring—Issues for Consideration. *Environ. Monit. Assess.* **1997**, *48*, 45–72. [CrossRef]
10. Guo, K.; Wang, B.; Niu, X. A Review of Research on Forest Ecosystem Quality Assessment and Prediction Methods. *Forests* **2023**, *14*, 317. [CrossRef]
11. Ding, Z.; Li, R.; O'Connor, P.; Zheng, H.; Huang, B.; Kong, L.; Xiao, Y.; Xu, W.; Ouyang, Z. An improved quality assessment framework to better inform large-scale forest restoration management. *Ecol. Indic.* **2021**, *123*, 107370. [CrossRef]
12. Chave, J.; Andalo, C.; Brown, S.; Cairns, M.A.; Chambers, J.Q.; Eamus, D.; Fölster, H.; Fromard, F.; Higuchi, N.; Kira, T.; et al. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* **2005**, *145*, 87–99. [CrossRef]
13. Wang, J.; Chen, X.; Cao, L.; An, F.; Chen, B.; Xue, L.; Yun, T. Individual Rubber Tree Segmentation Based on Ground-Based LiDAR Data and Faster R-CNN of Deep Learning. *Forests* **2019**, *10*, 793. [CrossRef]
14. Tockner, A.; Gollob, C.; Kraßnitzer, R.; Ritter, T.; Nothdurft, A. Automatic tree crown segmentation using dense forest point clouds from Personal Laser Scanning (PLS). *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *114*, 103025. [CrossRef]
15. Ravindranath, N.H.; Ostwald, M. Methods for Estimating Above-Ground Biomass. In *Carbon Inventory Methods Handbook for Greenhouse Gas Inventory, Carbon Mitigation and Roundwood Production Projects*; Advances in Global Change Research; Springer: Amsterdam, The Netherlands; 2008; pp. 113–147. [CrossRef]
16. Persson, H.J.; Ekström, M.; Ståhl, G. Quantify and account for field reference errors in forest remote sensing studies. *Remote Sens. Environ.* **2022**, *283*, 113302. [CrossRef]
17. Zang, J.; Jin, S.; Zhang, S.; Li, Q.; Mu, Y.; Li, Z.; Li, S.; Wang, X.; Su, Y.; Jiang, D. Field-measured canopy height may not be as accurate and heritable as believed: evidence from advanced 3D sensing. *Plant Methods* **2023**, *19*, 39. [CrossRef]
18. Clough, B.F.; Dixon, P.; Dalhaus, O. Allometric Relationships for Estimating Biomass in Multi-stemmed Mangrove Trees. *Aust. J. Bot.* **1997**, *45*, 1023–1031. [CrossRef]

19. Yin, D.; Wang, L. Individual mangrove tree measurement using UAV-based LiDAR data: Possibilities and challenges. *Remote Sens. Environ.* **2019**, *223*, 34–49. [CrossRef]

20. Thomas, N.; Bunting, P.; Lucas, R.; Hardy, A.; Rosenqvist, A.; Fatoyinbo, T. Mapping Mangrove Extent and Change: A Globally Applicable Approach. *Remote Sens.* **2018**, *10*, 1466. [CrossRef]

21. Samanta, S.; Hazra, S.; Mondal, P.P.; Chanda, A.; Giri, S.; French, J.R.; Nicholls, R.J. Assessment and Attribution of Mangrove Forest Changes in the Indian Sundarbans from 2000 to 2020. *Remote Sens.* **2021**, *13*, 4957. [CrossRef]

22. Hai, P.M.; Tinh, P.H.; Son, N.P.; Thuy, T.V.; Hanh, N.T.H.; Sharma, S.; Hoai, D.T.; Duy, V.C. Mangrove health assessment using spatial metrics and multi-temporal remote sensing data. *PLoS ONE* **2022**, *17*, e0275928. [CrossRef]

23. Wang, Z.; Li, J.; Tan, Z.; Liu, X.; Li, M. Swin-UperNet: A Semantic Segmentation Model for Mangroves and Spartina alterniflora Loisel Based on UperNet. *Electronics* **2023**, *12*, 1111. [CrossRef]

24. Ulku, I.; Akagündüz, E.; Ghamisi, P. Deep Semantic Segmentation of Trees Using Multispectral Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 7589–7604. [CrossRef]

25. Khan, A.; Khan, U.; Waleed, M.; Khan, A.; Kamal, T.; Marwat, S.N.K.; Maqsood, M.; Aadil, F. Remote Sensing: An Automated Methodology for Olive Tree Detection and Counting in Satellite Images. *IEEE Access* **2018**, *6*, 77816–77828. [CrossRef]

26. Flood, N.; Watson, F.; Collett, L. Using a U-net convolutional neural network to map woody vegetation extent from high resolution satellite imagery across Queensland, Australia. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *82*, 101897. [CrossRef]

27. G. Braga, J.R.; Peripato, V.; Dalagnol, R.; P. Ferreira, M.; Tarabalka, Y.; O. C. Aragão, L.E.; F. de Campos Velho, H.; Shiguemori, E.H.; Wagner, F.H. Tree Crown Delineation Algorithm Based on a Convolutional Neural Network. *Remote Sens.* **2020**, *12*, 1288. [CrossRef]

28. Lassalle, G.; Ferreira, M.P.; La Rosa, L.E.C.; de Souza Filho, C.R. Deep learning-based individual tree crown delineation in mangrove forests using very-high-resolution satellite imagery. *ISPRS J. Photogramm. Remote Sens.* **2022**, *189*, 220–235. [CrossRef]

29. Lassalle, G.; de Souza Filho, C.R. Tracking canopy gaps in mangroves remotely using deep learning. *Remote Sens. Ecol. Conserv.* **2022**, *8*, 890–903. [CrossRef]

30. Otero, V.; Van De Kerchove, R.; Satyanarayana, B.; Martínez-Espinosa, C.; Fisol, M.A.B.; Ibrahim, M.R.B.; Sulong, I.; Mohd-Lokman, H.; Lucas, R.; Dahdouh-Guebas, F. Managing mangrove forests from the sky: Forest inventory using field data and Unmanned Aerial Vehicle (UAV) imagery in the Matang Mangrove Forest Reserve, peninsular Malaysia. *For. Ecol. Manag.* **2018**, *411*, 35–45. [CrossRef]

31. Ruwaimana, M.; Satyanarayana, B.; Otero, V.; M. Muslim, A.; Syafiq A., M.; Ibrahim, S.; Raymaekers, D.; Koedam, N.; Dahdouh-Guebas, F. The advantages of using drones over space-borne imagery in the mapping of mangrove forests. *PLoS ONE* **2018**, *13*, e0200288. [CrossRef]

32. Castellanos-Galindo, G.A.; Casella, E.; Mejía-Rentería, J.C.; Rovere, A. Habitat mapping of remote coasts: Evaluating the usefulness of lightweight unmanned aerial vehicles for conservation and monitoring. *Biol. Conserv.* **2019**, *239*, 108282. [CrossRef]

33. Joyce, K.E.; Fickas, K.C.; Kalamandeen, M. The unique value proposition for using drones to map coastal ecosystems. *Camb. Prism. Coast. Futur.* **2023**, *1*, e6. [CrossRef]

34. Schiefer, F.; Kattenborn, T.; Frick, A.; Frey, J.; Schall, P.; Koch, B.; Schmidtlein, S. Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2020**, *170*, 205–215. [CrossRef]

35. Kattenborn, T.; Eichel, J.; Fassnacht, F.E. Convolutional Neural Networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. *Sci. Rep.* **2019**, *9*, 17656. [CrossRef]

36. Navarro, A.; Young, M.; Allan, B.; Carnell, P.; Macreadie, P.; Ierodiaconou, D. The application of Unmanned Aerial Vehicles (UAVs) to estimate above-ground biomass of mangrove ecosystems. *Remote Sens. Environ.* **2020**, *242*, 111747. [CrossRef]

37. Miraki, M.; Sohrabi, H.; Fatehi, P.; Kneubuehler, M. Individual tree crown delineation from high-resolution UAV images in broadleaf forest. *Ecol. Inform.* **2021**, *61*, 101207. [CrossRef]

38. La Rosa, L.E.C.; Sothe, C.; Feitosa, R.Q.; de Almeida, C.M.; Schimalski, M.B.; Oliveira, D.A.B. Multi-task fully convolutional network for tree species mapping in dense forests using small training hyperspectral data. *ISPRS J. Photogramm. Remote Sens.* **2021**, *179*, 35–49. [CrossRef]

39. Weinstein, B.G.; Marconi, S.; Bohlman, S.A.; Zare, A.; White, E.P. Cross-site learning in deep learning RGB tree crown detection. *Ecol. Inform.* **2020**, *56*, 101061. [CrossRef]

40. Wannasiri, W.; Nagai, M.; Honda, K.; Santitamnont, P.; Miphokasap, P. Extraction of Mangrove Biophysical Parameters Using Airborne LiDAR. *Remote Sens.* **2013**, *5*, 1787–1808. [CrossRef]

41. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [CrossRef]

42. Mölder, F.; Jablonski, K.P.; Letcher, B.; Hall, M.B.; Tomkins-Tinch, C.H.; Sochat, V.; Forster, J.; Lee, S.; Twardziok, S.O.; Kanitz, A.; et al. Sustainable data analysis with Snakemake. *F1000Research* **2021**, *10*, 33. [CrossRef]

43. Polidoro, B.A.; Carpenter, K.E.; Collins, L.; Duke, N.C.; Ellison, A.M.; Ellison, J.C.; Farnsworth, E.J.; Fernando, E.S.; Kathiresan, K.; Koedam, N.E.; et al. The Loss of Species: Mangrove Extinction Risk and Geographic Areas of Global Concern. *PLoS ONE* **2010**, *5*, e10095. [CrossRef] [PubMed]

44. Lee, Y.; Park, J. CenterMask : Real-Time Anchor-Free Instance Segmentation. *arXiv* **2020**, arXiv:1911.06667. [CrossRef]

45. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *arXiv* **2018**, arXiv:1802.02611. [CrossRef]

46. Wu, Y.; Kirillov, A.; Massa, F.; Lo, W.Y.; Girshick, R. Detectron2, 2019. Available online: https://github.com/facebookresearch/detectron2 (accessed on 26 June 2023).

47. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*; Curran Associates, Inc.: Red Hook, NY, USA; 2019; pp. 8024–8035.

48. Alonso, I.; Yuval, M.; Eyal, G.; Treibitz, T.; Murillo, A.C. CoralSeg: Learning coral segmentation from sparse annotations. *J. Field Robot.* **2019**, *36*, 1456–1477. [CrossRef]

49. Pavoni, G.; Corsini, M.; Callieri, M.; Fiameni, G.; Edwards, C.; Cignoni, P. On Improving the Training of Models for the Semantic Segmentation of Benthic Communities from Orthographic Imagery. *Remote Sens.* **2020**, *12*, 3106. [CrossRef]

50. Chiang, C.Y.; Barnes, C.; Angelov, P.; Jiang, R. Deep Learning-Based Automated Forest Health Diagnosis From Aerial Images. *IEEE Access* **2020**, *8*, 144064–144076. [CrossRef]

51. Hao, Z.; Lin, L.; Post, C.J.; Mikhailova, E.A.; Li, M.; Chen, Y.; Yu, K.; Liu, J. Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (Mask R-CNN). *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 112–123. [CrossRef]

52. Lee, Y.; Hwang, J.w.; Lee, S.; Bae, Y.; Park, J. An Energy and GPU-Computation Efficient Backbone Network for Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–20 June 2019; pp. 15–20.

53. van der Walt, S.; Schönberger, J.L.; Nunez-Iglesias, J.; Boulogne, F.; Warner, J.D.; Yager, N.; Gouillart, E.; Yu, T.; the scikit-image contributors. scikit-image: Image processing in Python. *PeerJ* **2014**, *2*, e453. [CrossRef]

54. Pavoni, G.; Corsini, M.; Ponchio, F.; Muntoni, A.; Edwards, C.; Pedersen, N.; Sandin, S.; Cignoni, P. TagLab: AI-assisted annotation for the fast and accurate semantic segmentation of coral reef orthoimages. *J. Field Robot.* **2022**, *39*, 246–262. [CrossRef]

55. Hoeser, T.; Kuenzer, C. Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends. *Remote Sens.* **2020**, *12*, 1667. [CrossRef]

56. Hafiz, A.M.; Bhat, G.M. A Survey on Instance Segmentation: State of the art. *Int. J. Multimed. Inf. Retr.* **2020**, *9*, 171–189. [CrossRef]

57. Urbazaev, M.; Thiel, C.; Cremer, F.; Dubayah, R.; Migliavacca, M.; Reichstein, M.; Schmullius, C. Estimation of forest aboveground biomass and uncertainties by integration of field measurements, airborne LiDAR, and SAR and optical satellite data in Mexico. *Carbon Balance Manag.* **2018**, *13*, 5. [CrossRef]

58. Simard, M.; Fatoyinbo, L.; Smetanka, C.; Rivera-Monroy, V.H.; Castañeda-Moya, E.; Thomas, N.; Van der Stocken, T. Mangrove canopy height globally related to precipitation, temperature and cyclone frequency. *Nat. Geosci.* **2019**, *12*, 40–45. [CrossRef]

59. Huang, B.; Reichman, D.; Collins, L.M.; Bradbury, K.; Malof, J.M. Tiling and Stitching Segmentation Output for Remote Sensing: Basic Challenges and Recommendations. *arXiv* **2018**, arXiv:1805.12219.

60. Fuchs, H.P. Ecological and palynological notes on Pelliciera rhizophorae. *Acta Bot. Neerl.* **1970**, *19*, 884–894. [CrossRef]

61. Allen, J.A. *Rhizophora mangle* L. In *Tropical Tree Seed Manual: Part II, Species Descriptions. Agricricultural Handbook*; Vozzo, J., Ed.; U.S. Department of Agriculture: Washington, DC, USA; 2002; Volume 712, pp. 690–692.

62. Suhardiman, A.; Tsuyuki, S.; Setiawan, Y. Estimating Mean Tree Crown Diameter of Mangrove Stands Using Aerial Photo. *Procedia Environ. Sci.* **2016**, *33*, 416–427. [CrossRef]

63. Alon, A.; Festijo, E.; Casuat, C. Tree Extraction of Airborne LiDAR Data Based on Coordinates of Deep Learning Object Detection from Orthophoto over Complex Mangrove Forest. *Int. J. Emerg. Trends Eng. Res.* **2020**, *8*, 2107. [CrossRef]

64. Schürholz, D.; Chennu, A. Digitizing the coral reef: Machine learning of underwater spectral images enables dense taxonomic mapping of benthic habitats. *Methods Ecol. Evol.* **2023**, *14*, 596–613. [CrossRef]

65. Cheng, B.; Collins, M.D.; Zhu, Y.; Liu, T.; Huang, T.S.; Adam, H.; Chen, L.C. Panoptic-DeepLab: A Simple, Strong, and Fast Baseline for Bottom-Up Panoptic Segmentation. *arXiv* **2020**, arXiv:1911.10194.