Accurate delineation of individual tree crowns in tropical forests from aerial RGB imagery using Mask R-CNN

James Ball, Sebastian Hickman, Tobias Jackson, Xian Jing Koay, James Hirst, William Jay, Matthew Archer, Mélaine Aubry-Kientz, Grégoire Vincent, David A. Coomes

Remote Sensing in Ecology and Conservation
Open Access
ZSL

RESEARCH ARTICLE

Accurate delineation of individual tree crowns in tropical forests from aerial RGB imagery using Mask R-CNN

James G. C. Ball[1,2,3], Sebastian H. M. Hickman[4,5], Tobias D. Jackson[1,2], Xian Jing Koay[2], James Hirst[6], William Jay[7], Matthew Archer[8], Mélaine Aubry-Kientz[9], Grégoire Vincent[3] & David A. Coomes[1,2]

[1]Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK
[2]Conservation Research Institute, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK
[3]UMR AMAP, University of Montpellier, IRD, CNRS, CIRAD, INRAE, Montpellier, France
[4]Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK
[5]The Alan Turing Institute, 96 Euston Road, London NW1 2DB, UK
[6]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK
[7]Plymouth Marine Laboratory, Prospect Place, Plymouth PL1 3DH, UK
[8]Research Software Engineering, University of Cambridge, Trinity Lane, Cambridge CB2 1TN, UK
[9]AgroParisTech, UMR EcoFoG, Kourou, French Guiana

Keywords
Convolutional neural networks, deep learning, Detectron2, forest monitoring, Mask R-CNN, tree crown delineation, tree crown segmentation, tree growth, tree mortality, tropical forests

Correspondence
James G. C. Ball and David A. Coomes, Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK. Tel: +44 (0) 1223 33900; E-mail: ball.jgc@gmail.com and dac18@cam.ac.uk

James G. C. Ball, Sebastian H. M. Hickman and Tobias D. Jackson contributed equally to the work (first authors).
Associate Editor: Nicola Clerici

Received: 10 January 2023; Revised: 6 March 2023; Accepted: 27 March 2023

doi: 10.1002/rse2.332

Abstract

Tropical forests are a major component of the global carbon cycle and home to two-thirds of terrestrial species. Upper-canopy trees store the majority of forest carbon and can be vulnerable to drought events and storms. Monitoring their growth and mortality is essential to understanding forest resilience to climate change, but in the context of forest carbon storage, large trees are underrepresented in traditional field surveys, so estimates are poorly constrained. Aerial photographs provide spectral and textural information to discriminate between tree crowns in diverse, complex tropical canopies, potentially opening the door to landscape monitoring of large trees. Here we describe a new deep convolutional neural network method, *Detectree2*, which builds on the Mask R-CNN computer vision framework to recognize the irregular edges of individual tree crowns from airborne RGB imagery. We trained and evaluated this model with 3797 manually delineated tree crowns at three sites in Malaysian Borneo and one site in French Guiana. As an example application, we combined the delineations with repeat lidar surveys (taken between 3 and 6 years apart) of the four sites to estimate the growth and mortality of upper-canopy trees. *Detectree2* delineated 65 000 upper-canopy trees across 14 km^2 of aerial images. The skill of the automatic method in delineating unseen test trees was good (F_1 score = 0.64) and for the tallest category of trees was excellent (F_1 score = 0.74). As predicted from previous field studies, we found that growth rate declined with tree height and tall trees had higher mortality rates than intermediate-size trees. Our approach demonstrates that deep learning methods can automatically segment trees in widely accessible RGB imagery. This tool (provided as an open-source Python package) has many potential applications in forest ecology and conservation, from estimating carbon stocks to monitoring forest phenology and restoration.Python package available to install at https://github.com/PatBall1/Detectree2.

Introduction

Intact tropical forests are an important component of the global carbon cycle: they are major carbon stores and significant carbon sinks (Pan et al., 2011). However, the strength of the carbon sink is diminishing as a result of global warming (Brienen et al., 2015; Hubau et al., 2020) and there are concerns that forests are reaching a tipping point beyond which they could switch irreversibly to open savanna systems (Chai et al., 2021). Forecasting the future of tropical forests is challenging because little is known about the ways different species will respond to changing climate, or the resilience provided by that diversity (Fisher et al., 2018; Gallup et al., 2021; Koven et al., 2020; Restrepo-Coupe et al., 2021). To understand the likely responses of forests to further climate change, ecosystem models need to represent growth and mortality processes of individual trees more accurately than is currently the case (Kellner et al., 2019; Piponiot et al., 2022; Zuidema & van der Sleen, 2022).

Remote sensing of individual upper-canopy trees can improve estimates of forest carbon (Dalponte & Coomes, 2016) and provide a means of tracking growth and mortality over large spatial scales. Traditional monitoring approaches rely on measuring stem dimensions in permanent inventory plots, and periodically revisiting those plots to assess recruitment, growth and mortality (Chave et al., 2019). However, the coverage of such plots in the tropics is limited ($\sim$0.0002% of tropical forests are sampled by the main plot networks) and their locations are often dictated by ease of access rather than by robust statistical sampling designs (Davies et al., 2021; ForestPlots.net et al., 2021; Marvin et al., 2014). Furthermore, upper-canopy trees store the majority of carbon in tropical forests, but few of them are sampled in inventory plots (Coomes et al., 2017; Lutz et al., 2018; Meakem et al., 2018). This under-sampling is particularly problematic when assessing impacts of climate change because upper-canopy trees are most vulnerable to periods of water shortage (Gora & Esquivel-Muelbert, 2021; Stovall et al., 2019) which are increasing in frequency (IPCC, 2021). Remote sensing has the potential to overcome these sampling challenges by providing wall-to-wall maps that can be used to monitor millions of upper-canopy trees.

Remote sensing of individual trees has mostly focused on airborne lidar data, which are used by the forestry industry to map trees at landscape scales (Zhen et al., 2016). Delineating individual trees from airborne lidar datasets is most successful for conifer because their apical dominance results in clear local height maxima that make tree crowns easily distinguishable (Dalponte & Coomes, 2016; Hastings et al., 2020), but complex tropical canopies have presented a far greater challenge for lidar delineation (Aubry-Kientz et al., 2019). Tropical forest canopies are often densely packed with partially interwoven crowns which point-cloud clustering algorithms can struggle to distinguish (Aubry-Kientz et al., 2021). Furthermore, lidar surveys require expensive aircraft (airplanes, helicopters or high-end drones) and sensors whereas standard RGB imagery can be collected cheaply with drones.

Automatic delineation of trees in RGB photographs can harness colour and texture information to distinguish trees, even if they are structurally similar (Almeida et al., 2021; Iglhaut et al., 2019). Most current methods of individual tree identification from RGB imagery use bounding boxes (Figure S3) (Aparecido et al., 2019; Weinstein et al., 2019, 2021), but more exact delineation of the edges of tree crowns would provide information on the crown area and lateral growth and avoid mixing signals from neighbouring vegetation. Recent advances in neural network approaches to computer vision provide opportunities to recognize individual trees from standard digital photographs taken from drones. A class of machine learning algorithms called deep convolutional neural networks (CNNs) is revolutionizing vegetation science through its ability to exploit spatial structures and automatically extract high-level features from image data (e.g. analyses of satellite imagery) (Kattenborn et al., 2021; Mugabowindekwe et al., 2022; Zhu et al., 2017). In the field of computer vision, exactly segmenting individual objects of interest from an image is known as *instance segmentation*. The Mask R-CNN algorithm has shown promise in tree crown identification and delineation in plantations (Hao et al., 2021; Kunyong et al., 2022), pine forests (Gensheng et al., 2022; Ocer et al., 2020), urban woodlands (Ocer et al., 2020) and forest fragments (Braga et al., 2020). Mask R-CNN has features that could allow it to overcome the challenges of delineating crowns in complex tropical canopies by discriminating based on the spectral and textural signals which are rich due to the phylogenetic diversity.

Here, we describe Detectree2, a system that automatically detects tree crowns from aerial RGB imagery. We adapted Facebook AI's Mask R-CNN algorithm (the *Detectron2* release), which has models that have been pre-trained on a wealth of available image data that can be transferred to new tasks (He et al., 2017; Wu et al., 2019). We trained and evaluated Detectree2 on four tropical forest sites. In total, 3797 manually delineated tree crowns were used of which 1530 spatially separated crowns were reserved to evaluate the model. We evaluated the performance with F_1 scores, which quantify the skill of the method in delineating individual tree crowns accurately. We expected a model trained at one site to drop in performance when transferred to making predictions of crowns at the other sites and that supplying a greater

variety of training data would boost performance. As an example ecological application, we deployed the trained model across 14 km^2 of airborne RGB imagery, automatically delineating 65 786 tree crowns. For context, this area is approximately 40% the total area of forest inventory plots in the main plot networks across the tropics (Davies et al., 2021; ForestPlots.net et al., 2021). We then combined these crowns with repeat airborne lidar data to investigate the growth and mortality rates of upper-canopy trees in relation to their height. Regional and global syntheses of forest inventory data suggest that growth slows down and mortality rates increase with tree size (Coomes et al., 2003; Hurst et al., 2011; Iida et al., 2014; Muller-Landau et al., 2006; Richardson et al., 2009). We therefore expected to find the tallest trees we sampled to have lower growth rates and higher mortality rates than shorter trees. The Detectree2 Python package is available to install and apply on new regions.[1]

Materials and Methods

Study sites

The analyses were conducted at four locations across three tropical field sites:

1. **Sepilok Forest Reserve (East and West), Sabah, Malaysia (5° 50′ N, 177° 56′ W)**
2. **Danum Valley Conservation Area, Sabah, Malaysia (4° 57′ N, 177° 41′ W)**
3. **Paracou Field Station, French Guiana (5° 16′ N 52° 55′ W)**

Danum Valley hosts lowland tropical rain forests dominated by dipterocarp species that are among the tallest forests on the planet (Shenkin et al., 2019). The available data from Sepilok included ecologically distinctive areas to the East and West. Sepilok West consists mostly of tall forest (similar to Danum), while Sepilok East is a heath forest growing on shallow soils overlying sandstone, containing smaller, more densely packed trees (Coomes et al., 2017). All three sites in Malaysia experience a similar climate with approximately 2300 mm rainfall per year with the wettest months between November and February (Nilus et al., 2011). Paracou contains lowland tropical rain forest mostly on shallow ferralitic soils that lay on a variably transformed loamy saprolite (Gourlet-Fleury et al., 2004). The mean annual rainfall is approximately 3000 mm with a 3-month dry season from mid-August to mid-November (Wagner et al., 2011). See Supplementary Note 1 for more details on the study sites.

Remote sensing data

Airborne RGB surveys were conducted at all four sites using manned aircraft (Table 1). Repeat lidar surveys were also conducted at all four location (see Table 1, noting different sensors and altitudes between flights in Sabah). We analysed RGB imagery from 3.85 km^2 of Malaysian forest, with a ground resolution of 10 cm. In Paracou, we sampled 10.2 km^2 of imagery, with an 8 cm ground resolution. The raw imagery was orthorectified, georeferenced and collated into homogeneous mosaics using structure from motion in AgiSoft Metashape (AgiSoft, 2021; Westoby et al., 2012) in Sabah. In Paracou the imagery was orthorectifed using TerraPhoto to the Canopy surface model derived from simultaneously acquired lidar data.

Manual tree crown data

To train and evaluate our automatic delineation approach, we created a manually labelled dataset of trees at all four sites. We generated our delineations using both

Table 1. Remote sensing data sources.

Scan dates	Region	Modality	Resolution	Pulse density	Beam divergence (mrad)	Scanning angle	Altitude (m)	Sensor
23-Oct-2014	Danum	RGB	10 cm	-	-	-	796	Leica RCD105
01-Nov-2014	Danum	Lidar	1 m	5 pls m^{-2}	<0.22	±14°	2000	Leica ALS50-II
19-Feb-2020	Danum	Lidar	1 m	35 pls m^{-2}	<0.5	±30°	200	RIEGL LMS-Q560
10-Oct-2014	Sepilok	RGB	10 cm	-	-	-	796	Leica RCD105
05-Nov-2014	Sepilok	Lidar	1 m	16 pls m^{-2}	<0.22	±14°	2000	Leica ALS50-II
15-Feb-2020	Sepilok	Lidar	1 m	42 pls m^{-2}	<0.5	±30°	200	RIEGL LMS-Q560
19-Sep-2016	Paracou	RGB	8 cm	-	-	-	800	IXA180 Phase One
19-Sep-2016	Paracou	Lidar	1 m	35 pls m^{-2}	<0.25	±30°	800	RIEGL LMS-Q780
19-Nov-2019	Paracou	Lidar	1 m	35 pls m^{-2}	<0.25	±30°	800	RIEGL LMS-Q780
15-Nov-2019	Paracou	Lidar	1 m	35 pls m^{-2}	<0.25	±30°	800	RIEGL LMS-Q780

The exact location of the sites is described in "Study sites" section. Resolution is given as ground resolution for the RGB imagery and as the processed CHM resolution for the lidar scans. Beam divergence is given at the $1/e^2$ points. Sepilok West and Sepilok East were separated for analysis due to the different characteristics of the Sepilok forest in these two areas.

RGB and lidar data and, in the case of Paracou, supplementary hyperspectral layers. We used several techniques to improve the accuracy of crown delineation, including manipulating the contrast and saturation of the RGB image to exaggerate differences between the crowns, and using a mask of the lidar data to remove irrelevant parts of the RGB imagery. These techniques meant that the vast majority of tree crowns were separable by eye but, it should be noted, that in rare cases, tree crowns were near impossible to delineate with certainty and the labeller's best estimate was used. See Supplementary Note 2 for further details.

We trained and tested our model with a total of 3797 manually delineated tree crowns across Paracou (1267), Danum (521), Sepilok West (1038) and Sepilok East (971). The crowns from Paracou were validated in the field with an expert local botanist, whereas the crowns in Malaysia were drawn by inspection of the remote sensing products. Four individuals performed the manual delineations which provided the network with variability in the inputs.

Data preparation

The orthomosaics and corresponding crown polygons were tiled into squares of approximately 100 m × 100 m to be ingested into the network (40 m core area, 30 m overlapping buffers). To be included in the training and test sets, a minimum crown polygon area coverage of a tile was set at 40%. Including overly sparse tiles was likely to lead to poor algorithm sensitivity while being too strict with coverage would have limited the amount of training and testing data available.

If training and test crowns are close to one another, spatial autocorrelative effects are likely to inflate the reported performance (Kattenborn et al., 2022). To avoid this, individual tiles (rather than individual crowns) were assigned to training and test sets ensuring spatial separation. Approximately 10% of the tiles from each site were reserved at random for testing. To avoid contamination of the test set, tiles with any overlap with the test tiles (including with the buffer) were excluded from the training set. The training tiles were further partitioned into 5-folds for cross validation. This allowed for the tuning of parameters and the implementation of early stopping (see "Training and model selection" section) without exposing the test set. Details of the data processing are described in Supplementary Note 3.

Model architecture and parameterization

Instance segmentation combines object detection with object segmentation. Once an object has been detected in a scene, a region of interest (as a bounding box) is established around the object. Then a 'segmentation' is then carried out to identify which pixels within the region of interest make up the object of interest (and which lie outside; see Figure S2 for an example).

We adapted Facebook AI's Mask R-CNN architecture as it was the best in class algorithm upon release for instance segmentation when tested on the Microsoft COCO (Common Objects in Context) benchmark (He et al., 2017; Lin et al., 2014) and has since been updated (as Detectron2) with improved training efficiency, documentation and transferability for integration into bespoke tools (Wu et al., 2019). We adapted the Detectron2 computer vision library to handle geospatial inputs/outputs and perform the delineation of individual tree crowns. The library performs instance segmentation by generating object 'masks' which exactly circumscribe the objects in the image (see Figure S2 for an example prediction). It also has a 'model zoo'[2] from which specific model architectures with weights from a variety of pre-training regimes can be loaded. Taking a pre-trained model (weights) and retraining it to perform a novel task (e.g. delineating trees from aerial imagery) is an example of *transfer learning* which can drastically reduce the amount of training data required to achieve acceptable performance on the new task (Weiss et al., 2016). We selected the R101-FPN configuration[3] as it has 'the best speed/accuracy tradeoff' of the architectures available (Wu et al., 2019). Each object predicted by Detectron2 is associated with a confidence score which relates to how sure the network is in its prediction. A suitable threshold can be selected to optimise accuracy or balance precision and recall. Additional details are given in Supplementary Note 4 and for full technical specifications, one should refer to the original papers and the Detectron2 repository (He et al., 2017; Wu et al., 2019).

Training and model selection

Training, tuning and model selection were performed with the five folds of training data tiles (see "Data preparation" section). To test the effect of volume and diversity of training data on performance we employed three training regimes: (1) Training on data of a single site, (2). Training on all sites ('combined'), (3) Training on all sites and then trained with a fixed training period on the single site. The idea behind (Hubau et al., 2020) was to train on the full available data and then 'hone' the delineator based on the local context.

Typically, a deep CNN would require several thousand training examples to learn a new task. This is a challenge in the case of tree crown delineation as manual delineation is time consuming. The burden was reduced by

transferring a model trained on a different instance segmentation task (Lin et al., 2014). Additionally, the training data were augmented by applying several randomly applied transformations to the training cases including vertical and horizontal flips, rotation, and varying the saturation and contrast of the image.

The model hyperparameters (Table S1) were tuned with a Bayesian hyperparameter sweep implemented on wandb.ai.[4] This is an automated process that allows an automated agent to iteratively adjust hyperparameters to optimize accuracy. The best performing models and optimal confidence threshold for a given model (see "Model architecture and parameterization" section) were selected based on the F_1 score (see "Performance evaluation" section) on the validation fold.

See Supplementary Note 4 for more details on model architecture, training and validation. The Colab (Jupyter) notebooks in the GitHub repository[1] illustrate the best practices for training and selecting models.

Performance evaluation

After tuning and training, the best performing models were taken forward to be evaluated against the test tiles. Matches between predictions and manual crowns (i.e. true positives) were identified by assessing the degree of spatial overlap between possible pairs. A minimum area threshold for valid crowns was set to 16 m^2. This removed fewer than 2% of manual crowns and introduced a level of consistency between sites and between the effort given by the manual delineators. The threshold was small enough to allow for an inclusive analysis of the variation in performance by tree height. Crown overlap was calculated as the area intersection over union (Fig. 1):

$$\mathrm{IoU}(A,B) = \frac{A \cap B}{A \cup B} \quad (1)$$

where A is the automatically delineated crown area and B is the manually delineated crown area. An IoU of an overlapping pair of more than 0.5 was considered a match. This is a commonly used threshold in similar studies (e.g. Aubry-Kientz et al., 2019; Hao et al., 2021) that allows for small discrepancies in alignment and outline. These 'true positives' as well as the unmatched predictions (false positives) and unmatched manual crowns (false negatives) were used to calculate the precision, recall and F_1 score of the automatic predictions.

Despite the best efforts of the manual delineators and selecting for tiles with high manual crown coverage, the manual crowns were inevitably an incomplete representation, so recall (fraction of relevant instances retrieved) was an insightful metric. However, to ensure balance with precision we used the balanced F_1 score metric to assess

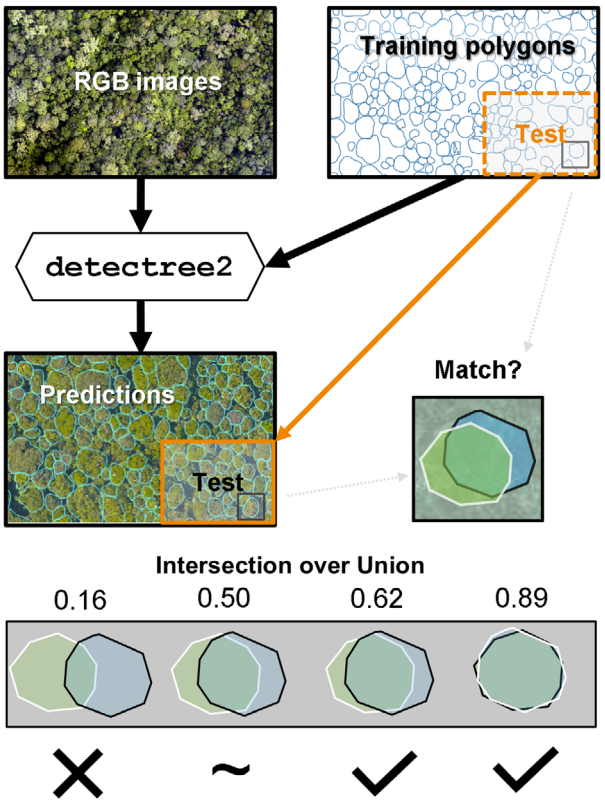

Figure 1. The automatic tree crown delineation workflow. Manually delineated crowns are randomly split into training and test sets (though the figure suggests that the sets were determined geographically, this is purely for visual clarity). The Mask R-CNN framework combines the training set with the RGB imagery to learn how to delineate automatically from RGB images. A set of automatic predictions are produced across the entire RGB image and compared with the test set to evaluate the performance of the automatic delineations. Intersection over union (IoU) is used to determine when an automatic crown has been successfully matched with a manual crown.

and compare the accuracy of the models. This approach is not biased by tree crown area and is widely used in tree crown segmentation studies (Braga et al., 2020; Hao et al., 2021). See Supplementary Note 5 for more details on the evaluation metrics.

To evaluate the performance of Detectree2 across tree heights, we assigned a height to each test crown (based on the median pixel value of the initial CHM within the crown) and arranged them into 5 m height bins. The shortest bin (0–5 m) at each site was iteratively merged with the next shortest bins until more than 10 individuals were represented (e.g. at Paracou 17 trees with a median height of 22.97 m fell into the lowest bin of 0–25 m). An equivalent process was used to define the highest bin at each site. The median tree height was calculated within each bin.

Transferability across sites

To determine whether models were able to generalize across different tropical forest areas, we evaluated the performance of the models when trained on one site and transferred to others. We compared these performances against the effect of using the 'combined' training regimes described in "Training and model selection" section.

Application to monitoring growth and mortality

We applied our best models for each site to their entire tiled orthomosaics (excluding the very edges where distortion is prominent) to generate site wide crown maps. We combined these crown delineations with repeat lidar surveys to determine the height changes in individual trees in our four sites. We determined the relationship between tree height and tree growth by fitting a robust least squares regression (Audibert & Catoni, 2011) to the data. Robust least squares was chosen to minimize the effects of outliers and mortality events on the regression. We note that here we are measuring the vertical growth of trees, instead of the growth in diameter at breast height (DBH) which is traditionally measured in forest inventory data.

To estimate mortality rates, we needed a suitable metric to identify mortality events. We took a statistical approach defining a mortality event as a negative change in height of more than three standard deviations below the robust least squares fit. This allowed for the possibility that a mortality event may uncover another layer of vegetation rather than the forest floor. This choice was ratified by manual inspection of trees meeting this threshold, and confirming that they constituted mortality events. Annual rates were determined by dividing by the time between lidar scans.

Differences in lidar scanning parameters (pulse density, scanning angle, flight height etc.) can bias height estimates (Roussel et al., 2017). For this reason, we resisted reporting a direct comparison of reported growth and mortality rates between sites. As our focus here was on demonstrating the use of Detectree2 for locating crowns, we considered that a detailed exploration of the potential biases from the lidar data beyond the scope of the current paper.

Computation

Training deep CNN models can be computationally expensive and benefits from the availability of GPUs. Model training and evaluation was performed on the Google Colab (Pro) platform which employed Intel(R) Xeon(R) CPU at 2.30 GHz with 12.8 GB RAM and Tesla P100-PCIE-16GB GPUs. On this platform, model training always completed within 2 h.

Results

Performance by site and tree height

Detectree2 located and delineated trees well ($F_1 > 0.56$) across all sites (Table 2). It performed better in the tall dipterocarp dominated forests of Danum and Sepilok West and worse in the more compact forests of Sepilok East and Paracou. Indeed, Danum, the site with the best performance, had the greatest proportion of the tallest class of trees of any of the sites (see "Extended results" for a full table of results). There was no apparent relationship between the amount of training data available at a site and the performance of the automatic delineator suggesting forest structure was the key determinant of accuracy. Where predictions were not accurate, it was slightly more likely to be from under-segmention (0.23–0.45) than over-segmentation (0.13–0.23) (Clinton et al., 2010) across all sites (see "Extended results") (Fig. 2).

Across all sites, accuracy improved with tree height (Fig. 3). This is likely due to the increased crown visibility of tall trees in the RGB images. Paracou has the least well differentiated canopy of all sites which may explain relatively poor performance there.

Performance between forest types

Danum and Sepilok West have tall dipterocarp dominated forests whereas Paracou and Sepilok East have a more compact forest structure. As we expected, performance degrades when testing a model on a different forest type to the one it was trained on (Fig. 4A). For example, the performance at the forests of Sepilok West is significantly degraded when a model trained on Sepilok East or Paracou is used. In contrast, there is no drop in performance

Table 2. Precision, recall and F_1 score of Detectree2 tree crown delineations by site as measured against the manual crowns of the test set tiles.

	# test trees	Precision	Recall	F_1 score
Paracou	381	0.595	0.543	0.568
Danum	278	0.713	0.662	0.687
Sepilok East	167	0.612	0.653	0.632
Sepilok West	704	0.640	0.656	0.648
Average (sum)	(1530)	0.640	0.629	0.634

The unweighted means of the metrics across individual sites are given as a summary overall performance.

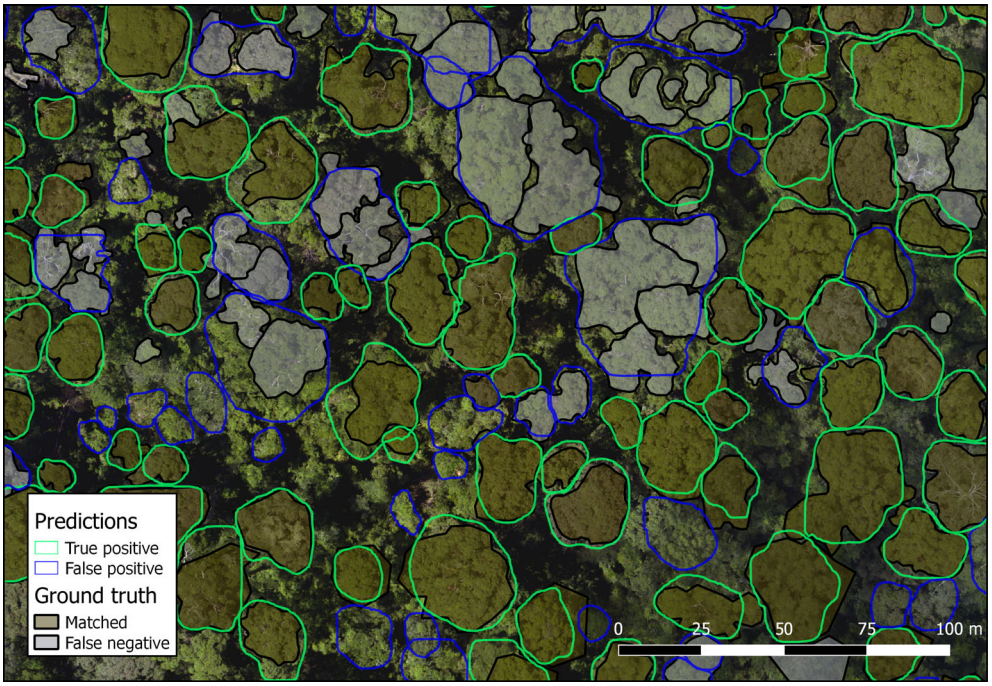

Figure 2. An area of predicted crowns (transparent) overlaid on ground truth crowns (shaded with black outlines) at Danum. Colours and shading are used to indicate whether individual crowns have been successfully delineated. Some examples of under-segmentation (where a single prediction encompasses multiple ground truth crowns) and over-segmentation (where multiple predictions false try to split a single ground truth crown) are visible.

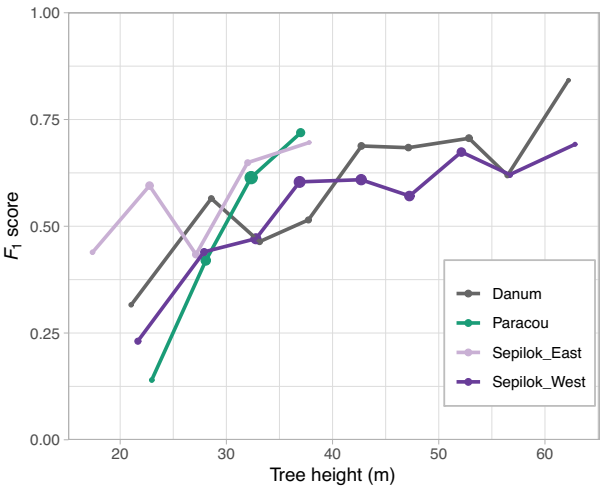

Figure 3. F_1 scores of the tree crown delineations at the four different sites across tree heights. Bins of 5 m width were used to calculate F_1 score and corresponding median tree height. Point area is scaled by the number of test trees in the bin.

for predictions at Danum when the Sepilok West model is used and there is even a slight increase in performance for Sepilok East predictions when the Paracou model is used.

In general, the model that was trained on all sites at once ('combined') outperformed the models that were trained on just a single site with the exception of Paracou (Fig. 4B). Across the board, the best performing models were those that were exposed to data from all sites before being trained for a fixed number of iterations at the site to be predicted on. This suggests that providing a broad range of input data helps the networks to learn the key visual features but further tuning for local context helps maximize performance.

Application: growth and mortality

One application of Detectree2 is to study tall tree growth and mortality rates. To do this, we overlaid Detectree2's tree crown predictions at the start date for each site on repeat lidar data (as canopy height models described in Supplementary Note 1) to retrieve the tree height dynamics over time.

We were able to estimate the relationships between tree height and tree growth for each site by fitting robust least squares linear relationships between the two variables for Danum, Paracou, Sepilok East and Sepilok West (Fig. 5). The regression coefficients and intercepts are given in Table S2. The growth rate decreased with tree height in all sites.

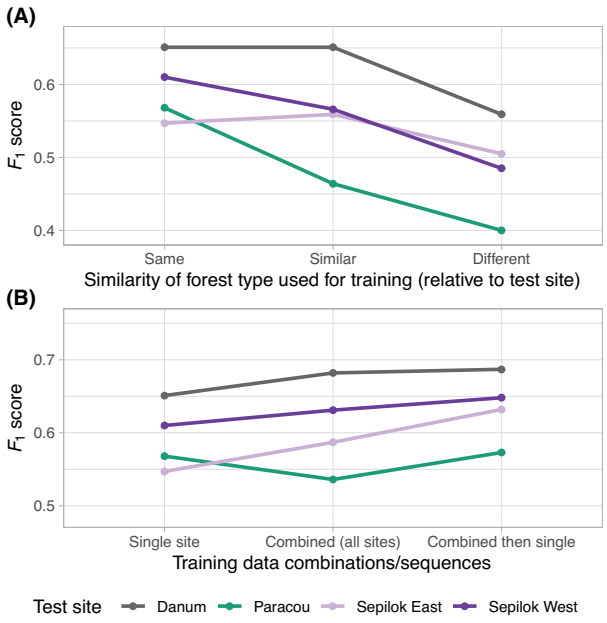

Figure 4. Sensitivity of Detectree2 delineation accuracy to the training data used. In (A), 'Same' indicates that training and testing took place at the same site. Sepilok West and Danum are 'similar' forest types in that they are tall dipterocarp dominated forests in contrast to Sepilok East and Paracou that are shorter forests with a larger number of trees per hectare. As each site has two 'different' sites and an average was calculated for the F_1 score. (B) shows the change in performance that occurs through employing different combinations of the training data. That can be just a single site, all sites at once ('combined') or all sites at once followed by a limited number of iterations on the site to be tested on.

We assumed trees had died when their height decreased substantially. To evaluate this quantitatively, we fitted a robust least squares to the height change, against the original height of the tree, taking trees that were three standard deviations below the mean of the regression fit to be mortality events. The robust least squares regression differs from ordinary least squares as outliers contribute less to the regression fit. Therefore the robust least squares weights the fit towards those trees which did not suffer large height loss, and by taking the threshold to be three standard deviations we aim to identify only those trees that are outside the assumed normal distribution of typical tree growth and measurement error. Furthermore, as the robust least squares still incorporates outliers when fitting the data, three standard deviations was considered sufficient to identify mortality events. Figure 5 illustrates how certain trees were identified as mortality events and some visual examples of mortality events from Paracou are given alongside (see Figure S12 for the other sites).

The mortality rates increased with tree height (Fig. 6). The given uncertainty estimates were determined by bootstrapping. A detailed table of the growth and mortality rates can be found in "Extended results."

Discussion

Improved tropical crown delineation

Accurately delineating trees in remote sensing data is a long-standing problem in ecology and conservation, and would enable us to efficiently monitor large areas of forests. Detectree2 addresses this problem, delineating individual trees in aerial RGB imagery with high precision and recall. We used Detectree2 to automatically delineate 65 786 trees across three tropical forests. We found that the accuracy of Detectree2 increased with tree height, meaning that that the tall trees which store the most carbon are also the most reliably delineated.

Detectree2 performed well across a range of challenging, dense, closed canopy forests. It is able to exactly delineate highly irregular crowns within the jigsaw of the canopy rather than simply identifying a bounding box. This opens up new opportunities for tracking dynamic processes including growth and demographics (as demonstrated here) as well as phenology (where bounding boxes would risk mixing signals). Furthermore, Detectree2 is relatively accessible since it requires a low number of manually delineated trees as training data compared with other methods (Braga et al., 2020; Weinstein et al., 2019). These advantages are partly due to Detectree2 being built on a state-of-the-art pre-trained model. While direct comparison is impossible due to the different test data and the differing tasks (instance detection vs. segmentation), our method performs comparably with the results of Weinstein et al. (2019), which reported a tree crown recall of 0.69, a precision of 0.61 and an F_1 score of 0.65. Our results did not match the Mask R-CNN performance reported in Braga et al. (2020) but this study is based on semi-synthetic images (i.e. constructed by stitching together existing images) of forests and so is not directly comparable.

Generalizability across sites

There was no obvious relationship between the amount of training data available at a site and the accuracy attained there. Rather, forest type and tree height distribution seemed to be the key factors for determining accuracy. The well differentiated forest at Sepilok West and Danum were the easiest to delineate while the lowest accuracy was in Paracou which has little variation in the height of the visible canopy. Furthermore, at Paracou it is common to observe crowns mixing and growing into each other which makes visually separating the crowns

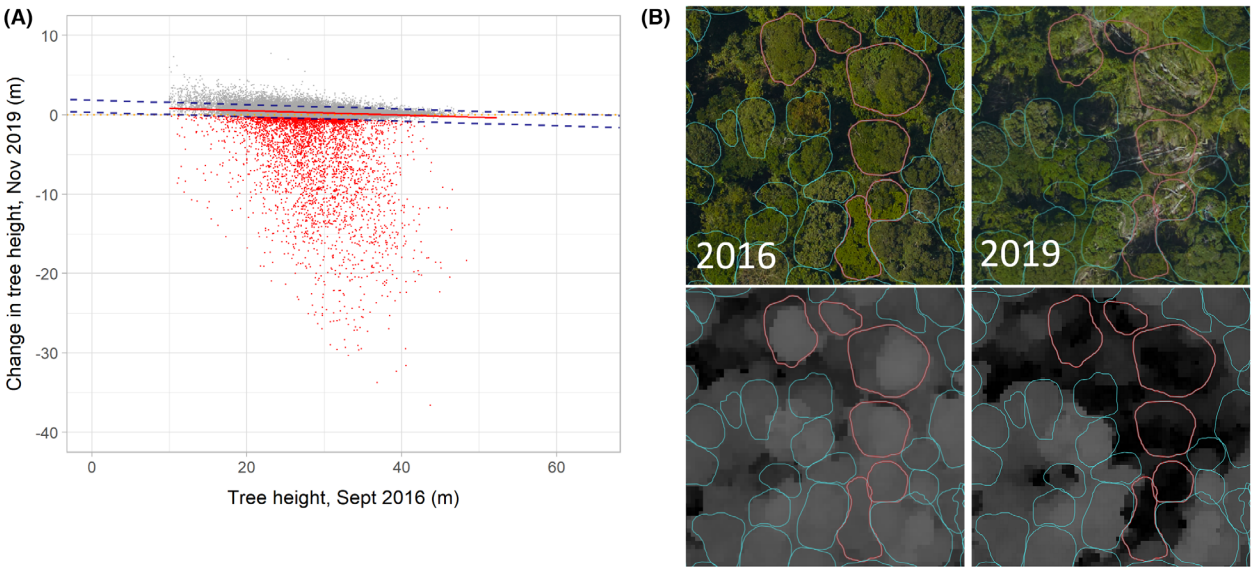

Figure 5. (A) shows the robust least squares fit for change in height and tree height for Paracou. The dashed lines indicate three standard deviations either side of the best fit and red points below the lower bound indicate likely mortality events. (B) Illustrates how predictions were overlaid on lidar data and shows mortality events clearly visible in the lidar and the RGB imagery. The 2016 imagery is shown on the left, 2019 on the right. Crown delineations are based on the earlier imagery.

challenging. This in turn is down to soil type and other biogeographic factors.

We found that the accuracy dropped when transferring a model trained on one forest type to predict on another. However, we found that Detectree2 can be quickly trained to perform well on new areas of forest using around 10 images (each $\sim$1 ha in scale) with all visible tree crowns manually delineated. This manual delineation represents approximately 4 h work. The best performing models were those that were exposed to training data from all the sites and then 'honed' with a limited number of training iterations on the site to be predicted on. This suggests that our trained models (provided freely with the Python package[5]) could be transferred to a new site with very little manual data or training iterations.

We note that the manual delineations were done by different people focusing on different parts of the sites. There was no clear effect of different delineators on the results but this would be somewhat confounded with site differences.

Application: growth and mortality rates

Tall trees store the majority of forest carbon and dominate many important forest nutrient cycles. However, they are rare and therefore poorly represented in traditional field inventories (Hurst et al., 2011) which makes estimating their growth and mortality rates particularly challenging (Coomes et al., 2003; Iida et al., 2014; Muller-Landau et al., 2006; Richardson et al., 2009). Tall trees are also particularly sensitive to the effects of climate change, such as increased wind speeds and drought (Gora & Esquivel-Muelbert, 2021), and as such tracking their dynamics over time is increasingly important. Recent remote sensing studies are bringing new insights into disturbance patterns by mapping the gaps left in the forest canopy after a tree (or multiple trees or branches) have fallen (Araujo et al., 2021; Cushman et al., 2022; Huertas et al., 2022). Tracking individual trees over time instead of gaps will make it easier to interpret our results in an ecological context and also to compare the results more directly with the available field inventory data.

Across all sites taller trees had higher mortality rates and lower growth rates. This aligns with large scale analyses of field-based studies (Iida et al., 2014). The apparent higher growth and mortality rates in French Guiana compared with the sites in Malaysia was potentially a result of biases introduced to the variation in scan parameters (flight height, pulse density, time of year etc.) and so the values should not be directly compared across sites. Inventory data show that mean DBH growth for trees at Paracou was 1.2 mm $year^{-1}$ (Wagner et al., 2010) compared with 0.9 mm $year^{-1}$ in Sepilok East, 1.1 mm $year^{-1}$ in Sepilok West and 0.5 mm $year^{-1}$ in Danum (Ordway et al., 2022; Piponiot et al., 2022). We note that these inventory measured DBH growth rates may not be directly comparable with the height growth measured in this study. Another caveat is that we defined mortality as a drop in height of more than a statistically determined threshold. We do not verify

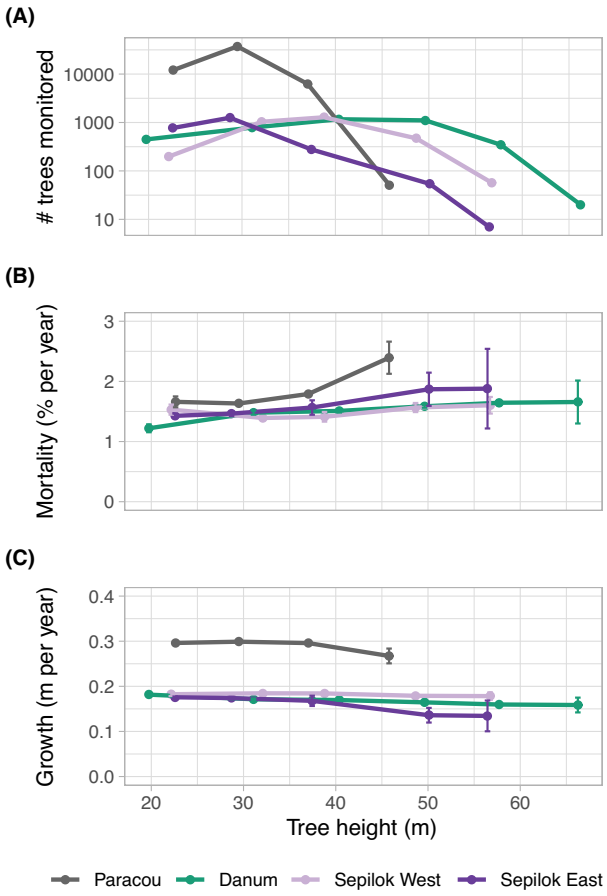

Figure 6. (A) shows the distribution of tree heights per site. (B) shows the mortality rates of trees of different heights in each site, and (C) gives the growth rate of trees split by height bin. Due to biases in tree height measurements that can arise from differences in lidar scan parameters we advise against a direct comparison of growth and mortality rates between sites. Uncertainty estimates were determined by bootstrapping.

directly that the tree has died, although it is likely that it has snapped or uprooted. Further analysis would help to understand the discrepancy in observed height change at Paracou in comparison with the other sites but is not the focus of the current study. Nevertheless, we believe this example application demonstrates the utility of Detectree2 in expanding the sample of trees under observation.

Future methodological developments and applications

Detectree2 performs impressively when delineating tall trees but it fails to delineate a significant proportion of trees. There is considerable scope to increase the quantity and variety of training data by labelling more trees by hand. A more robust approach to compensating for shadowed regions may also support the detection of trees otherwise obscured by their neighbours.

The fact that Detectree2 can be quickly trained to perform well on a new type of forest and imagery demonstrates that it is a useful tool for forest management. Many conservation or restoration projects have access to low-cost imagery from drones or satellites. Detectree2 would allow them to quickly quantify and track the number and size distribution of trees across an entire landscape. In combination with other remote sensing data sources, this could allow for improved carbon stock and dynamics estimation. Estimating carbon stocks in forests has traditionally been done using area-based methods which discard considerable granular information at the individual tree level (Coomes et al., 2017).

We focused on aerial RGB imagery which is the cheapest and most widely available imaging source for tropical forests. We also benefited from the variety of pre-trained models that come with this data type. However, different data sources may provide additional information that would help to discern differences between crowns. In particular, multi-spectral imagery that typically includes additional bands in the near-infrared is commonly used to study differences between trees due to the optical properties of vegetation (Knipling, 1970). Alternatively, the canopy surface (a raster expressing the height of the canopy) is commonly used in traditional segmentation techniques (e.g. watershed algorithms) and is produced with photogrammetry as a step in generating an orthomosaic. Including this as a layer would add an additional dimension of information that could help to distinguish fine differences in structure. It would be straightforward to include additional (or different) bands to the Detectree2 framework but it would forego the utility of the pre-trained models. Therefore, it is likely that significantly more training data and computational resources would be required to train a model (from scratch) to the desired performance.

Ideally, we could apply this approach to satellite imagery to perform global analyses. Preliminary tests suggest that Detectree2 can accurately delineate trees in RGB imagery at 2 m resolution (Supplementary Note 7) which is equivalent to modern high-resolution satellite imagery. If this proves possible it will help answer many long-standing questions in forest ecology as well as provide an important tool for forest management. A 'random resizing' augmentation step would further help improve generalizability across resolutions and incorporating 'small object' detection features (Tong & Yiquan, 2022) would improve the sensitivity to shorter trees.

While we studied the delineation of a single class (*tree*), Detectree2 can be trained and make predictions on multiple classes. This may allow for low-cost species identification and mapping. It may also help to automatically

assess liana infestation occurrence. Previously, hyperspectral data have been employed to address this problem but with limited success due to the phylogenetic and spectral diversity of lianas and relatively low spatial resolution of hyperspectral imagery (Grabska et al., 2020; Wessel et al., 2018). The availablility of Detectree2 as an open-source Python package means other research groups can test its efficacy on their own research questions.

Acknowledgements

We thank the NERC Earth Observation Data Acquisition and Analysis Service (NEODAAS) for supplying data and computational resources for this study. We thank CEDA for the 2014 Sepilok and Danum remote sensing data. S. H. M. H. received funding from the Centre for Doctoral Training in Application of Artificial Intelligence to the study of Environmental Risks (AI4ER, EP/S022961/1), which is supported by the Engineering and Physical Sciences Research Council (EPSRC). J. G. C. B. was supported by the NERC C-CLEAR doctoral training programme (PDAG/501). T. D. J. and D. A. C. were supported by NERC grant (NE/S010750/1). D. A. C. was supported by the Franklinia Foundation. We thank Stephen Goult for contributing to discussions. Data collection in French Guiana was supported by CNES who funded the 2016 hyperspectral, RGB and lidar data over Paracou and Labex CEBA (ANR-10-LABX-25) for contributing financial resource for the field validation of manual crown segmentations. The 2019 data in Paracou and 2020 data in Sabah were funded by NERC (NE/S010750/1). The 2014 Sabah data were also funded by NERC (NE/K016377/1).

Data Availability Statement

Airborne LiDAR and RGB imagery collected in Malaysia in 2020 are available here: https://doi.org/10.5285/dd4d20c8626f4b9d99bc14358b1b50fe. Crown data for Danum and Sepilok are available here https://doi.org/10.5281/zenodo.7843154

Endnotes

1 https://github.com/PatBall1/Detectree2.
2 https://github.com/facebookresearch/detectron2/blob/main/MODEL_ZOO.md.
3 https://github.com/facebookresearch/detectron2/blob/main/configs/COCO-Detection/faster_rcnn_R_101_FPN_3x.yaml.
4 https://wandb.ai/detectree/tune/sweeps/.
5 See https://github.com/PatBall1/Detectree2.

References

AgiSoft. (2021) *Agisoft metashape (version 1.7.3)*.

Almeida, A., Gonçalves, F., Silva, G., Mendonça, A., Gonzaga, M., Silva, J. et al. (2021) Individual tree detection and qualitative inventory of a *Eucalyptus* sp. stand using UAV photogrammetry data. *Remote Sensing*, **13**(18), 3655. Available from: https://doi.org/10.3390/rs13183655

Aparecido, A., dos Santos, J., Junior, M., Araújo, M.S., Di Martini, D.R., Tetila, E.C. et al. (2019) Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVs. *Sensors*, **19** (16), 3595.

Araujo, R.F., Grubinger, S., Celes, C.H.S., Garcia, M., Dandois, J.P. & Muller-Landau, H.C. (2021) Strong temporal variation in treefall and branchfall rates in a tropical forest is related to extreme rainfall: results from 5 years of monthly drone data for a 50 ha plot. *Biogeosciences*, **18**(24), 6517–6531.

Aubry-Kientz, M., Dutrieux, R., Ferraz, A., Saatchi, S., Hamraz, H., Williams, J. et al. (2019) A comparative assessment of the performance of individual tree crowns delineation algorithms from ALS data in tropical forests. *Remote Sensing*, **11**(9), 1086. Available from: https://doi.org/10.3390/rs11091086

Aubry-Kientz, M., Laybros, A., Weinstein, B., Ball, J., Jackson, T., Coomes, D. et al. (2021) Multisensor data fusion for improved segmentation of individual tree crowns in dense tropical forests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **14**, 3927–3936. Available from: https://doi.org/10.1109/JSTARS.2021.3069159

Audibert, J.-Y. & Catoni, O. (2011) Robust linear least squares regression. *The Annals of Statistics*, **39**(5), 2766–2794.

Braga, J.R.G., Peripato, V., Dalagnol, R., Ferreira, M.P., Tarabalka, Y. & Aragão, L.E.O.C. (2020) Tree crown delineation algorithm based on a convolutional neural network. *Remote Sensing*, **12**(8), 1288.

Brienen, R.J.W., Phillips, O.L., Feldpausch, T.R., Gloor, E., Baker, T.R., Lloyd, J. et al. (2015) Long-term decline of the Amazon carbon sink. *Nature*, **519**, 344–348. Available from: https://doi.org/10.1038/nature14283

Chai, Y., Martins, G., Nobre, C., von Randow, C., Chen, T. & Dolman, H. (2021) Constraining Amazonian land surface temperature sensitivity to precipitation and the probability of forest dieback. *NPJ Climate and Atmospheric Science*, **4**, 6. Available from: https://doi.org/10.1038/s41612-021-00162-1

Chave, J., Davies, S.J., Phillips, O.L., Lewis, S.L., Sist, P., Schepaschenko, D. et al. (2019) Ground data are essential for biomass remote sensing missions. *Surveys in Geophysics*, **40**, 863–880. Available from: https://doi.org/10.1007/s10712-019-09528-w

Clinton, N., Holt, A., Scarborough, J., Yan, L. & Gong, P. (2010) Accuracy assessment measures for object-based image segmentation goodness. *Photogrammetric Engineering & Remote Sensing*, **76**(3), 289–299.

Coomes, D.A., Dalponte, M., Jucker, T., Asner, G.P., Banin, L.F., Burslem, D.F.R.P. et al. (2017) Area-based vs tree-

centric approaches to mapping forest carbon in southeast Asian forests from airborne laser scanning data. *Remote Sensing of Environment*, **194**, 77–88.
Coomes, D.A., Duncan, R.P., Allen, R.B. & Truscott, J. (2003) Disturbances prevent stem size-density distributions in natural forests from following scaling relationships. *Ecology Letters*, **6**(11), 980–989.
Cushman, K.C., Detto, M., Garca, M. & Muller-Landau, H.C. (2022) Soils and topography control natural disturbance rates and thereby forest structure in a lowland tropical landscape. *Ecology Letters*, **25**, 1126–1138.
Dalponte, M. & Coomes, D.A. (2016) Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. *Methods in Ecology and Evolution*, **7**(10), 1236–1245. Available from: https://doi.org/10.1111/2041-210X.12575
Davies, S.J., Abiem, I., Salim, K.A., Aguilar, S., Allen, D., Alonso, A. et al. (2021) Forestgeo: understanding forest diversity and dynamics through a global observatory network. *Biological Conservation*, **253**, 108907. Available from: https://doi.org/10.1016/j.biocon.2020.108907
Fisher, R.A., Koven, C.D., Anderegg, W.R.L., Christoffersen, B.O., Dietze, M.C., Farrior, C.E. et al. (2018) Vegetation demographics in Earth system models: a review of progress and priorities. *Global Change Biology*, **24**(1), 35–54. Available from: https://doi.org/10.1111/gcb.13910
ForestPlots.net, Blundo, C., Carilla, J., Grau, R., Malizia, A., Malizia, L. et al. (2021) Taking the pulse of Earth's tropical forests using networks of highly distributed plots. *Biological Conservation*, **260**, 108849. Available from: https://doi.org/10.1016/j.biocon.2020.108849
Gallup, S.M., Baker, I.T., Gallup, J.L., Restrepo-Coupe, N., Haynes, K.D., Geyer, N.M. et al. (2021) Accurate simulation of both sensitivity and variability for Amazonian photosynthesis: is it too much to ask? *Journal of Advances in Modeling Earth Systems*, **13**(8), e2021MS002555. Available from: https://doi.org/10.1029/2021MS002555
Gensheng, H., Wang, T., Wan, M., Bao, W. & Zeng, W. (2022) UAV remote sensing monitoring of pine forest diseases based on improved Mask R-CNN. *International Journal of Remote Sensing*, **43**(4), 1274–1305. Available from: https://doi.org/10.1080/01431161.2022.2032455
Glorot, X. & Bengio, Y. (2010) Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. JMLR workshop and conference proceedings. pp. 249–256.
Gora, E.M. & Esquivel-Muelbert, A. (2021) Implications of size-dependent tree mortality for tropical forest carbon dynamics. *Nature Plants*, **7**, 384–391. Available from: https://doi.org/10.1038/s41477-021-00879-0
Gourlet-Fleury, S., Guehl, J.-M. & Laroussinie, O. (2004) *Ecology and management of a neotropical rainforest: lessons drawn from Paracou, a long-term experimental research site in French Guiana*. Paris: Elsevier.
Grabska, E., Frantz, D. & Ostapowicz, K. (2020) Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the polish carpathians. *Remote Sensing of Environment*, **251**, 112103.
Hao, Z., Lin, L., Post, C.J., Mikhailova, E.A., Li, M., Chen, Y. et al. (2021) Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (Mask R-CNN). *ISPRS Journal of Photogrammetry and Remote Sensing*, **178**, 112–123.
Hastings, J.H., Ollinger, S.V., Ouimette, A.P., Sanders-DeMott, R., Palace, M.W., Ducey, M.J. et al. (2020) Tree species traits determine the success of lidar-based crown mapping in a mixed temperate forest. *Remote Sensing*, **12**(2), 309. Available from: https://doi.org/10.3390/rs12020309
He, K., Gkioxari, G., Dollár, P. & Girshick, R. (2017) Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy. pp. 2961–2969.
He, K., Zhang, X., Ren, S. & Sun, J. (2016) Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 770–778.
Hubau, W., Lewis, S.L., Phillips, O.L., Affum-Baffoe, K., Beeckman, H., Cuní-Sanchez, A. et al. (2020) Asynchronous carbon sink saturation in African and Amazonian tropical forests. *Nature*, **579**, 80–87. Available from: https://doi.org/10.1038/s41586-020-2035-0
Huertas, C., Sabatier, D., Derroire, G., Ferry, B., Jackson, T.D., Pélissier, R. et al. (2022) Mapping tree mortality rate in a tropical moist forest using multi-temporal lidar. *International Journal of Applied Earth Observation and Geoinformation*, **109**, 102780. Available from: https://doi.org/10.1016/j.jag.2022.102780
Hurst, J.M., Allen, R.B., Coomes, D.A. & Duncan, R.P. (2011) Size-specific tree mortality varies with neighbourhood crowding and disturbance in a Montane Nothofagus forest. *PLoS One*, **6**(10), e26670.
Iglhaut, J., Cabo, C., Puliti, S., Piermattei, L., O'Connor, J., Rosette, J. et al. (2019) Structure from motion photogrammetry in forestry: a review. *Current Forestry Reports*, **5**, 155–168. Available from: https://doi.org/10.1007/s40725-019-00094-3
Iida, Y., Poorter, L., Sterck, F., Kassim, A.R., Potts, M.D., Kubo, T. et al. (2014) Linking size-dependent growth and mortality with architectural traits across 145 co-occurring tropical tree species. *Ecology*, **95**(2), 353–363.
IPCC. (2021) *Climate change 2021: the physical science basis*. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press.
Kattenborn, T., Leitloff, J., Schiefer, F. & Hinz, S. (2021) Review on convolutional neural networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry*

and Remote Sensing, **173**, 24–49. Available from: https://doi.org/10.1016/j.isprsjprs.2020.12.010

Kattenborn, T., Schiefer, F., Frey, J., Feilhauer, H., Mahecha, M.D. & Dormann, C.F. (2022) Spatially autocorrelated training and validation samples inflate performance assessment of convolutional neural networks. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, **5**, 100018. Available from: https://doi.org/10.1016/j.ophoto.2022.100018

Kellner, J.R., Albert, L.P., Burley, J.T. & Cushman, K.C. (2019) The case for remote sensing of individual plants. *American Journal of Botany*, **106**(9), 1139–1142. Available from: https://doi.org/10.1002/ajb2.1347

Knipling, E.B. (1970) Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment*, **1**(3), 155–159. Available from: https://doi.org/10.1016/S0034-4257(70)80021-9

Koven, C.D., Knox, R.G., Fisher, R.A., Chambers, J.Q., Christoffersen, B.O., Davies, S.J. et al. (2020) Benchmarking and parameter sensitivity of physiological and vegetation dynamics using the functionally assembled terrestrial ecosystem simulator (fates) at Barro Colorado Island, Panama. *Biogeosciences*, **17**(11), 3017–3044.

Kunyong, Y., Hao, Z., Post, C.J., Mikhailova, E.A., Lin, L., Zhao, G. et al. (2022) Comparison of classical methods and Mask R-CNN for automatic tree detection and mapping using UAV imagery. *Remote Sensing*, **14**(2), 295. Available from: https://doi.org/10.3390/rs14020295

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B. & Belongie, S. (2017) Feature pyramid networks for object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA. pp. 2117–2125.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J. et al. (2014) Microsoft COCO: common objects in context. In: *European conference on computer vision*. Zurich: Springer, pp. 740–755.

Lutz, J.A., Furniss, T.J., Johnson, D.J., Davies, S.J., Allen, D., Alonso, A. et al. (2018) Global importance of large-diameter trees. *Global Ecology and Biogeography*, **27**(7), 849–864.

Marvin, D.C., Asner, G.P., Knapp, D.E., Anderson, C.B., Martin, R.E., Sinca, F. et al. (2014) Amazonian landscapes and the bias in field studies of forest structure and biomass. *Proceedings of the National Academy of Sciences of the United States of America*, **111**(48), E5224–E5232. Available from: https://doi.org/10.1073/pnas.1412999111

Meakem, V., Tepley, A.J., Gonzalez-Akre, E.B., Herrmann, V., Muller-Landau, H.C., Wright, S.J. et al. (2018) Role of tree size in moist tropical forest carbon cycling and water deficit responses. *New Phytologist*, **219**(3), 947–958. Available from: https://doi.org/10.1111/nph.14633

Mugabowindekwe, M., Brandt, M., Chave, J., Reiner, F., Skole, D.L., Kariryaa, A. et al. (2022) Nation-wide mapping of tree-level aboveground carbon stocks in Rwanda. *Nature Climate Change*, **13**, 91–97. Available from: https://doi.org/10.1038/s41558-022-01544-w

Muller-Landau, H.C., Condit, R.S., Chave, J., Chave, J., Thomas, S.C., Bohlman, S.A. et al. (2006) Testing metabolic ecology theory for allometric scaling of tree size, growth and mortality in tropical forests. *Ecology Letters*, **9**(5), 575–588.

Nilus, R., Maycock, C.R., Majalap-Lee, N. & Burslem, D.F.R.P. (2011) Nutrient limitation of tree seedling growth in three soil types found in Sabah. *Journal of Tropical Forest Science*, **23**, 133–142.

Ocer, N.E., Kaplan, G., Erdem, F., Kucuk Matci, D. & Avdan, U. (2020) Tree extraction from multi-scale UAV images using Mask R-CNN with FPN. *Remote Sensing Letters*, **11**(9), 847–856. Available from: https://doi.org/10.1080/2150704X.2020.1784491

Ordway, E.M., Asner, G.P., Burslem, D.F.R.P., Lewis, S.L., Nilus, R., Martin, R.E. et al. (2022) Mapping tropical forest functional variation at satellite remote sensing resolutions depends on key traits. *Communications Earth & Environment*, **3**(1), 247.

Pan, Y., Birdsey, R.A., Fang, J., Houghton, R., Kauppi, P.E., Kurz, W.A. et al. (2011) A large and persistent carbon sink in the world's forests. *Science*, **333**(6045), 988–993. Available from: https://doi.org/10.1126/science.1201609

Piponiot, C., Anderson-Teixeira, K.J., Davies, S.J., Allen, D., Bourg, N.A., Burslem, D.F.R.P. et al. (2022) Distribution of biomass dynamics in relation to tree size in forests across the world. *New Phytologist*, **234**(5), 1664–1677. Available from: https://doi.org/10.1111/nph.17995

Restrepo-Coupe, N., Albert, L.P., Longo, M., Baker, I., Levine, N.M., Mercado, L.M. et al. (2021) Understanding water and energy fluxes in the Amazonia: lessons from an observation-model intercomparison. *Global Change Biology*, **27**(9), 1802–1819. Available from: https://doi.org/10.1111/gcb.15555

Richardson, S.J., Smale, M.C., Hurst, J.M., Fitzgerald, N.B., Peltzer, D.A., Allen, R.B. et al. (2009) Large-tree growth and mortality rates in forests of the central North Island, New Zealand. *New Zealand Journal of Ecology*, **33**(2), 208–215.

Roussel, J.-R., Caspersen, J., Béland, M., Thomas, S. & Achim, A. (2017) Removing bias from lidar-based estimates of canopy height: accounting for the effects of pulse density and footprint size. *Remote Sensing of Environment*, **198**, 1–16.

Shenkin, A., Chandler, C.J., Boyd, D.S., Jackson, T., Disney, M., Majalap, N. et al. (2019) The world's tallest tropical tree in three dimensions. *Frontiers in Forests and Global Change*, **2**, 32.

Stovall, A.E.L., Shugart, H. & Yang, X. (2019) Tree height explains mortality risk during an intense drought. *Nature Communications*, **10**, 4385. Available from: https://doi.org/10.1038/s41467-019-12380-6

Tong, K. & Yiquan, W. (2022) Deep learning-based detection from the perspective of small or tiny objects: a survey. *Image and Vision Computing*, **123**, 104471. Available from: https://doi.org/10.1016/j.imavis.2022.104471

Wagner, F., Hérault, B., Stahl, C., Rossi, V., Bonal, D. & Rossi, V. (2011) Modeling water availability for trees in tropical forests. *Agricultural and Forest Meteorology*, **151**(9), 1202–

1213. Available from: https://doi.org/10.1016/j.agrformet.2011.04.012
Wagner, F., Rutishauser, E., Blanc, L. & Herault, B. (2010) Effects of plot size and census interval on descriptors of forest structure and dynamics. *Biotropica*, **42**(6), 664–671. Available from: https://doi.org/10.1111/j.1744-7429.2010.00644.x
Weinstein, B.G., Marconi, S., Bohlman, S.A., Zare, A., Singh, A., Graves, S.J. et al. (2021) A remote sensing derived data set of 100 million individual tree crowns for the national ecological observatory network. *eLife*, **10**, e62922.
Weinstein, B.G., Marconi, S., Bohlman, S., Zare, A. & White, E.P. (2019) Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. *Remote Sensing*, **11**(11), 1309.
Weiss, K., Khoshgoftaar, T.M. & Wang, D.D. (2016) A survey of transfer learning. *Journal of Big Data*, **3**(1), 1–40.
Wessel, M., Brandmeier, M. & Tiede, D. (2018) Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on Sentinel-2 data. *Remote Sensing*, **10**(9), 1419.
Westoby, M.J., Brasington, J., Glasser, N.F., Hambrey, M.J. & Reynolds, J.M. (2012) Structure-from-motion'photogrammetry: a low-cost, effective tool for geoscience applications. *Geomorphology*, **179**, 300–314.
Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y. & Girshick, R. (2019) *Detectron2*. Available from: https://github.com/facebookresearch/detectron2 [Accessed 18th April 2023].
Zhen, Z., Quackenbush, L.J. & Zhang, L. (2016) Trends in automatic individual tree crown detection and delineation evolution of lidar data. *Remote Sensing*, **8**(4), 333. Available from: https://doi.org/10.3390/rs8040333
Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F. et al. (2017) Deep learning in remote sensing: a comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, **5**(4), 8–36. Available from: https://doi.org/10.1109/MGRS.2017.2762307
Zuidema, P.A. & van der Sleen, P. (2022) Seeing the forest through the trees: how tree-level measurements can help understand forest dynamics. *New Phytologist*, **234**(5), 1544–1546. Available from: https://doi.org/10.1111/nph.18144

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. An original RGB image is displayed on the left, while a stretched RGB image is displayed on the right of this figure. Stretching the colours of the image to the values allows for easier identification of individual trees when carrying out manual tree crown delineations. The effect is particularly noticeable in the lower left corner of the images.
Figure S2. Illustration of the Mask R-CNN predictions and architecture from He et al. (2017).
Figure S3. (A) Crown bounding boxes predicted by DeepForest (Weinstein et al., 2019), and (B) crowns predicted by Detectree2. The colours in plot B merely distinguish predicted trees. A comparison of manually delineated crowns, overlaid on lidar (C) and RGB (D).
Figure S4. Examples of training data provided to Mask R-CNN. The different colours help to distinguish between trees.
Figure S5. These plots are taken from He et al. (2016) and they illustrate that deeper neural networks do not necessarily learn as well as shallower neural networks.
Figure S6. The total training and validation loss of Mask R-CNN as the model trained. Both total training and validation loss were calculated every 20 iterations.
Figure S7. Example delineation results at Danum.
Figure S8. Example delineation results at Sepilok West.
Figure S9. Example delineation results at Sepilok West.
Figure S10. Example delineation results at Paracou.
Figure S11. The sensitivity of the accuracy of the segmentations to the resolution of images used in training and testing.
Figure S12. Shows the robust least squares fit for change in height and tree height for Sabah (Danum, Sepilok West and Sepilok East). The dashed lines indicate three standard deviations either side of the best fit and red points below the lower bound indicate likely mortality events.
Table S1. Tunable hyperparameters (with their optimised value) and a description of their purpose.
Table S2. The coefficients and intercepts for the robust least squares fit between original tree height and the change in tree height.
Supplementary Note 1. Study sites and remote sensing data collection.
Supplementary Note 2. Tree crown data.
Supplementary Note 3. Data preparation and processing.
Supplementary Note 4. Model architecture, tuning and training.
Supplementary Note 5. Evaluation metrics.
Supplementary Note 6. Maps of predictions.
Supplementary Note 7. Sensitivity to image resolution.
Supplementary Note 8. Growth and mortality details.
Table S3. Model accuracies and parameters across sites.
Table S4. A comparison of the contribution of over/undersegmentation to the accuracies across sites.
Table S5. The accuracy of predictions by tree height.
Table S6. The estimated growth and mortality by tree height across the sites.