Ecological Informatics 56 (2020) 101061

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Cross-site learning in deep learning RGB tree crown detection

Ben G. Weinstein[a,*], Sergio Marconi[a], Stephanie A. Bohlman[b], Alina Zare[c], Ethan P. White[a]

[a] Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, FL, USA
[b] School of Forest Resources and Conservation, University of Florida, Gainesville, FL, USA
[c] Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

ARTICLE INFO

Keywords:
Tree crown detection
RGB deep learning
Object detection
Airborne LiDAR

ABSTRACT

Tree crown detection is a fundamental task in remote sensing for forestry and ecosystem ecology. While many individual tree segmentation algorithms have been proposed, the development and testing of these algorithms is typically site specific, with few methods evaluated against data from multiple forest types simultaneously. This makes it difficult to determine the generalization of proposed approaches, and limits tree detection at broad scales. Using data from the National Ecological Observatory Network, we extend a recently developed deep learning approach to include data from a range of forest types to determine whether information from one forest can be used for tree detection in other forests, and explore the potential for building a universal tree detection algorithm. We find that the deep learning approach works well for overstory tree detection across forest conditions. Performance was best in open oak woodlands and worst in alpine forests. When models were fit to one forest type and used to predict another, performance generally decreased, with better performance when forests were more similar in structure. However, when models were pretrained on data from other sites and then fine-tuned using a relatively small amount of hand-labeled data from the evaluation site, they performed similarly to local site models. Most importantly, a model fit to data from all sites performed as well or better than individual models trained for each local site.

1. Introduction

Tree detection is a critical step in remote sensing of forested landscapes. Identifying individual crowns in airborne imagery allows ecologists, foresters, and land managers to increase the extent of sampling compared to terrestrial surveys. While many LIDAR-based tree segmentation algorithms have been proposed (Aubry-Kientz et al., 2019), the field has been slow to adopt automated methods due to concerns over accuracy, transferability and transparency (Vaglio Laurin et al., 2019). As a result, existing methods are rarely evaluated on multiple forests simultaneously, making it unclear how they will perform in the novel contexts required for large scale application. This is critical since the vast majority of future users will want to take methods designed for one site and apply them to a broad range of geographic areas. These areas are likely to include diverse forest characteristics, such as differences in crown structure, species diversity, canopy openness. A lack of knowledge about the flexibility of algorithms to new forest conditions is a major impediment to widespread adoption.

Current tree segmentation approaches are primarily based on user-defined algorithms that describe the appearance of trees in a hierarchical sequence of rules. These rule-based approaches rely on combinations of shape features (Gomes et al., 2018), template matching (Dai et al., 2018), network analysis (Williams et al., 2019), and watershed routines (Silva et al., 2016) that are applied to either LIDAR point clouds or RGB photogrammetric imagery (Brieger et al., 2019). By describing the parameters that define an individual tree, these algorithms attempt to match these rules when predicting unlabeled data. The algorithms are largely unsupervised, as they often contain just a handful of important parameters for tuning to local data. These parameters can often have large effects on results. The combination of few parameters of large effect makes it challenging to apply these algorithms to large areas, and often leads to overfitting during visual calibration to a test dataset. For example, some methods use allometric relationships between crown area and tree height to improve algorithm performance (Coomes et al., 2017; Williams et al., 2019), but these relationships vary with forest type and species. Recent attempts to mitigate this variation have used approaches that choose from a pool of potential tree shapes (Gomes et al., 2018). However, the need to define the full pool of possible tree shapes before analyzing each new site will be prohibitive over large geographic areas that incorporate diverse assemblages. As a result of these limitations, most tree detection algorithms have been applied and tested on similar forest types with little

* Corresponding author.
E-mail address: ben.weinstein@weecology.org (B.G. Weinstein).

https://doi.org/10.1016/j.ecoinf.2020.101061
Received 11 December 2019; Received in revised form 24 January 2020; Accepted 27 January 2020
Available online 30 January 2020

exploration of how the algorithms generalize to other natural settings. Therefore, despite the intense work in airborne tree detection over the last decade (Coomes et al., 2017; Heinzel and Huber, 2018; Jakubowski et al., 2013; Li et al., 2012; Williams et al., 2019), there remains no clear consensus on best practices (Aubry-Kientz et al., 2019).

Within the field of computer vision, there has been a broad shift away from user-designed features toward approaches that learn features from data using neural networks (Agarwal et al., 2018). There have been few attempts to use learned features in tree detection (Dai et al., 2018) due to the need for large amounts of labeled training data, which is often difficult or impossible to collect in ecological contexts. Overall, generalization of deep learning algorithms across applications in airborne remote sensing remains a challenging task (Zhu et al., 2017). A typical neural network has millions of parameters and is therefore at risk of overfitting when using the small datasets usually available for training. Given the diversity of trees, finding general features will require a combination of large training datasets and algorithmic approaches that allow the neural networks to learn the combination of features that characterize trees across forest types.

Weinstein et al. (2019) recently developed a deep learning approach for tree detection using RGB (red-green-blue) data, which was applied at one site, but has the potential to address these requirements for identifying trees across forest types (Fig. 1). The method uses unsupervised LiDAR-based tree detection (e.g., Silva et al., 2016) to generate millions of labeled trees. These low to moderate quality annotations are used to pretrain the neural network. This pretraining stage is followed by retraining based on a small number of high-quality hand-annotations. Whether this method can learn general features across forest types is unknown. This challenge represents an important step from demonstrating the utility of the algorithm as a proof-of-concept to creating a flexible approach that can adapt to the tremendous geographic diversity in tree shapes, appearance and landscape structure. More generally, this problem reflects the ongoing synthesis between data science and ecology. Ecological imagery is often more complex than other computer vision applications due to variability in image acquisition environments and extreme intra-class variation. Finding solutions in ecological machine learning that perform well at large geographic scales is a key factor in delivering reliable results.

Here we explore the potential of this tree detection method to generalize across sites by evaluating its performance on a range of forest types, assessing the transferability of tree features across forest types, and exploring the possibility of building a single unified tree detection model. Our aim is to test a deep learning approach 1) for identifying trees in four different forest types when trained on that forest type ('within-site'); 2) for identifying trees when trained on data from other forest types ('cross-site'); 3) for combining pretraining data from other sites with hand-annotated data from a new site ('transfer learning'); and 4) for comparing the performance of a within-site model to a universal model fit to data on all forest types simultaneously ('universal'). By universal we mean training single model with data from all geographic locations. We also explore the sensitivity of the approach to the number of hand annotations, to determine the amount of time-intensive work needed to produce accurate results. By answering these questions, we will improve our understanding of the potential for universal tree detection methods and potentially advance RGB-based tree detection from algorithm development to large scale application for better understanding forests at scale.

2. Methods

Data collection and site descriptions

The aerial remote sensing data products were provided by the National Ecology Observation Network (NEON) Airborne Observation Platform. We used the NEON 2018 "classified LiDAR point cloud" data product (NEON ID: DP1.30003.001) and the "orthorectified camera mosaic" (NEON ID: DP1.30010.001). The LiDAR data consist of 3D spatial point coordinates with an average of 4–6 points/m^2. These data provide high resolution information about crown shape and height. The RGB data are a 1 km × 1 km mosaic of individual images with a cell size of 0.1 m. All data are publicly available on the NEON Data Portal (http://data.neonscience.org/). For hand-annotations, we selected two 1 km × 1 km RGB tiles and used the program RectLabel (https://rectlabel.com/) to draw bounding boxes around each visible tree. For a count of tree annotations per site, see Table 1. All code for this project is available on GitHub (https://github.com/weecology/DeepLiDAR) and archived on Zenodo, and all annotations are available as part of the forthcoming NEON Tree Benchmark (https://github.com/weecology/NeonTreeEvaluation).

We selected four sites from the NEON network to capture a range of crown shapes, canopy complexity and forest types. The 'Oak Woodland' is the San Joaquin Experimental Range, California. The site contains live oak (*Quercus agrifolia*), blue oak (*Quercus douglasii*) and foothill pine (*Pinus sabiniana*) forest. The majority of the site is relatively open, has a single-story canopy, rounded crowns with mixed understory of herbaceous vegetation. The "Mixed Pine" site is Lower Teakettle, California (37.00583, −119.00602) which contains red fir (*Abies magnifica*) and white fir (*Abies concolor*), Jeffrey pine (*Pinus jeffreyi*) and lodgepole Pine (*Pinus contorta*). This site has a closed canopy of conically shaped conifers crowns. The "Alpine" site is Niwot Ridge Mountain Research Station, Colorado (40.05425, −105.58237). This high elevation site (3000 m) is near treeline with clusters of subalpine fir (*Abies lasciocarpa*) and Englemann spruce (*Picea engelmanii*). This site is very open with small, conically shaped crowns often occurring in tight clumps. Finally, the "Eastern Deciduous" site is the Mountain Lake Biological Station, Virginia (37.37828, −80.52484). Here the dense canopy is dominated by red maple (*Acer rubrum*) and white oak (*Quercus alba*). The canopy is closed with rounded to flat-topped abutting crowns and often a developed understory. Each site presents its own challenges, with broad flat-topped trees in the Oak Woodland, tight clusters of trees in the Mixed Pine forest, thin conifers in the

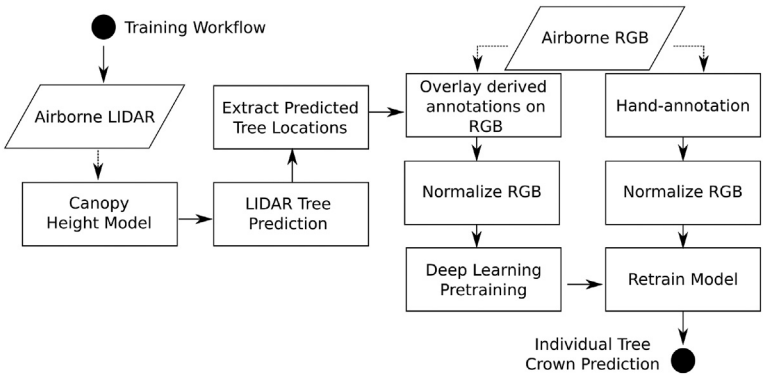

Fig. 1. Conceptual workflow of proposed approach for airborne detection of individual tree crowns. Pretraining data is generated by overlaying predicted trees from a LiDAR-based unsupervised algorithm on to RGB imagery. These RGB images are used to pretrain a deep learning neural network. The resulting model is retrained based on RGB hand-annotations.

Table 1
The number of tree annotations used for pretraining, retraining and evaluation. Pretraining annotations are generated automatically using a LiDAR-based unsupervised algorithm. Training and evaluation annotations were hand-drawn.

Forest type	Pretraining annotations	Training annotations	Evaluation annotations
Oak Woodland	550,905	2533	293
Mixed Pine	2,522,855	3405	747
Alpine	3,121,036	9730	1699
Eastern Deciduous	3,131,283	1231	489

Alpine forest, and completely connected crowns in the Eastern Deciduous forest.

For each site, we manually annotated training tiles using the program RectLabel (Table 1). Training tiles were selected at random from the NEON data portal. At higher tree density sites, we cropped the $1km^2$ tiles to create more tractable sizes for hand-annotation. To enforce a minimum size threshold for tree annotations, we compared the hand-annotations to a LiDAR canopy height model and removed any trees less than 3 m in height. The resulting annotations were compared to the LiDAR point cloud for further assessment. No attempt was made to delineate understory trees that were not visible in the RGB imagery. Since these annotations were made using only remotely sensed imagery, there is some uncertainty in the bounding box extents. While methods exist for combining imagery and field-collected labels (e.g. Aubry-Kientz et al., 2019; Graves et al., 2018), these are difficult to implement at large scales. Associated uncertainty should be considered when interpreting our results and future efforts to quantify label uncertainty produced by both field and remote-sensing based methods is an important research direction.

For model evaluation, we used the NEON "tower" plots, which are a set of 40x40m plots placed throughout each site. For the Eastern Deciduous site, it was difficult to determine tree boundaries in both the RGB and LiDAR images. For this site, we overlaid a 1 m resolution three-band hyperspectral composite image to highlight differences among co-occurring tree species in the area. The composite image came from NEON's orthorectified surface reflectance (ID: DP1.30006.001) and contained bands in the infrared (940 nm), red (650 nm), and blue (430 nm) spectrum. This allowed us to more accurately annotate the training and evaluation data in closed canopy conditions.

2.1. LiDAR tree detection

We tested three existing unsupervised LiDAR algorithms (Dalponte and Coomes, 2016; Li et al., 2012; Silva et al., 2016), as implemented in the lidR R package (Roussel and Auty, 2019), as both a comparison to the deep learning approach, and as potential algorithms to generate tree labels for model pretraining. We selected the best performing method (Silva et al., 2016) to create initial tree predictions in the LiDAR point cloud. This approach uses a canopy height model and an allometry of tree height to crown width to cluster the LiDAR cloud into individual trees. We used a canopy height model of 0.5 m horizontal resolution to generate local treetops and an allometry of 90% of crown diameter to height for deciduous forests (Oak Woodland and Eastern Deciduous) and 20% of crown diameter to height for the coniferous forests (Mixed Pine and Alpine). These parameters were based on visual testing on algorithm performance. LiDAR algorithms perform segmentation on a per-point basis, so we converted the output to a bounding box that covered the entire set of LiDAR points assigned to each tree to create training data equivalent to the hand-annotated bounding boxes.

2.2. Deep learning

We used our previously developed algorithm for RGB-based tree identification, which was used at a single site (Weinstein et al., 2019). This method uses the Retinanet one-stage object detector (Gaiser et al., 2018) with a Resnet-50 classification backbone, which allows pixel information to be shared at multiple scales, from individual pixels to groups of connected objects. We used a Resnet-50 classification backbone pretrained on the ImageNet dataset (He et al., 2016). Since the entire 1 km RGB tile cannot fit into GPU memory, we cut each tile into 40 m by 40 m windows with an overlap of 5% (n = 729). The 40mx40m window size was adopted since this is the spatial extent of the NEON tower plots. The order of tiles and windows were randomized before training to minimize overfitting among epochs. To reduce potential spatial autocorrelation in tree appearance between evaluation plots and pretraining data, we removed any training tiles within 1 km of an evaluation tile. Using the pool of unsupervised LiDAR-based tree predictions, we pretrained the network with a batch size of 20 on 2 Tesla K80 GPU for 5 epochs. To align these unsupervised classifications with the ImageNet pretraining weights, we normalized the RGB channels by subtracting the ImageNet mean from each channel. We then retrained the network using the hand-annotated data for 40 epochs. For more details of this approach see Weinstein et al. (2019). Data augmentation of random flips and translations was tested and found to have little effect on the final score.

2.3. Model evaluation

Using the evaluation plots, we chose two metrics to assess model performance. For comparison with the existing LiDAR-only implementations, we used precision and recall statistics with a bounding box marked as true positive if it had an intersection-over-union (IoU) of greater than 0.5. Intersection-over-union is the ratio of the area of bounding box overlap to the area of bounding box union between the predicted tree crown and the visually annotated crowns in the evaluation data. For each bounding box prediction, the deep learning model reports a confidence score between 0 and 1. To transform these scores into precision and recall statistics, we need to define a threshold of box scores to accept. As we lower the threshold for acceptance, a greater number of trees will be captured, but at the expense of decreased precision. To highlight this relationship, we showed the performance of the deep learning approach across all bounding box probability thresholds between 0 and 1 with an interval of 0.1. IoU precision and recall are reported separately and do not capture differences in bounding box confidence scores. When comparing the different generalization approaches, it is useful to have a single metric to compare. We used the Average Precision (AP) metric commonly used for object detection tasks in computer vision, which is the area under the precision-recall curve computed at the 11 fixed 0.1 intervals between 0 and 1 (Lin et al., 2017).

2.4. Assessing generalization, transferability, and universal model fit

To assess generalization among sites, we performed three types of experiments that used different combinations for hand-annotations and pretraining data (Fig. 2). The first experiment is to use pretraining and hand-annotated data to predict the evaluation data from the same site ('within-site'). The next setup is to use the pretraining data and hand-annotated from the same site to predict the evaluation data from a different site ('cross-site'). For example, using each of the within-site models, we can test the ability for a model to predict tree conditions in each of the other geographic sites, creating a matrix of cross-site predictions. To assess generalization without local pretraining data, we tested a model training using pretraining data from all other sites, but hand annotations from the same site as the evaluation data ('transfer-learning'). For example, the transfer learning model for Oak Woodland used the hand-annotations from Oak Woodland, but the pretraining data for Alpine, Mixed Pine, and Eastern Deciduous. Finally, to test the potential for a universal model, we tested a model pretrained on all sites, followed by retraining on all hand-annotations. We then

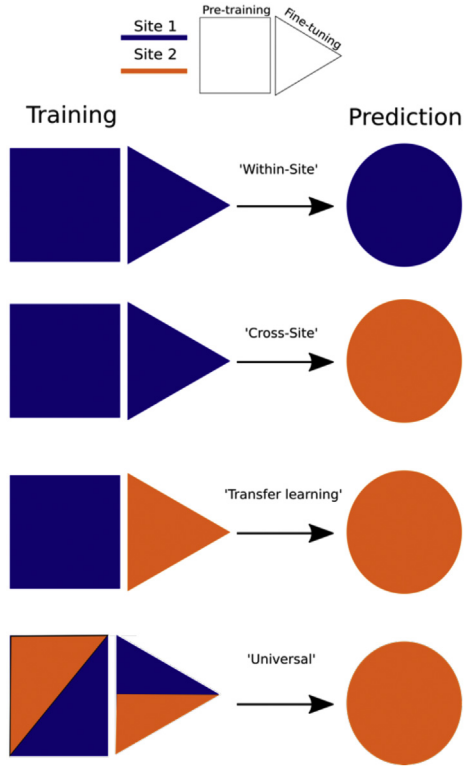

Fig. 2. Approaches to geographic generalization in model training: 1) 'Within-site' training in which training data from site 1 is used to predict site 1; 2) 'Cross-site' training in which training data from site 1 is used to predict site 2; 2) 'Transfer learning' in which a model is first trained on site 1 data, followed by finetuning on site 2 training data, and 3) 'Universal' model in which training data from both site 1 and site 2 are used to predict evaluation data from site 2.

compared this model with each of the within-site model to test whether the addition of data from other sites improved predictions of trees from the same site.

2.5. Sensitivity to the number of hand-annotations

Collecting a sufficient number of training samples will often be a bottleneck in developing supervised methods in airborne imagery. It is therefore useful to test the number of local training samples needed to achieve maximum performance. We performed a sensitivity study by training models using different proportions of training data. We selected 5%, 25%, 50% and 75% of the total hand-annotations to compare to the full dataset for the within-site results for each site. We reran this experiment five times to account for the random subsampling of annotations. In addition, we ran the evaluation plots for the pretraining model only (i.e. 0% hand-annotated data) to assess whether the addition of hand-annotated data improved the within-site pretraining.

3. Results

Within-site predictions ranged from 0.60 recall and 0.75 precision in Mixed Pine to 0.34 recall and 0.55 precision in Alpine. The Oak Woodland and Mixed Pine sites consistently performed better than the Eastern Deciduous and Alpine sites. Visual inspection of the results showed that the majority of false positives were positively identified trees, but whose crown boundaries were either too large or too small for the intersection-over-union score of 0.5. Repeated training runs for each model showed relatively little variance, despite heterogeneity in tree types at all sites (Fig. 3).

When qualitatively comparing model performance with the LiDAR-based algorithms used to generate the pretraining data (Fig. 3), the deep learning model was more successful at delineating boxes around complex crown boundaries and avoiding clumping together small trees with narrow gaps (Fig. 4). For the Oak Woodland site, the deep learning model was better able to capture crown area for the flat-topped canopy and avoided erroneously labeling bushes as trees (defined as woody vegetation > 3 m in height). For the Eastern Deciduous site, the deep learning model more accurately found trees in the closed canopies, despite strong overlap in bounding box predictions and similarity in neighboring tree appearance. To view predictions overlaid on each of the plots for the within-site models, see supplemental dataset S1. Note that the comparison with the LiDAR methods is complicated by differences in data types (RGB versus LiDAR), and the uncertainty in hand-annotations when viewing the RGB image. However, it is important to note that the majority of errors in the LiDAR methods was not in the extent of the bounding boxes, but in joining multiple trees together or splitting trees apart (Fig. 4). Therefore, while we would need significantly more data and analysis to state that a hybrid RGB-based method was superior to LiDAR-only methods, the types of errors made by the LiDAR algorithms cannot be attributed solely to the hand-annotation process.

When applying a model fit at one site to make predictions at other sites, we found generalization of the single-site models to be weak (Fig. 5). Tree stems were often correctly identified among sites with similar forest crown structures (Coniferous versus Deciduous), but the resulting crown boundaries were rarely accurate (Fig. 5 – "Cross-Site"). The one exception was the prediction of Alpine evaluation plots using a model built from the Mixed Pine site. This model outperformed all other cross-site experiments and was superior even to the Alpine within-site model.

Combining local hand-annotated data with unsupervised pretraining data from the other three sites demonstrated good transferability, with performance almost as good as using local pretraining data (Fig. 5). The transfer learning experiments performed better than cross-site predictions for every site. This suggests that the pretraining model allows for generalized features that can be fine-tuned to local conditions.

Fitting a single universal model using data from all sites resulted in the best predictions for every individual site (Fig. 5), except the Alpine site, which was best predicted by the Mixed Pine site. Compared to a model trained at the target site, the average precision of the universal model for the Eastern Deciduous site improved from 0.44 to 0.54 (22%), Mixed Pine from 0.56 to 0.59 (5.4%), Alpine from 0.24 to 0.26 (8.3%) and Oak Woodland from 0.6 to 0.61 (1.6%). Fig. 6 shows a concrete example of the universal model can performance compared to within site models. In Fig. 6B, the within-site model (Mixed Pine) erroneously labels a large boulder in the bottom right hand corner of the image as a tree. This error was made in all other cross-site models, except for Oak Woodland. In the Universal model, this error was not made, suggesting that either the universal model learned information about the background from other sites to improve predictions, or that more data, regardless of locality, led to higher performance.

Assessment of the number of hand-annotations needed to improve model performance indicated that while some hand-annotated data was important at all sites, the number of hand-annotated trees needed to improve model performance was typically relatively small. For example, the recall in the Mixed Pine site was < 0.2 with no hand-annotated data and was over 0.6 using approximately 2000 hand labeled crowns. Only minimal gains in performance occurred using up to an additional 2000 hand-annotated crowns. Overall, the shape of the ablation curves suggest that the model is fairly robust and needs only approximately 1000 crowns in most cases to create a model close to full performance. The exception is the Alpine model, which improved by more than 30% after 3000 crowns. In general, the precision was more robust than recall, suggesting that the hand annotations mostly improve the predictions of crown boundaries rather than additional tree locations.

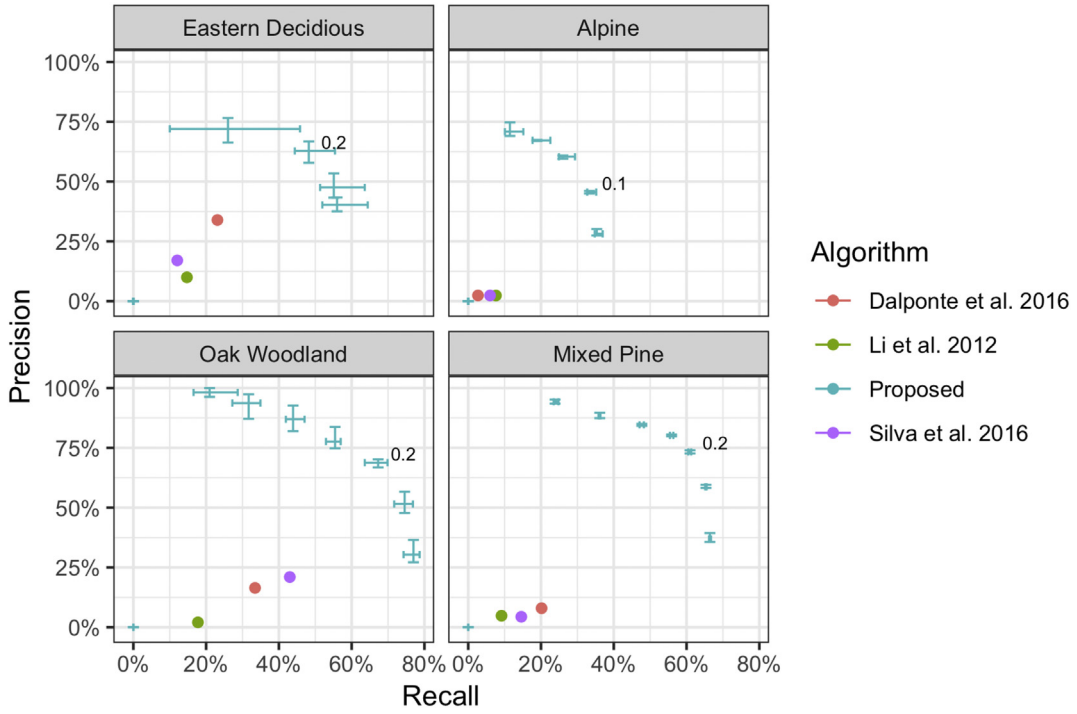

Fig. 3. For each site, results of our proposed workflow for the evaluation data. The proposed model is compared to three existing LiDAR-only implementations from the commonly used lidR package. The proposed approach was evaluated at each of the 0.1 probability score intervals between 0 and 1. The probability threshold of the best performing model in our approach, calculated by f-score, is shown in black. Error bars show the variance in recall and precision based on five runs of hand-annotation training for each probability cutoff at each site.

4. Discussion

Airborne tree detection can unlock ecological and forestry data at unprecedented spatial extents. When combined with traditional ground surveys, these data will inform forest dynamics, ecosystem services, and natural resource management at broad scales. To turn remote sensing data into ecological information, there is a need for a unified tree detection model that can be applied to a broad array of forest conditions. Using a deep learning approach (Weinstein et al., 2019), we trained individual tree detection models for four geographic sites and studied the transferability of learned features among four forest types that encompass a wide range of the variability encountered in temperate forests. Despite large variation in forest structure from open to closed, overlapping canopies, different tree shapes and allometries, and different levels of tree clumping the proposed approach holds promise for automated tree location and size detection at scale. On average across sites, a universal model trained on all sites together correctly identified crown bounding boxes with 65% recall and 70% precision. The remaining false positives were almost always detected as individual trees, but with crown boundaries that did not meet the specified intersection over union threshold of of > 0.5. The success of this algorithm over unsupervised methods highlights the power of supervised approaches for addressing the large geographic diversity of tree forms. Using small amounts of hand-annotated data from the target site, in combination with large amounts of pretraining data from unsupervised LiDAR implementations, is a promising approach for addressing the challenge of tree detection at scale.

One goal was to assess the proposed crown detection approach in variety of canopy conditions to better understand which factors limit performance. We find performance is best in open canopy forests with large, well-spaced, trees as in the Oak Woodland site. We had anticipated the performance of the algorithm would be worst at the closed canopy Eastern Deciduous site. However, it was at the Alpine site that the algorithm had the poorest performance, suggesting that short clusters of trees, rather than complex, interconnected tree boundaries are the biggest challenge among the four forest types. One possible explanation is that the trees in the Alpine site are more sensitive to the resolution of the RGB image due to their small size. Since we use an evaluation metric of intersection-over-union of 0.5, a difference of one pixel is inconsequential for large trees but may push small trees under the threshold for being predicted positively.

One of the advantages of deep learning approaches to tree detection is the potential to learn cross-site features simultaneously. We conducted three types of generalization experiments to assess the transferability among forest types. The first was to use models trained from one site to predict an unseen site. Prediction to unseen conditions is a challenging task in computer vision, especially when the sites were specifically chosen to represent distinct forest types. Overall, we saw a significant decrease in performance between cross-site and within-site models. This means that fitting to a single forest type and applying the model to predict a distinct forest type without local training data remains unlikely to provide acceptable results. The one exception was the prediction of the Alpine site, which had superior performance when predicted by the Mixed Pine site, rather than using the Alpine hand annotations. This may stem from the difficulty of hand annotating the small trees that are common in the Alpine site. It is possible that the model was better at transferring the features from the large conifers in Mixed Pine to the smaller conifers in Alpine than a human was in annotating the crown boundaries in Alpine. A second possibility is that the significant heterogeneity in the pretraining data for the Alpine site led to poor results. The LiDAR-based pretraining algorithm did not perform well at this site, with consistent under-segmentation among small trees. It is possible that the superior quality of the pretraining data at the Mixed Pine site allowed for better predictions in the Alpine site, compared to using lower quality data from the same site. This suggests that improvement in the pretraining algorithm may yield increased performance when combined with hand-annotated data.

To provide the cross-site model with more information on local tree conditions, we conducted transfer learning experiments to assess whether models pretrained at other sites could be used with training data

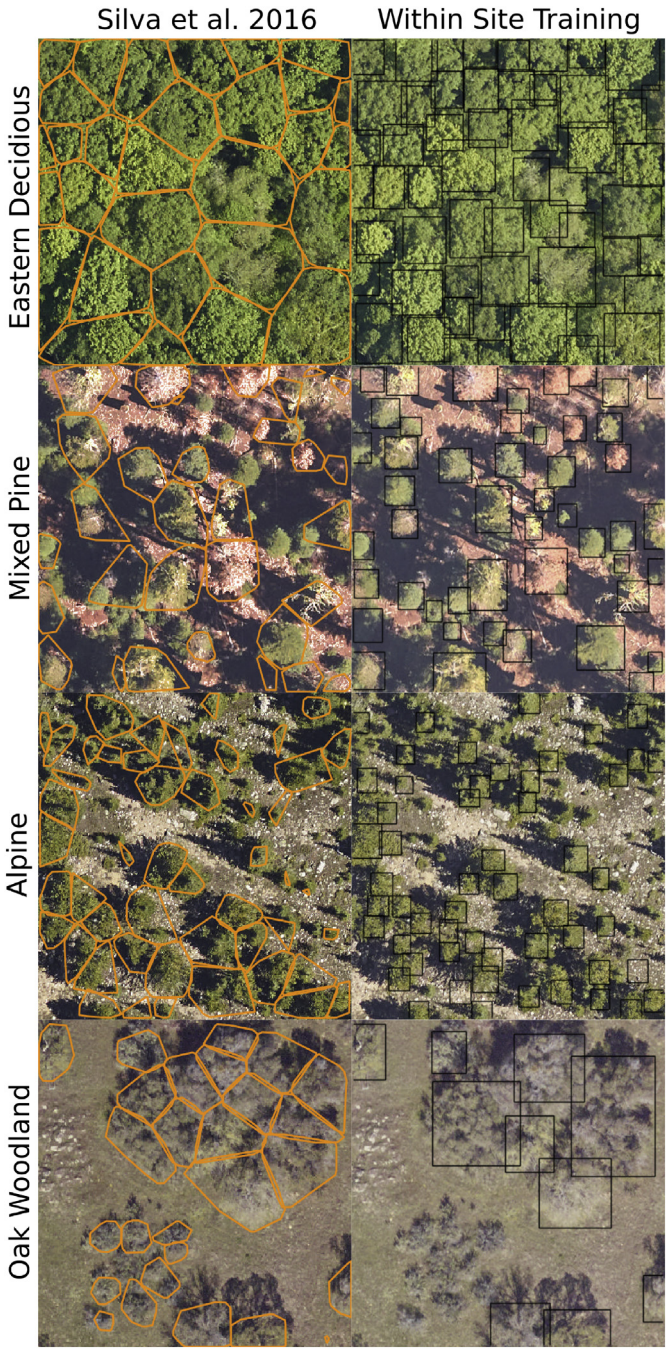

Fig. 4. Example predictions for the LiDAR-only pretraining algorithm, and the deep learning detection network trained within-site.

from a local to site to fine tune the model to that site. We find that building from existing models of tree detection is a promising avenue toward cross-site generalization. Adding only a small amount of local training data (typically < 1000 trees requiring ~3–5 h to label; Fig. 7) greatly increased performance and nearly recovered performance of the within-site model. This opens up the possibility of tree detection models that connect forest types based on their dominant canopy structure and species.

The ultimate goal of the proposed approach is to move toward a single unified model that can produce individual tree predictions in a variety of ecosystems. Our analysis shows promising results for a universal model trained from all pretraining and hand annotations from every site. In all sites, a universal model was equivalent or better than a model train on data from the same site, with improvements of up to

Fig. 5. Comparison single-site, cross-site, transfer, and universal model performance based on Average Precision (AP). Single site predictions are on the bolded diagonal of the cross-site section and represent fitting and predicting on the same site. Cross-site predictions are for models trained on the one site (listed on the left side of the results matrix) and evaluated on a second site (listed across the bottom of the results matrix). Transfer learning takes a model pretrained on all sites except the focal site and retrained using the hand-annotations of the evaluation site. The universal model uses pretraining and hand-annotation data from all sites.

20%. Given that the sites were selected to be as different as possible, and encompass a range of tree canopy conditions, this result underscores the ability of convolutional neural networks to learn flexible deep features. We expect that as more sites are included, the universal model will continue to improve. This means that a way forward is to combine pretraining from as many sites as possible. Given that each NEON site has millions of trees, and there are dozens of sites with trees collected annually, there is a possibility of pretraining on continental scale. Because NEON sites are intended to represent all of the major biogeographic regions in the United States, this broad scale pre-training (in combination with existing local training data) has the potential to reduce the need for new local training data by giving the model the potential to learn the general suite of features characterizing trees (at least those within the United States).

While our results point toward a general strategy for RGB tree detection using LiDAR derived pretraining labels, there are many remaining questions to explore. At broad scales, it is unclear whether hand annotations from every site are needed to generate robust continental scales models, or whether a representative sample of sites, combined with extensive pretraining, will yield adequate results. Furthermore, connecting computer vision measures of uncertainty into more familiar ecological metrics, such as tree height and biomass estimation, will be important for determining the level of precision needed to answer ecological questions. In addition, our evaluation methods deal exclusively with tree crowns that can be annotated by hand, and therefore ignore subcanopy trees. Finally, it is necessary to understand the influence of both hand-annotation accuracy and the use of rectangular bounding boxes instead of convex hulls. Given the degree of inaccuracy in current algorithms these details are secondary to the overall need for broad improvement in tree detection and segmentation performance. Once an algorithm is found that performs well across a broad array of forest types, this method can be refined by incorporating uncertainty in labeled trees and refining bounding box predictions using methods like raster-based segmentations (e.g. Mask RCNN (He et al., 2017)) or (when LiDAR data is available) draping predictions over point cloud data.

In addition to universal model development, transferring knowledge

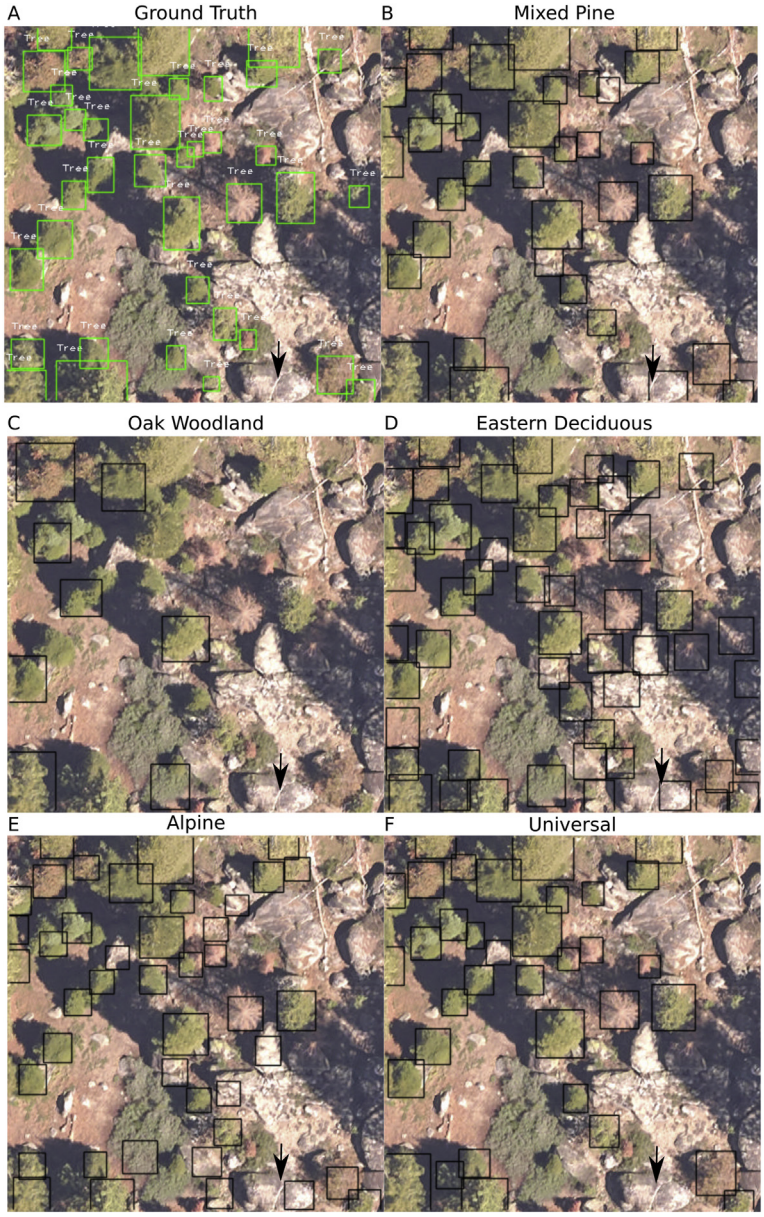

Fig. 6. A sample evaluation plots from the Mixed Pine site predicted by a model built from training data from the same site, from each other site, and a universal model. Ground truth boxes are shown in green (A). Individual trees with a predicted probability greater than 15% are shown in black. B) Predictions from the model trained on mixed-pine annotations. C) Predictions from the model trained Oak Woodland annotations. D) Predictions from the model trained on Eastern Deciduous annotations. E) Predictions from the model trained on Alpine annotations. F) Predicted from the model trained on annotations from all sites. The universal model (F) built from all annotations slightly outperformed all other models, including the model trained only from the Mixed Pine site. For example, the boulder in the bottom right corner is incorrectly classified as a tree by the models trained from Mixed Pine, Alpine, and Eastern Deciduous sites, but is correctly ignored in the Oak Woodland and Universal models.

beyond the NEON sites may be useful for many applied problems. It is currently unknown to what extent features learned from the 0.1 m resolution data used here can be applied to lower resolution satellite data (Karlson et al., 2014) or higher resolution UAV data (Brieger et al., 2019). Cross resolution training has not been fully explored in environmental remote sensing, but Li et al. (2018) recently showed that deep learning networks can learn scale invariant land classifications that can be matched among data sources. Given the ability to collect virtually unlimited pretraining data using our data generation approach, NEON sites can be seen as an ideal training sources for RGB tree models that could then be applied to other data types.

Our deep learning approach uses LiDAR-based pretraining and RGB deep learning to perform individual tree segmentation (Weinstein et al., 2019). The NEON Airborne platform also collects hyperspectral information that may improve generalization across sites with similar species composition. Due to foliar and structural properties, tree species often have distinct spectral signatures which may facilitate distinguishing adjacent tree crowns. Hyperspectral features for tree species classification are relatively common (e.g. Maschler et al., 2018), but few papers have focused on integrating hyperspectral data into tree detection alongside data from other sensors. Hyperspectral data is available for all NEON sites, and we utilized a three-band composite image to assist in annotating the Eastern Deciduous site (Fig. 8), illustrating the usefulness of hyperspectral data to distinguish adjacent tree crowns with human vision. Choosing the best way to represent high-dimensional hyperspectral data in conjunction with the LiDAR and RGB data is non-trivial and will be important for improvements in individual tree detection at broad scales.

Methods to extract ecological information from airborne sensors are maturing due to advancements in computer vision, data availability and sensor quality. Given our results, what are the strengths and limitations ecologists should consider when adding airborne-derived data to their analyses of ecological questions? Remote sensing methods have limited ability to quantify small and subcanopy trees (Aubry-Kientz et al., 2019), and RGB only methods are particularly susceptible to this limitation because they cannot see trees below the sun-exposed canopy. We therefore expect that ecological questions that are strongly determined by canopy trees will benefit the most from remote sensing at broad scales. For example, the total amount of biomass in most forests depends strongly on the largest trees and will be less sensitive to potential

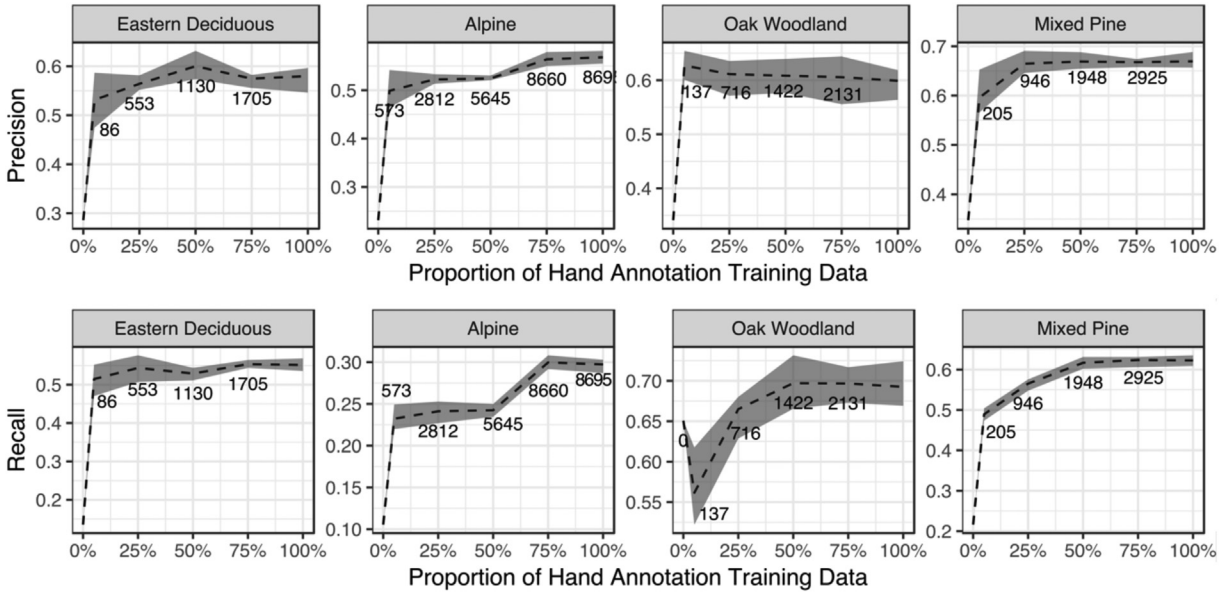

Fig. 7. Ablation curves of the proportion of hand-annotation training data for each site. Values indicate the number of trees in the training dataset for each cutoff. Shaded area is the range of results from rerunning the analysis five times for each site. Note that due to the random sampling among runs, the exact number of trees will vary slightly. For simplicity, we show the mean number of training trees for each threshold.

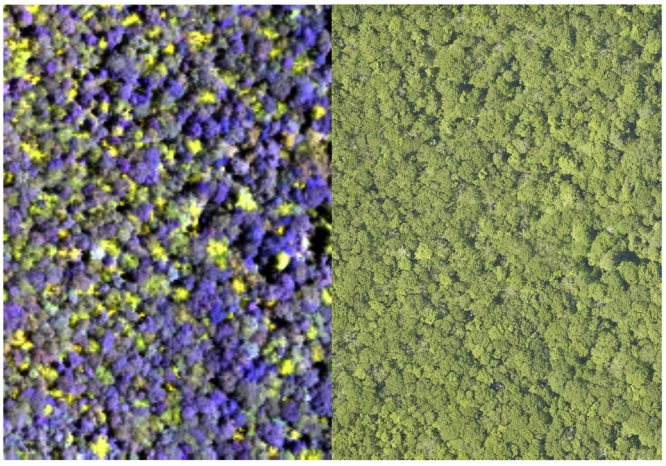

Fig. 8. Composite hyperspectral image and corresponding RGB image for the Eastern Deciduous site. The composite image contained near infrared (940 nm), red (650 nm), and blue (430 nm) bands. Forests that are difficult to segment in RGB imagery may be more separable in hyperspectral imagery due to the differing foliar chemical properties of co-occurring trees.

non-detections of smaller subcanopy trees (Asner et al., 2012; Bastin et al., 2018; Stegen et al., 2011). The inclusion of RGB data may benefit existing large-scale LiDAR-based studies of tree growth since improved individual segmentation will lead to a more accurate matching of individual trees to metadata on taxonomy and health status. Studies of post-landscape disturbance, such as after fires, may be aided by the broader perspective of airborne data (Meng et al., 2018). Most disturbances, such as fire and windstorms, alter the size distribution of forests, including large trees, and thus our approach can provide valuable, detailed landscape scale information about disturbance intensity and impacts (Knapp et al., 2018). To address these questions, we envision a future in which airborne data on tree locations and sizes are a complement to local field surveys in broadening the scale of sampling in complex landscapes.

Data availability

All data and code are made available through a github repo (https://github.com/weecology/DeepLidar/) and archived in Zenodo (doi:https://doi.org/10.5281/zenodo.3347164). We are also actively building a larger dataset as part of a publicly available NeonTreeEvaluation benchmark (https://github.com/weecology/NeonTreeEvaluation).

Acknowledgements

This research was supported by the Gordon and Betty Moore Foundation's Data-Driven Discovery Initiative through grant GBMF4563 to E.P. White and by the National Science Foundation through grant 1926542 to E.P. White, S.A. Bohlman, A. Zare, D.Z. Wang, and A. Singh. This work was supported by the USDA National Institute of Food and Agriculture, McIntire Stennis project 1007080.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2020.101061.

References

Agarwal, S., Du Terrail, J.O., Jurie, F., 2018. Recent advances in object detection in the age of deep convolutional. Neural Netw. 1–106. https://doi.org/10.1016/B0-12-369397-7/00385-X.

Asner, G.P., Clark, J.K., Mascaro, J., Garc, G.A.G., 2012. High-Resolution Mapping of Forest Carbon Stocks in the Colombian Amazon. pp. 2683–2696. https://doi.org/10.5194/bg-9-2683-2012.

Aubry-Kientz, M., Dutrieux, R., Ferraz, A., Saatchi, S., Hamraz, H., Williams, J., Coomes, D., Piboule, A., Vincent, G., 2019. A comparative assessment of the performance of individual tree crowns delineation algorithms from ALS data in tropical forests. Remote Sens. 11, 1086. https://doi.org/10.3390/rs11091086.

Bastin, J.F., Rutishauser, E., Kellner, J.R., Saatchi, S., Pélissier, R., Hérault, B., Slik, F., Bogaert, J., De Cannière, C., Marshall, A.R., Poulsen, J., Alvarez-Loyayza, P., Andrade, A., Angbonga-Basia, A., Araujo-Murakami, A., Arroyo, L., Ayyappan, N., de Azevedo, C.P., Banki, O., Barbier, N., Barroso, J.G., Beeckman, H., Bitariho, R., Boeckx, P., Boehning-Gaese, K., Brandão, H., Brearley, F.Q., Breuer Ndoundou Hockemba, M., Brienen, R., Camargo, J.L.C., Campos-Arceiz, A., Cassart, B., Chave, J., Chazdon, R., Chuyong, G., Clark, D.B., Clark, C.J., Condit, R., Honorio Coronado, E.N., Davidar, P., de Haulleville, T., Descroix, L., Doucet, J.L., Dourdain, A., Droissart, V., Duncan, T., Silva Espejo, J., Espinosa, S., Farwig, N., Fayolle, A., Feldpausch, T.R.,

Ferraz, A., Fletcher, C., Gajapersad, K., Gillet, J.F., do Amaral, I.L., Gonmadje, C., Grogan, J., Harris, D., Herzog, S.K., Homeier, J., Hubau, W., Hubbell, S.P., Hufkens, K., Hurtado, J., Kamdem, N.G., Kearsley, E., Kenfack, D., Kessler, M., Labrière, N., Laumonier, Y., Laurance, S., Laurance, W.F., Lewis, S.L., Libalah, M.B., Ligot, G., Lloyd, J., Lovejoy, T.E., Malhi, Y., Marimon, B.S., Marimon Junior, B.H., Martin, E.H., Matius, P., Meyer, V., Mendoza Bautista, C., Monteagudo-Mendoza, A., Mtui, A., Neill, D., Parada Gutierrez, G.A., Pardo, G., Parren, M., Parthasarathy, N., Phillips, O.L., Pitman, N.C.A., Ploton, P., Ponette, Q., Ramesh, B.R., Razafimahaimodison, J.C., Réjou-Méchain, M., Rolim, S.G., Romero-Saltos, H., Rossi, L.M.B., Spironello, W.R., Rovero, F., Saner, P., Sasaki, D., Schulze, M., Silveira, M., Singh, J., Sist, P., Sonke, B., Soto, J.D., de Souza, C.R., Stropp, J., Sullivan, M.J.P., Swanepoel, B., ter Steege, H., Terborgh, J., Texier, N., Toma, T., Valencia, R., Valenzuela, L., Ferreira, L.V., Valverde, F.C., Van Andel, T.R., Vasque, R., Verbeeck, H., Vivek, P., Vleminckx, J., Vos, V.A., Wagner, F.H., Warsudi, P.P., Wortel, V., Zagt, R.J., Zebaze, D., 2018. Pan-tropical prediction of forest structure from the largest trees. Glob. Ecol. Biogeogr. 27, 1366–1383. https://doi.org/10.1111/geb.12803.

Brieger, F., Herzschuh, U., Pestryakova, L.A., Bookhagen, B., Zakharov, E.S., Kruse, S., 2019. Advances in the derivation of northeast Siberian Forest metrics using high-resolution UAV-based photogrammetric point clouds. Remote Sens. 11, 1447. https://doi.org/10.3390/rs11121447.

Coomes, D.A., Dalponte, M., Jucker, T., Asner, G.P., Banin, L.F., Burslem, D.F.R.P., Lewis, S.L., Nilus, R., Phillips, O.L., Phua, M.H., Qie, L., 2017. Area-based vs tree-centric approaches to mapping forest carbon in Southeast Asian forests from airborne laser scanning data. Remote Sens. Environ. 194, 77–88. https://doi.org/10.1016/j.rse.2017.03.017.

Dai, W., Yang, B., Dong, Z., Shaker, A., 2018. A new method for 3D individual tree extraction using multispectral airborne LiDAR point clouds. ISPRS J. Photogramm. Remote Sens. 144, 400–411. https://doi.org/10.1016/j.isprsjprs.2018.08.010.

Dalponte, M., Coomes, D.A., 2016. Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. Methods Ecol. Evol. 7, 1236–1245. https://doi.org/10.1111/2041-210X.12575.

Gaiser, Hans, de Vries, Maarten, Lacatusu, Valeriu, Williamson, Ashley, Enrico Liscio, D.D., 2018. fizy-r/Keras-retinanet. (https://doi.org/zenodo.1464720).

Gomes, M.F., Maillard, P., Deng, H., 2018. Individual tree crown detection in sub-meter satellite imagery using marked point processes and a geometrical-optical model. Remote Sens. Environ. 211, 184–195. https://doi.org/10.1016/j.rse.2018.04.002.

Graves, S.J., Caughlin, T.T., Asner, G.P., Bohlman, S.A., 2018. A tree-based approach to biomass estimation from remote sensing data in a tropical agricultural landscape. Remote Sens. Environ. 218, 32–43. https://doi.org/10.1016/j.rse.2018.09.009.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Comput. Vis. Pattern Recognit. (CVPR) 2016, 770–778. https://doi.org/10.3389/fpsyg.2013.00124.

He, K., Gkioxari, G., Dollar, P., Girshick, R., 2017. Mask R-CNN. Proc. IEEE Int. Conf. Comput. Vis. 2017-Octob. pp. 2980–2988. https://doi.org/10.1109/ICCV.2017.322.

Heinzel, J., Huber, M.O., 2018. Constrained Spectral Clustering of Individual Trees in Dense Forest Using Terrestrial Laser Scanning Data. https://doi.org/10.3390/rs10071056.

Jakubowski, M.K., Li, W., Guo, Q., Kelly, M., 2013. Delineating individual trees from lidar data: a comparison of vector- and raster-based segmentation approaches. Remote Sens. 5, 4163–4186. https://doi.org/10.3390/rs5094163.

Karlson, M., Reese, H., Ostwald, M., 2014. Tree crown mapping in managed woodlands (Parklands) of semi-arid West Africa using WorldView-2 imagery and geographic object based image analysis. Sensors (Switzerland) 14, 22643–22669. https://doi.org/10.3390/s141222643.

Knapp, N., Fischer, R., Huth, A., 2018. Linking lidar and forest modeling to assess biomass estimation across scales and disturbance states. Remote Sens. Environ. 205, 199–209. https://doi.org/10.1016/j.rse.2017.11.018.

Li, W., Guo, Q., Jakubowski, M.K., Kelly, M., 2012. A new method for segmenting individual trees from the lidar point cloud. Photogramm. Eng. Remote. Sens. 78, 75–84. https://doi.org/10.14358/PERS.78.1.75.

Li, Y., Zhang, Y., Huang, X., Ma, J., 2018. Learning source-invariant deep hashing convolutional neural networks for cross-source remote sensing image retrieval. IEEE Trans. Geosci. Remote Sens. 56, 6521–6536. https://doi.org/10.1109/TGRS.2018.2839705.

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P., 2017. Focal Loss for Dense Object Detection. Proc. IEEE Int. Conf. Comput. Vis. 2017-Octob. pp. 2999–3007. https://doi.org/10.1109/ICCV.2017.324.

Maschler, J., Atzberger, C., Immitzer, M., 2018. Individual tree crown segmentation and classification of 13 tree species using airborne hyperspectral data. Remote Sens. 10, 1218. https://doi.org/10.3390/rs10081218.

Meng, R., Wu, J., Zhao, F., Cook, B.D., Hanavan, R.P., Serbin, S.P., 2018. Measuring short-term post-fire forest recovery across a burn severity gradient in a mixed pine-oak forest using multi-sensor remote sensing techniques. Remote Sens. Environ. 210, 282–296. https://doi.org/10.1016/j.rse.2018.03.019.

Roussel, J.-R., Auty, David, 2019. lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications.

Silva, C.A., Hudak, A.T., Vierling, L.A., Loudermilk, E.L., O'Brien, J.J., Hiers, J.K., Jack, S.B., Gonzalez-Benecke, C., Lee, H., Falkowski, M.J., Khosravipour, A., 2016. Imputation of individual longleaf pine (Pinus palustris mill.) tree attributes from field and LiDAR data. Can. J. Remote. Sens. 42, 554–573. https://doi.org/10.1080/07038992.2016.1196582.

Stegen, J.C., Swenson, N.G., Enquist, B.J., White, E.P., Phillips, O.L., Jørgensen, P.M., Weiser, M.D., Monteagudo Mendoza, A., Núñez Vargas, P., 2011. Variation in above-ground forest biomass across broad climatic gradients. Glob. Ecol. Biogeogr. 20, 744–754. https://doi.org/10.1111/j.1466-8238.2010.00645.x.

Vaglio Laurin, G., Ding, J., Disney, M., Bartholomeus, H., Herold, M., Papale, D., Valentini, R., 2019. Tree height in tropical forest as measured by different ground, proximal, and remote sensing instruments, and impacts on above ground biomass estimates. Int. J. Appl. Earth Obs. Geoinf. 82, 101899. https://doi.org/10.1016/j.jag.2019.101899.

Weinstein, B.G., Marconi, S., Bohlman, S., Zare, A., White, E., 2019. Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. Remote Sens. 11, 1309. https://doi.org/10.3390/rs11111309.

Williams, J., Schönlieb, C.-B., Swinfield, T., Lee, J., Cai, X., Qie, L., Coomes, D.A., 2019. Three-Dimensional Segmentation of Trees through a Flexible Multi-Class Graph Cut Algorithm (MCGC). pp. 1–33.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geosci. Remote Sens. Mag. 5, 8–36. https://doi.org/10.1109/MGRS.2017.2762307.