Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks

Ben. G. Weinstein[1], Sergio Marconi[1], Stephanie Bohlman[2], Alina Zare[3], Ethan White[1]

[1] Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, Florida, USA

[2] School of Forest Resources and Conservation, University of Florida, Gainesville, Florida, USA

[3] Department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida, USA

Abstract

Remote sensing can transform the speed, scale, and cost of biodiversity and forestry surveys. Data acquisition currently outpaces the ability to identify individual organisms in high resolution imagery. We outline an approach for identifying tree-crowns in RGB imagery using a semi-supervised deep learning detection network. Individual crown delineation has been a long-standing challenge in remote sensing and available algorithms produce mixed results. We show that deep learning models can leverage existing lidar-based unsupervised delineation to create generated trees to train an initial RGB crown detection model. Despite limitations in the original unsupervised detection approach, this noisy training data may contain information from which the neural network can learn initial tree features. We then refine the initial model using a small number of higher-quality hand-annotated RGB images. We validate our proposed approach using an open-canopy site in the National Ecological Observation Network. Our results show that a model using 434,551 self-generated trees with the addition of 2,848 hand-annotated trees yields accurate predictions in natural landscapes. Using an intersection-over-union threshold of 0.5, the full model had an average tree crown recall of 0.69, with a precision of 0.61 for visually-annotated data. The model had an average tree detection rate of 0.82 for field collected stems. The addition of a small number of hand-annotated trees improved performance over the initial self-supervised model. This semi-supervised deep learning approach demonstrates that remote sensing can overcome a lack of labeled training data by generating noisy data for initial training using unsupervised methods and retraining the resulting models with high quality labeled data.

Keywords: Deep Learning; Trees; Detection; Remote Sensing; LIDAR; RGB; NEON

1. Introduction

The cost of human observation limits our ability to understand the natural world. Image-based artificial intelligence can advance our understanding of individual organisms, species, and ecosystems by greatly increasing the scale and efficiency of data collection [1]. The growing availability of sub-meter airborne imagery brings opportunities for remote sensing of biological landscapes that scales from individual organisms to global systems. However, the use of this imagery remains limited by the laborious, non-reproducible, and costly annotation of these datasets [2].

Tree detection is a central task in forestry and ecosystem research and both commercial and scientific applications rely on delineating individual tree crowns from imagery [3,4]. While there has been considerable research in unsupervised tree detection using airborne LIDAR (Light Detection and Ranging; a sensor that uses laser pulses to map three dimensional structure) [3,5,6], less is known about tree detection in RGB (red, green, blue) orthophotos. Compared to LIDAR, two dimensional RGB orthophotos are less expensive to acquire and easier to process but lack direct three-dimensional information on crown shape. Effective RGB-based tree detection would unlock data at much larger scales due to increasing satellite-based RGB resolution and the growing use of uncrewed aerial vehicles. Initial studies of tree detection in

RGB imagery focused on pixel-based methods and watershed algorithms to find local maxima among pixels to create potential tree crowns [7]. Combined with hand-crafted rules on tree geometries, these approaches separately performed tree-detection and crown delineation [8,9]. The need to hand-craft tree geometry rules makes it a challenge to create a single approach that encompass a range of tree types [10].

Deep learning is a well-established method for detecting and identifying objects in RGB images but has only recently been applied to vegetation detection [11,12]. Compared to previous rule-based approaches, deep learning has three features that make it ideal for tree detection. First, convolutional neural networks (CNNs) delineate objects of interest directly from training data rather than using hand-crafted pixel features. This reduces the expertise required for each use-case and improves transferability among projects [13]. Second, CNNs learn hierarchical combinations of image features that focus on object-level, rather than pixel-level, representations of objects. Finally, neural networks are re-trainable to incorporate the idiosyncrasies of individual datasets. This allows models to be refined with data from new local areas without discarding information from previous training sets.

The challenge for applying deep learning to natural systems is the need for large training datasets. A lack of training data is a pervasive problem in remote sensing due to the cost of data collection and annotation [14]. In addition, the spatial extent of training data often prohibits field-based verification of annotated objects. For tree detection, the high variation in tree crown appearance due to taxonomy, health status, and human management increases the risk of overfitting when using small amounts of training data [11]. One approach to addressing data limitation in deep learning is “self-supervised learning” (*sensus* [15]), which uses unsupervised methods to generate training data that is used to train supervised models [16]. This approach has recently been applied to remote sensing for hyperspectral image classification [10]. Self-supervision, which relies only on unlabeled data, can be combined with labeled data in a semi-supervised framework (*sensu* Zu 2005), which may improve deep learning on limited training data by providing neural networks the opportunity to learn generalized features on a wider array of training examples, followed by retraining on a smaller number of high quality annotations [17]. Given the imperfect nature of existing unsupervised tree delimitation approaches, it is unknown whether moderate to low quality annotations can be used to generate trees for model training.

In this paper, we propose a semi-supervised pipeline for detecting tree crowns based on RGB data. This pipeline is outlined in Fig. 1. In the proposed workflow, a LIDAR unsupervised algorithm generates initial tree predictions. The bounding box for each tree is extracted and the corresponding RGB crop is used to train an initial deep learning model. Then, using this self-supervised model as a starting point, we retrain the model using a small number of hand-annotations to correct errors from the unsupervised detection. The LIDAR data is used only to improve the initial training of the network. It is not used for the final prediction step. The result is a deep learning neural network that combines unsupervised and supervised approaches to perform tree delineation in new RGB imagery without the need for co-registered LIDAR data. This provides the potential for expanding the use of deep learning in remote sensing applications with limited labeled data by exploring whether generating hundreds of thousands of noisy labels will yield improved performance even though these labeled data are imperfect due to the limitations of the generative algorithm [18].

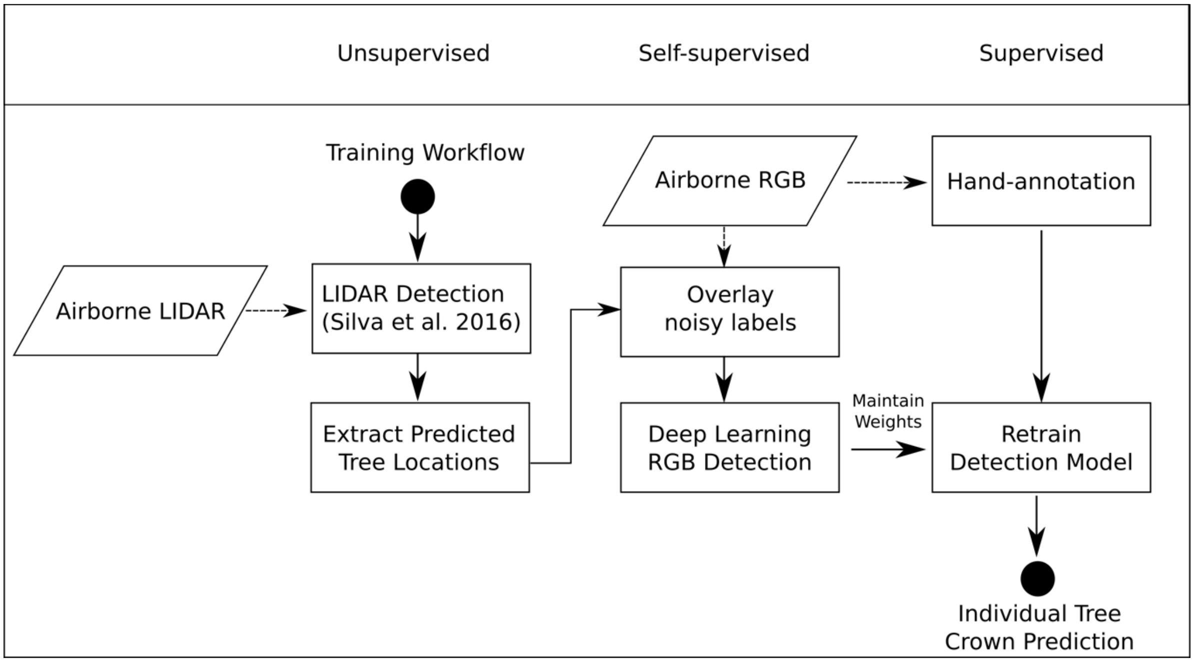

Figure 1. A conceptual figure of the proposed semi-supervised pipeline. A LIDAR-based unsupervised detection generates initial training data for a self-supervised RGB deep learning model. The model is then retrained based on a small number of hand-annotated trees to create the full model.

2. Materials and Methods

2.1. Study Site and Field Data

We used data from the National Ecological Observatory Network (NEON) site at the San Joaquin Experimental Range in California to assess our proposed approach (Figure 2). The site contains open woodland of live oak (*Quercus agrifolia*), blue oak (*Quercus douglasii*) and foothill pine (*Pinus sabiniana*) forest. The majority of the site is a single-story canopy with mixed understory of herbaceous vegetation. All aerial remote sensing data products were provided by the 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 (4-6 points/m^2) which provides high resolution information about crown shape and height. The RGB data are a 1km x 1km mosaic of individual images with a cell size of 0.1 meters. Both data products are georeferenced in the UTM projection Zone 11. In addition to airborne data, NEON field teams semi-annually catalog “Woody Plant Vegetation Structure” (NEON ID: DP1.10098.001), which lists the tag and species identity of trees with DBH > 10cm in 40m x 40m plots at the site. For each tagged tree, the trunk location was obtained using the azimuth and distance to the nearest georeferenced point within the plot. All data are publicly available on the NEON Data Portal (http://data.neonscience.org/). All code for this project is available on GitHub (https://github.com/weecology/DeepLidar) and archived on Zenodo (Weinstein and White 2019).

Figure 2. The San Joaquin, CA (SJER) site (C) in National Ecological Observation Network contains 148 1km^2 tiles (B), each with a spatial resolution of 0.1m. For our analysis, we further divided each tile in 40x40m windows (A) for individual tree prediction (n=729 per 1km^2 tile).

For hand annotations, we selected two 1km x 1km RGB tiles and used the program RectLabel (https://rectlabel.com/) to draw bounding boxes around each visible tree. We chose not to include snags, or low bushes that appeared to be non-woody. In total, we hand-annotated 2,848 trees for the San Joaquin site. In addition to the 1km tile, we hand-annotated canopy bounding boxes on the cropped RGB images for each NEON field plot (n=35), which were withheld from training and used as a validation dataset.

2.2. Unsupervised LIDAR Detection

We tested three existing unsupervised algorithms for use in generating trees for the self-supervised portion of the workflow [19–21]. Existing unsupervised algorithms yield imperfect crown delineations in part because: 1) the algorithms are not designed to learn the specifics of different regions and datasets; 2) it is difficult to design hand-crafted features that are flexible enough to encompass the high variability in tree appearance; 3) distinguishing between trees and vertical objects such as boulders and artificial structures can be difficult with only three-dimensional LIDAR data. We evaluated three available unsupervised LIDAR detection algorithms in order to choose the best performing algorithm to generate training labels [19–21]. We then used the best performing method ([21]) to create initial self-supervised tree predictions in the LIDAR point cloud. This algorithm uses a canopy height model and threshold of tree height to crown width to cluster the LIDAR cloud into individual trees (Figure 3). We used a canopy height model of 0.5m resolution to generate local tree tops, and a maximum crown diameter of 60% of tree height. A bounding box was automatically drawn over the entire set of points assigned to each tree to create the training trees. In total, we generated 434,551 unsupervised tree labels to use during model training.

Figure 3. Example results from the Silva et al. 2016 unsupervised lidar algorithm [21], as implemented in the R liDR package [22]. Two plots from the San Joaquin NEON site are shown (SJER_009, SJER_010).

2.3. Deep Learning RGB detection

Convolutional neural networks are often used for object detection, due to their ability to represent semantic information as combinations of image features. Early applications passed a sliding window over the entire image and treated each window as a separate classification problem. This approach was slow and enforced arbitrary decisions for window size and shape. This was improved by considering potential detection boxes generated by image segmentation techniques [23] or by combining the bounding box proposal and classification into a single deep learning framework [24]. We chose the retinanet one-stage detector [25,26], which allows pixel information to be shared at multiple scales, from individual pixels to groups of connected objects for learning both bounding boxes and image classes. We used a resnet-50 classification backbone pretrained on the ImageNet dataset [27]. We experimented with deeper architectures (resnet-101 and resnet-152) but found no improvement that offset the increased training time.

Since the entire 1km RGB tile cannot fit into GPU memory, we first cut the tile into smaller windows for model training. We experimented with a number of different window sizes and found optimal performance at 400 X 400 pixels due to a balance between memory constraints and providing the model sufficient spatial context for tree detection. This resulted in 729 windows per 1km tile. The order of tiles and windows were randomized before training. Using the pool of unsupervised tree predictions, we trained the network with a batch size of 6 on a Tesla K80 GPU for 8 epochs. After prediction, we passed each image through a non-max suppression filter to remove predicted boxes that overlapped by more than 15%, maintaining only the box with the superior predicted score. One advantage of this neural network approach is that each predicted bounding box has an associated confidence score. We removed boxes within confidence scores less than 0.2.

2.4. Model Evaluation

We used the NEON woody vegetation data to evaluate model recall using field-collected points corresponding to individual tree stems. A field-collected tree point was considered correctly predicted if the point fell within a predicted bounding box. This is a more conservative approach than most over studies, where the field-collected tree point is considered correctly predicted if an edge of the bounding box falls within a horizontal search radius (e.g 3m in [28] to 8m in [29]). Due to these variations in accuracy measurement, it is difficult to establish state-of-art performance, but 70-80% detection rate between predicted trees and field located trees is typical [5,10]. Given the variation in tree appearance and segmentation difficulty, there are too few previous attempts at individual tree crown prediction to provide an expectation for accuracy.

To evaluate the hand-annotated crown areas, we computed recall and precision based on an intersection-over-union score of greater than 0.5 for each predicted crown. The intersection-over-union evaluation metric measures the area of overlap divided by the area of union of the ground truth bounding box and the predicted bounding box. Direct comparisons of predicted and observed crown overlap are rarely performed due to the difficulty of collecting data for a sufficient number of validation examples. The most common approach is to compare the predicted crown area to a matched tree, such as in [30] or use per pixel overlap in visually annotated data [10,31]. Compared to previous works, our use of a minimum 0.5 intersection-over-union sore is more stringent. We chose this value because it more closely resembles the required accuracy for forestry and ecological investigations [32].

3. Results

Initial exploration of existing lidar-based tree detection tools showed that the best performing algorithm [21] was able to correctly recall the crown area of 14% of trees at intersection-over-union score of 0.5 (Table 1). Challenges included over-segmentation of large individual trees, erroneous predicted tree objects based on imperfections in the ground model, and inclusion of non-tree vertical objects (Figure 3).

Table 1. Exploratory analysis of lidar-based unsupervised algorithms. Recall and precision statistics are shown for intersection-over-union with a threshold of 0.5 overlap for the hand annotated trees on the NEON field plots (n=271 trees).

LIDAR Algorithm	Recall	Precision
Li et. al (2012)	0.107	0.021
Dalponte et al. (2016)	0.138	0.083
Silva et al. (2016)	0.142	0.071

Using the bounding boxes from the Silva et al. (2016) predictions, we extracted RGB crops and pretrained the RGB neural network. This self-supervised network had a field collected stem recall of 0.83, and a hand-annotated crown area recall of 0.53 with a precision of 0.32. Retraining the self-supervised model with hand-annotated trees increased the recall of the hand annotated tree crowns to 0.69 with a precision of 0.61 (Table 2, Figure 4). The field collected stem recall did not meaningfully change among models.

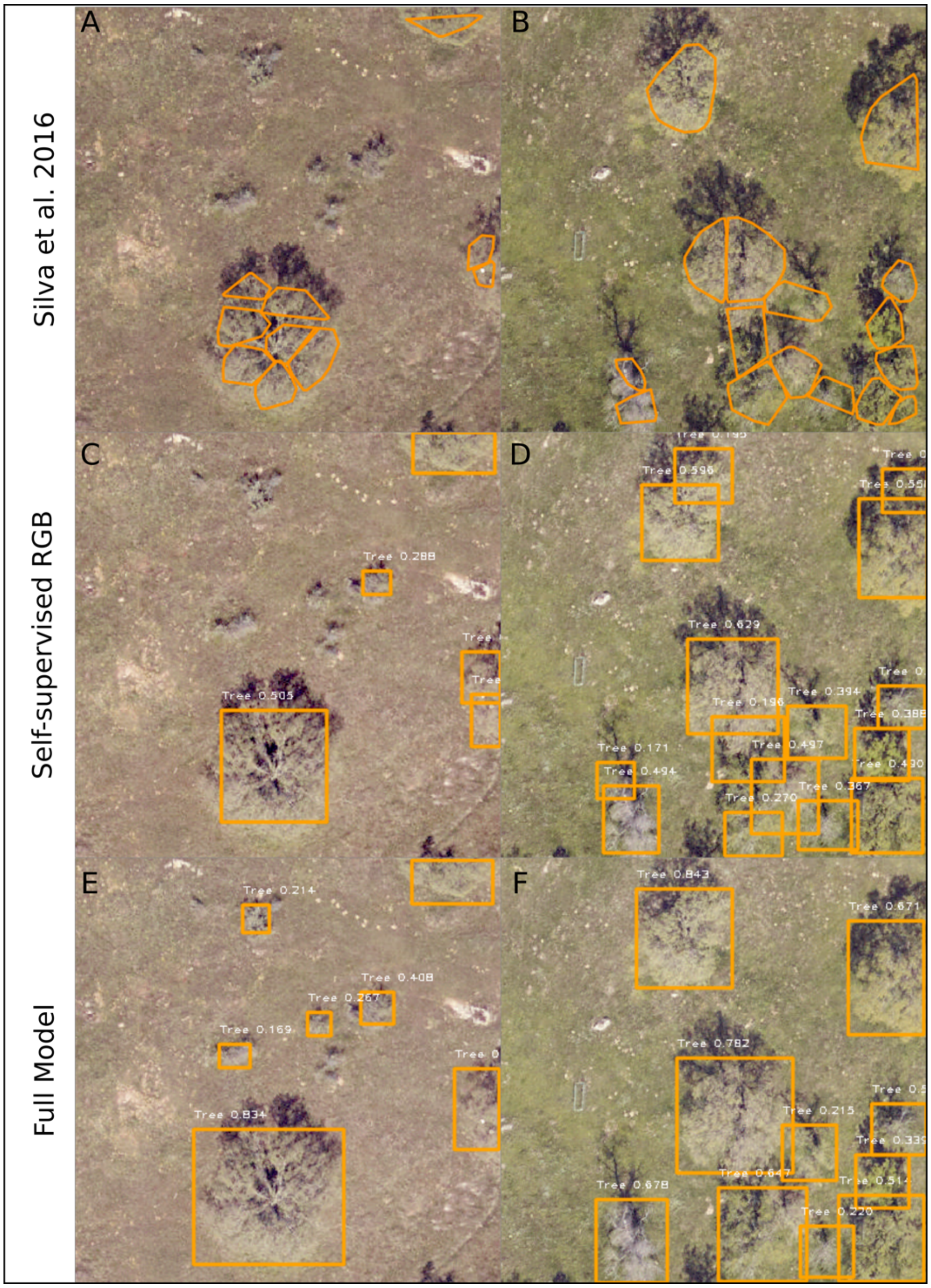

Figure 4. Predicted individual tree crowns for the unsupervised lidar (A, B), self-supervised RGB (C, D) and full (semi-supervised) model (E, F) for two NEON tower plots, SJER_015 (A, C, E), and SJER_053 (B, D, F) at the San Joaquin, CA site. For each tree prediction, the detection probability is shown in white.

Table 2. Evaluation metrics for each of the models. All evaluation was conducted on the 34 NEON field plots. Stem recall was calculated using the field-collected tree stem locations (n=111 trees). Precision and recall for crown overlap was calculated on hand-annotated bounding boxes around each tree crown (n=271 trees) with a minimum predicted probability threshold of 0.5.

Model	**Hand-annotated crown overlap (>50%)**		**Stem Recall**
	Recall	Precision	
Silva et al. 2016	0.14	0.07	0.79
Hand-annotation only	0.38	0.60	0.79
Self-supervised RGB	0.53	0.32	0.83
Full Model	0.69	0.61	0.81

By comparing images of the predictions from the unsupervised lidar detection, the self-supervised RGB deep learning model, and the combined full model, we can learn about the contributions of each stage of the pipeline. The LIDAR unsupervised detection does a good job of identifying trees versus background based on height. Most small trees are well segmented, but there is consistent over-segmentation of the large trees, with multiple crown predictions abutting together. Visual inspection shows that these predictions represent multiple major branches of a single large tree, rather than multiple small trees (Figure 4a). In the self-supervised RGB model, these large trees are more accurately segmented, but there is a proliferation of bounding boxes, and overall lower confidence scores for even well-resolved trees (Figure 4d). This is shown in the precision-recall curves for the hand-annotated validation data, in which the self-supervised model more rapidly declines in precision at higher score thresholds (Figure 5).

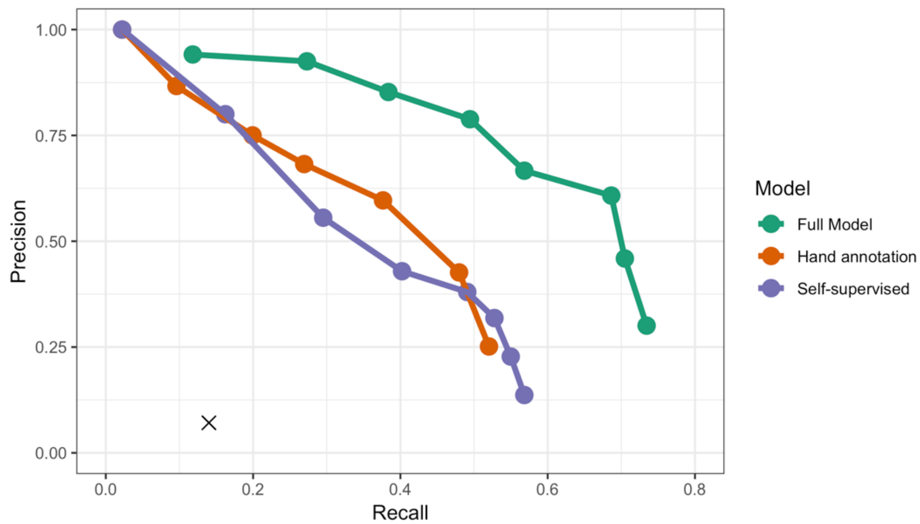

Figure 5. Precision-recall curves for the hand-annotated NEON plots. For each model, we calculated the proportion of correctly predicted boxes for score thresholds [0,0.1,..,0.7]. An annotation was considered correctly predicted if the intersection-over-union (IoU) score was greater than 0.5. The recall and precision scores for the initial lidar-based unsupervised algorithm is shown in black X.

By combining the self-supervised and the hand annotated datasets, the full model reduces the extraneous boxes and improves the segmentation of large trees (Figure 6). The full model has optimal performance in areas of well-spaced large trees (Figure 6b) but tends to under-segment small clusters of trees (Figure 6c).

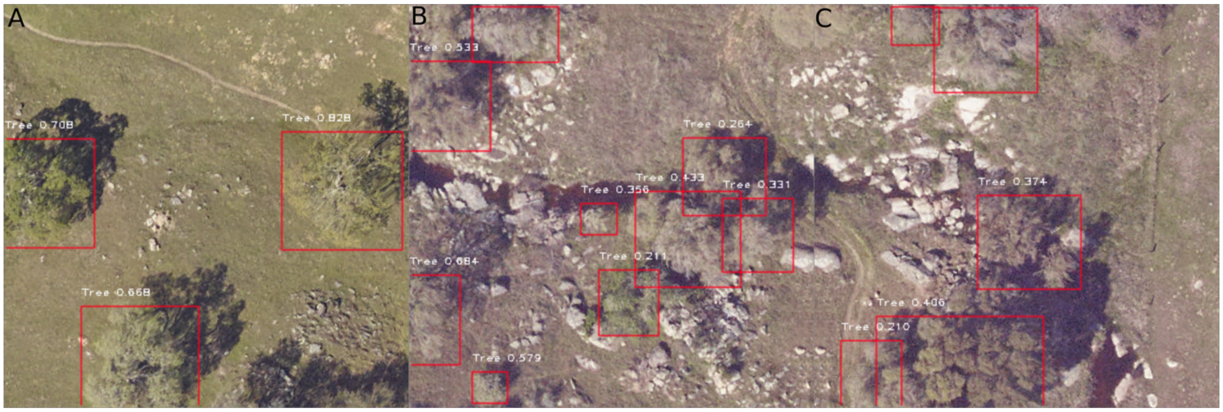

Figure 6. Predictions from the full model on the validation 1km^2 tile. Canopy complexity increases from a) well-defined large trees to B) mixed-species canopies to c) tightly packed clusters of trees. As canopy complexity increases, the full model tends to under-segment small tree clusters.

4. Discussion

Using recent developments in deep learning, we built a neural network-based pipeline for identifying individual trees in RGB imagery. Commercial high resolution RGB data is increasingly available at near global scales, meaning that accurate RGB based crown delineation methods could be used to detect overstory trees at unprecedented extents. To address the long-standing challenge of a lack of labeled training data, we used an unsupervised LIDAR tree detection algorithm to generate labels for initial training. This self-supervised approach allows the network to learn the general features of trees even if the LIDAR-based unsupervised detection is imperfect. The addition of only 2,848 hand-annotated trees generated a final model that performed well when applied to a large geographic area. This approach opens the door for the use of deep learning in airborne biodiversity surveys, despite the persistent lack of annotated data in forestry and ecology datasets.

Many of the false positives in our evaluation dataset were due to disagreements between the hand annotations, unsupervised LIDAR pretraining and RGB prediction in what defines a tree. For example, small trees were often considered too low for inclusion in the LIDAR algorithm (Figure 4a), whereas they were included in the full model based on the hand-annotations (Figure 2b). Similarly, large bushes were sometimes included in hand annotations due to the difficulty of determining overall woody structure. When deploying these models to applied problems, it will be important to have strict quantitative guidelines that define class definitions. Where LIDAR data is available, draping the 2D boxes over the 3D point cloud to filter out points based on vertical height should be useful for improving precision. It should be noted that the quantitative results are likely biased toward the RGB model, since the hand-annotations were made by looking at the RGB, and not the LIDAR data. However, the good recall rate for the field-collected stems suggests that hand annotations were useful in capturing field conditions. An unexpected benefit of the RGB model was the ability to discriminate trees from other vertical objects, such as houses or poles, despite a lack of distinction in the unsupervised LIDAR training data (Figure 7). This may be useful in urban tree detection and other non-forested sites.

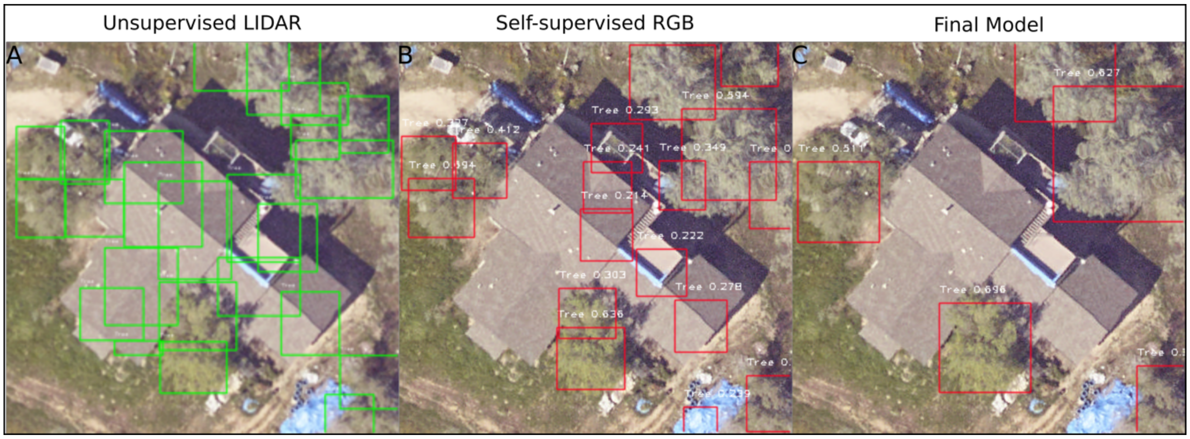

Figure 7. Improvement in prediction quality during the training pipeline. A) Bounding boxes from the lidar-based unsupervised detection erroneously identified artificial structures as trees. B) Predictions from the self-supervised RGB model showed that the addition of RGB data diminished the effect of incorrectly labeled training data, with only edges of the artificial structures maintained as tree predictions. C) In the full semi-supervised model, combining the self-supervised RGB data with hand-annotations eliminated the influence of the original misclassification in the training data, while still capturing the majority of trees in the image.

It is likely that accurate tree detection will be region specific, and that the best model will vary among environments. This will require training a new model for each geographic area using both RGB and LIDAR training data. The proposed approach could save resources by allowing a smaller scale LIDAR flight to generate training data, and then cover a much larger area with less expensive RGB orthophotos. Uncrewed aerial vehicles (UAVs) can be used for capturing LIDAR at high resolution, but at a limited spatial extent. In combination with our method, these UAVs may allow cost effective development of custom regional tree detection models. In addition, the National Ecological Observatory Network, which provided the data for this analysis, has 45 forested NEON sites selected to cover the major ecoclimatic domains in the United States. These sites could serve as pools of LIDAR and RGB data at 10,000 ha scales for regional model training. Combining these two detectors together could produce accurate individual level tree maps at broad scales, with potential applications to forest inventory, ecosystem health, post-natural disaster recovery, and carbon dynamics.

While the semi-supervised deep learning method performed well at the open-canopy test site, geographic areas with complex canopy conditions will be more challenging. The current model only uses LIDAR in the pretraining step. Where available, directly incorporating a LIDAR canopy height model into the deep learning approach should allow the model to simultaneously learn the vertical features of individual trees in addition to the two-dimensional color features in the RGB data. Recent applications of three-dimensional CNNs [33], as well as point-based semantic segmentation [34], provide new avenues for joint multi-sensor modeling. These developments will be crucial in segmenting complex canopies that overlap in the two-dimensional RGB imagery. In addition, recent extensions of region-proposal networks refine bounding boxes to identify the individual pixels that belong to a class [35]. This will provide a better estimate of tree crown area, as trees typically have a non-rectangular shape.

5. Conclusions

Applying deep learning models to natural landscapes opens new opportunities in ecology, forestry, and land management. Despite a lack of high-quality training data, deep learning algorithms can be deployed for tree prediction using unsupervised detection to produce generated trees for pretraining the neural network. Although the lidar-based algorithm used to generate the pretraining data achieved less than 20% recall of hand-annotated tree crowns, the deeply learned RGB features from those data achieved greater than 50% recall. When combined with a small number of hand-annotated images, recall increased to 69% with 60% precision. As shown by the comparison with field-collected stems, the majority of the remaining predictions represent valid trees (>80%), but the overlap with hand-estimated crown area was less than the desired 50%. Many previous papers have used a lower overlap threshold (e.g., 20% overlap in [36]), and we expect this value to improve with a combination of better validation data and more hand-annotated training samples.

In addition to scaling tree detection at much lower costs, there is the potential for this method to provide additional important information about natural systems. The current model could be expanded from a single class, “Tree”, to one that provides more detailed classifications based on taxonomy and health status. For example, splitting the “Tree” class into living and dead trees would provide management insight when surveying for outbreaks of tree pests and pathogens [37], as well as post-fire timber operations [38]. With the addition of hyperspectral data, dividing the tree class into species labels yields additional insights into the economic value, ecological habitat, and carbon storage capacity for large geographic areas [39]. As such, deep learning-based approaches provide the potential for large scale actionable information on natural systems to be derived from remote sensing data.

6. Author Contributions

BGW, EPW, SB and AZ conceived of project design. EW and SM collected the preliminary data. BGW performed the analysis and wrote the text. All authors contributed to the text.

7. Funding

This research was supported by the Gordon and Betty Moore Foundation's Data-Driven Discovery Initiative through grant GBMF4563 to E.P. White. The authors declare no conflict of interest.

8. References

1. Anderson, C.B. Biodiversity monitoring, earth observations and the ecology of scale. *Ecol. Lett.* **2018**.
2. Weinstein, B.G. A computer vision for animal ecology. *J. Anim. Ecol.* **2018**, *87*, 533–545.
3. Wu, B.; Yu, B.; Wu, Q.; Huang, Y.; Chen, Z.; Wu, J. Individual tree crown delineation using localized contour tree method and airborne LiDAR data in coniferous forests. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *52*, 82–94.
4. Caughlin, T.T.; Graves, S.J.; Asner, G.P.; Van Breugel, M.; Hall, J.S.; Martin, R.E.; Ashton, M.S.; Bohlman, S.A. A hyperspectral image can predict tropical tree growth rates in single-species stands. *Ecol. Appl.* **2016**, *26*, 2367–2373.
5. Ayrey, E.; Fraver, S.; Kershaw, J.A.; Kenefic, L.S.; Hayes, D.; Weiskittel, A.R.; Roth, B.E. Layer Stacking: A Novel Algorithm for Individual Forest Tree Segmentation from LiDAR Point Clouds. *Can. J. Remote Sens.* **2017**, *43*, 16–27.
6. Liu, T.; Im, J.; Quackenbush, L.J. A novel transferable individual tree crown delineation model based on Fishing Net Dragging and boundary classification. *ISPRS J. Photogramm. Remote Sens.* **2015**, *110*, 34–47.
7. Gougeon, F.A.; Leckie, D.G. The Individual Tree Crown Approach Applied to Ikonos Images of a Coniferous Plantation Area. *Photogramm. Eng. Remote Sens.* 2006, *72*, 1287–1297.

8. Liu, T.; Im, J.; Quackenbush, L.J. A novel transferable individual tree crown delineation model based on Fishing Net Dragging and boundary classification. *ISPRS J. Photogramm. Remote Sens.* **2015**, *110*, 34–47.

9. Weinmann, M.; Weinmann, M.; Mallet, C.; Brédif, M. A classification-segmentation framework for the detection of individual trees in dense MMS point cloud data acquired in urban areas. *Remote Sens.* **2017**, *9*.

10. Gomes, M.F.; Maillard, P.; Deng, H. Individual tree crown detection in sub-meter satellite imagery using Marked Point Processes and a geometrical-optical model. *Remote Sens. Environ.* **2018**, *211*, 184–195.

11. Li, W.; Fu, H.; Yu, L.; Cracknell, A. Deep Learning Based Oil Palm Tree Detection and Counting for High-Resolution Remote Sensing Images. *Remote Sens.* **2016**, *9*, 22.

12. Guirado, E.; Tabik, S.; Alcaraz-Segura, D.; Cabello, J.; Herrera, F. Deep-learning Versus OBIA for Scattered Shrub Detection with Google Earth Imagery: Ziziphus lotus as Case Study. *Remote Sens.* **2017**, *9*, 1220.

13. Ayrey, E.; Hayes, D. The Use of Three-Dimensional Convolutional Neural Networks to Interpret LiDAR for Forest Inventory. *Remote Sens.* **2018**, *10*, 649.

14. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36.

15. Dahlkamp, H.; Kaehler, A.; Stavens, D.; Thrun, S.; Bradski, G. Self-supervised Monocular Road Detection in Desert Terrain. In Proceedings of the Robotics: Science and Systems II; Robotics: Science and Systems Foundation, 2006.

16. Wu, H.; Prasad, S. Semi-Supervised Deep Learning Using Pseudo Labels for Hyperspectral Image Classification. *IEEE Trans. Image Process.* **2018**, *27*, 1259–1270.

17. Romero, A.; Ballas, N.; Kahou, S.E.; Chassang, A.; Gatta, C.; Bengio, Y. FitNets: Hints for Thin Deep Nets. **2014**, 1–13.

18. Erhan, D.; Manzagol, P.-A.; Bengio, Y.; Bengio, S.; Vincent, P. The Difficulty of Training Deep Architectures and the Effect of Unsupervised Pre-Training. *Twelfth Int. Conf. Artif. Intell. Stat. (AISTATS), JMLR Work. Conf. Procedings* **2009**, *5*, 153–160.

19. Dalponte, M.; Coomes, D.A. Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. *Methods Ecol. Evol.* **2016**, *7*, 1236–1245.

20. Li, W.; Guo, Q.; Jakubowski, M.K.; Kelly, M. A New Method for Segmenting Individual Trees from the Lidar Point Cloud. *Photogramm. Eng. Remote Sens.* **2012**, *78*, 75–84.

21. 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.; et al. Imputation of Individual Longleaf Pine (Pinus palustris Mill.) Tree Attributes from Field and LiDAR Data. *Can. J. Remote Sens.* **2016**, *42*, 554–573.

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

23. Uijlings, J.R.R.; Van De Sande, K.E.A.; Gevers, T.; Smeulders, A.W.M. Selective search for object recognition. *Int. J. Comput. Vis.* **2013**, *104*, 154–171.

24. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Nips* **2015**, 91–99.

25. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection.

Proc. IEEE Int. Conf. Comput. Vis. **2017**, *2017–Octob*, 2999–3007.

26. Hans Gaiser, Maarten de Vries, Valeriu Lacatusu, Ashley Williamson, Enrico Liscio, D.D. fizy-r/Keras-retinanet 2018.

27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *Comput. Vis. Pattern Recognit. (CVPR), 2016* **2016**, 770–778.

28. Vastaranta, M.; Kankare, V.; Holopainen, M.; Yu, X.; Hyyppä, J.; Hyyppä, H. Combination of individual tree detection and area-based approach in imputation of forest variables using airborne laser data. *ISPRS J. Photogramm. Remote Sens.* **2012**, *67*, 73–79.

29. Duncanson, L.I.; Cook, B.D.; Hurtt, G.C.; Dubayah, R.O. An efficient, multi-layered crown delineation algorithm for mapping individual tree structure across multiple ecosystems. *Remote Sens. Environ.* **2014**, *154*, 378–386.

30. 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.; et al. Area-based vs tree-centric approaches to mapping forest carbon in Southeast Asian forests from airborne laser scanning data. *Remote Sens. Environ.* **2017**, *194*, 77–88.

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

32. Jeronimo, S.M.A.; Kane, V.R.; Churchill, D.J.; McGaughey, R.J.; Franklin, J.F. Applying LiDAR Individual Tree Detection to Management of Structurally Diverse Forest Landscapes. *J. For.* **2018**, *116*, 336–346.

33. Zhou, Y.; Tuzel, O. VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection. **2017**.

34. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. PointNet: Deep learning on point sets for 3D classification and segmentation. *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017* **2017**, *2017–Janua*, 77–85.

35. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. *Proc. IEEE Int. Conf. Comput. Vis.* **2017**, *2017–Octob*, 2980–2988.

36. Wallace, L.; Lucieer, A.; Watson, C.S. Evaluating tree detection and segmentation routines on very high resolution UAV LiDAR ata. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7619–7628.

37. Wulder, M.A.; Dymond, C.C.; White, J.C.; Leckie, D.G.; Carroll, A.L. Surveying mountain pine beetle damage of forests: A review of remote sensing opportunities. *For. Ecol. Manage.* **2006**, *221*, 27–41.

38. Vogeler, J.C.; Yang, Z.; Cohen, W.B. Mapping post-fire habitat characteristics through the fusion of remote sensing tools. *Remote Sens. Environ.* **2016**, *173*, 294–303.

39. Deng, S.; Katoh, M.; Yu, X.; Hyyppä, J.; Gao, T. Comparison of tree species classifications at the individual tree level by combining ALS data and RGB images using different algorithms. *Remote Sens.* **2016**, *8*.