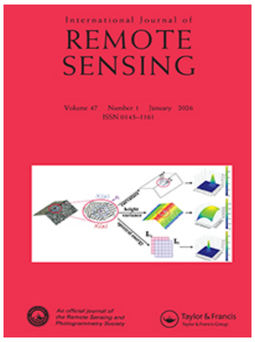

International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: www.tandfonline.com/journals/tres20

Applying a convolutional neural network (CNN) to Virginia's forests: how forest type and age can impact individual tree segmentation

Alison L. Ritz, Randolph H. Wynne, Valerie A. Thomas, Fabien H. Wagner, P. Corey Green, Todd A. Schroeder & Sassan Saatchi

To cite this article: Alison L. Ritz, Randolph H. Wynne, Valerie A. Thomas, Fabien H. Wagner, P. Corey Green, Todd A. Schroeder & Sassan Saatchi (2026) Applying a convolutional neural network (CNN) to Virginia's forests: how forest type and age can impact individual tree segmentation, International Journal of Remote Sensing, 47:1, 218-245, DOI: 10.1080/01431161.2025.2598075

To link to this article: https://doi.org/10.1080/01431161.2025.2598075

Published online: 07 Dec 2025.

INTERNATIONAL JOURNAL OF REMOTE SENSING
2026, VOL. 47, NO. 1, 218–245
https://doi.org/10.1080/01431161.2025.2598075

Applying a convolutional neural network (CNN) to Virginia's forests: how forest type and age can impact individual tree segmentation

Alison L. Ritz [a], Randolph H. Wynne[a], Valerie A. Thomas[a], Fabien H. Wagner[b,c], P. Corey Green[a], Todd A. Schroeder[d] and Sassan Saatchi[b,c]

[a]Forest Resources and Environmental Conservation, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA; [b]Institute of Environment and Sustainability, University of California, Los Angeles, CA, USA; [c]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA; [d]USDA Forest Service, Southern Research Station, Knoxville, TN, USA

ABSTRACT

Convolutional neural networks (CNNs), specifically U-Nets, successfully detect and segment individual tree crowns when applied to high spatial resolution remotely sensed data. However, much of the prior work for individual tree detection and segmentation has been in arid climates such as the western United States and the African Sahel. In this study, we used the U-Net-id CNN architecture and National Agriculture Imagery Program (NAIP) imagery to investigate forest-cover-specific training on individual tree crown detection in Virginia, U.S. We trained three models using hand-delineated crowns: one model with mixed/deciduous forests (3,457 trees), one with pine plantations (10,680 trees), and one that combined all the training data (14,137 trees) from both taxonomic groups. The training data was selected using 128×128 pixel patches across the study area where one interpreter hand-delineated all the tree crowns in the patch. Accuracy for the models was assessed over the entire model application area, by forest cover, and plantation age groups. Accuracy assessment included comparing the hand delineation and model results for tree counts and crown area as well as precision, recall, and *F*1 scores calculated. The Combined model performed poorly, achieving an *F*1 score of 0.4. However, further assessment of the forest cover and plantation age level showed increased *F*1 scores. The Plantation model's *F*1 scores indicated the best-performing age class was the 16–19-year-olds, with a score of 0.74. However, the younger plantations had *F*1scores of 0.42, 0.28, and 0.07 for the 8–11, 4–7, and 0–3-year-olds, respectively. The best count and crown area relationships came from the Mixed/Deciduous model, which achieved an *F*1 score of 0.64. In this study, we found that the representative quality of the training data versus a larger quantity of data more heavily impacts this U-Net-id CNN architecture and demonstrates the impact of training data quality on a CNN model performance.

ARTICLE HISTORY
Received 7 April 2025
Accepted 26 November 2025

KEYWORDS
Convolutional neural network; individual tree crown (ITC) segmentation; instance segmentation

CONTACT Alison L. Ritz alritz20@vt.edu Forest Resources and Environmental Conservation, Virginia Polytechnic Institute and State University, 310 West Campus Dr., Blacksburg, VA 24061, USA

1. Introduction

There is a growing focus on using remote sensing for individual tree level measurements. Monitoring forests at the individual tree level can provide more detailed information on the forest, such as stem counts and trees per acre (TPA) (Pont et al. 2015), crown area (Freudenberg, Magdon, and Nölke 2022; Weinstein et al. 2019), and biomass estimates (Dalponte et al. 2018; Graves et al. 2018). There is also evidence of using tree level mapping to detect trees outside of forests that have been missed or excluded by previous studies (Reiner et al. 2023). These tree level measurements not only help local forest managers make informed decisions about their management practices (Keefe, Zimbelman, and Picchi 2022), but they also provide information on forest health at the global level (Trumbore, Brando, and Hartmann 2015).

The National Agricultural Imagery Program (NAIP) is sub-metre, high-resolution imagery that has previously been used for mapping individual trees outside of forests (Meneguzzo, Liknes, and Nelson 2013), detecting tree mortality from fire and pests (Bishop et al. 2014; Gartner et al. 2015), estimating tree cover and biomass of evergreen woodlands (Hulet et al. 2014), and mapping dead trees for forest loss (Cheng et al. 2024). The combination of NAIP's high-resolution imagery with machine learning techniques for tree cover detection has shown promise, however, there is evidence of misclassification (Meneguzzo, Liknes, and Nelson 2013) and processing problems due to the size of the data (Basu et al. 2015). In a comparison of satellite and aerial imagery like NAIP, Freudenberg, Magdon, and Nölke (2022) found that aerial imagery performed better at tree crown segmentation when combined with a convolutional neural network (CNN), a deep learning network for object detection (Z. Li et al. 2021).

CNNs can delineate objects of interest from training data without the need of thousands of hand-crafted pixel features. This allows them to focus on object-level detection versus pixel-level. There are many different types of CNN backends that have been created for various image recognition and detection methodologies, including AlexNet (Krizhevsky, Sutskever, and Hinton 2017), ResNet (He et al. 2016), DenseNet (Huang et al. 2017), and GhostNet (Han et al. 2020). In Weinstein et al. (2019) they used ResNet-50 image classification to identify individual oak and pine tree crowns at a California National Ecological Observation Network (NEON) site, noting that the larger trees were more accurately segmented in their model. Using the You Only Look Once (YOLO) (Redmon and Farhadi 2018) and Mask R-CNN (He et al. 2017) models, Pérez-Carrasco et al. (2022) counted and segmented individual pine tree crowns with an average F1 score (the harmonic mean of precision and recall) of 0.69 for the YOLO model and 0.80 for the Mask R-CNN model. Although the Mask R-CNN is one of the fastest CNN models for image segmentation, all training data for this model is created by hand. For larger areas where more training is needed, the U-Net CNN is a better model structure as it includes steps for data augmentation so more training data can be acquired from a smaller number of hand-delineated samples (Ronneberger, Fischer, and Brox 2015). The U-Net model also typically has a higher Dice score when predicting the segmentation mask compared to the Mask R-CNN (Vuola, Akram, and Kannala 2019)

The U-Net semantic segmentation model, first described by Ronneberger, Fischer, and Brox (2015), uses pooling and upsampling operators to increase the resolution of the images, which allows for seamless segmentation of large images by using an overlappingtile strategy. Then, U-Net CNN has proven its capability for tree crown detection and demonstrated its high accuracy when compared to other machine learning and AI architectures (Kattenborn et al. 2021; Zhao et al. 2023). In Freudenberg et al. (2019), their two U-Net architectures received F1 scores of over 0.93 for tree segmentation in both a forested and city area, which outperformed the AlexNet with an F1 score of 0.85. In Schiefer et al. (2020) they used the U-Net model to segment and classify tree species from 2 cm resolution imagery with F1 scores as high as 0.73, however, the authors found large differences by cover type with the lowest F1 score of 0.27 for deciduous crowns and the highest F1 score of 0.91 for pine crowns. However, since the U-Net model is a semantic segmentation model it identifies tree and non-tree pixels, but it does not separate them into individual tree crowns with their own geometry, which is a problem for tree counts in dense canopies. Some researchers have used the Deep Watershed (Bai and Urtasun 2017) algorithm to separate the tree crowns after the U-Net model is applied to the imagery (Cheng et al. 2024). Korznikov et al. (2021) have adapted the U-Net model into a 'U-Net-like' architectures where they combine the original semantic segmentation model with pixel-based classifiers, like random forest, for image recognition and seperation of tree species. (Freudenberg, Magdon, and Nölke 2022) used the U-Net to create a tree cover mask and then addad a polygon extraction step from the watershed transform model to delienate the individual tree crowns. However, these methods were completed with two or more separate models which can be combersome to replicate. One recent adaptation of the U-Net model comes from Wagner et al. (2020) where the U-Net sematic segmention model has an instance segmentaion backend embedded within the model. It is called the U-Net-id and it begins with three interconnected U-Net paths that predict the object border, object segment, and the object inner segment and then extracts the inner segments and adds a buffer to individualize the objects. This allows the model to both identify and individualize trees so that we can get crown counts and crown area measurements from one model.

Overall, the combination of both high-resolution imagery and CNN object detection models have displayed promising results for individual tree detection. However, crown type and crown size has been shown to have an impact on CNN object detection success. In Pérez-Carrasco et al. (2022), the authors found that the younger trees with smaller crowns performed poorly in both models with lower F1 scores, compared to the older trees. Freudenberg, Magdon, and Nölke (2022) found that larger trees tended to be over-segmented by the model and only 40% of trees smaller than 10 m^2 were detected. The evergreen trees also show greater success overall with the model than the deciduous. In S. Li et al. (2023), the U-Net adapted architecture for tree crown segmentation over-counted in the dense conifer forests but undercounted in the dense deciduous forests, which also underestimating crown area across all forest types. These studies demonstrate that individual crown segmentation may be influenced by tree crown size and cover type.

Based on the review completed by Zhao et al. (2023) and our own literature review, other studies are limited in their exploration the use of high-resolution imagery and U-Net CNNs with an instance segmentation backend on individual tree crown segmentation over forests that are highly diverse in terms of species and structure, such as the eastern U.S. The ecosystem composition of the eastern U.S. not only includes an array of species and crown structures but also homogeneous stands of pine plantations with a variety of crown sizes. Based on the findings discussed by previous studies with similar methodologies and data sources, we believe that there is potential for continued research on the specifics behind localized training such as species, structure, and age for the CNN object detection with instance segmentation over high-resolution imagery for individual tree crown mapping in more diverse tree canopies, such as those in the eastern U.S.

Considering the need for individual crown counting and segmentation across forest cover types, the objective of this work was to determine the impact of training data on the U-Net's object detection and U-Net-id's instnace segmentation in the eastern U.S., which has been successful in delineating small objects in very high-resolution imagery, such as buildings. Specifically, we will assess various forest cover types and crown size/age to investigate the effects on the U-Net-id CNN model when applied to the diverse canopy cover of Virginia, U.S.

2. Materials and methods

2.1. Study area

The study area was comprised of Brunswick, Lunenberg, and Mecklenburg counties from Virginia, U.S. (Figure 1, Table 1). The selection criteria for the counties used in the study area included: the most common forest type of loblolly/shortleaf pine (Pinus taeda/Pinus echinata) or oak/hickory (Quercus/Carya), greater than 50% of the Forest Inventory and Analysis (FIA) plots were forested plots (to indicate the approximate fraction of forested area), and the date of NAIP imagery acquisition was between May and September to ensure leaf on conditions. These counties were selected using information from the FIA DataMart webpage, the Landscape Change Monitoring System (LCMS), the National Landcover Dataset (NLCD), and NAIP imagery metadata. FIA data used in this analysis was downloaded from the (FIA DataMart 2023; Burrill et al. 2024). LCMS is a data suite that highlights change in the United States using the latest remote sensing technology (USDA Ag Data Commons 2022). Specifically, the Fast Loss product shows different classifications of change yearly over the past four decades. NLCD is produced by the U.S. Geological Survey (USGS), in association with the Multi-Resolution Land Characteristics (MRLC) Consortium, for the U.S. (EROS Center 2018).

The most common forest types for each county were calculated using plot level forest condition information. These conditions were divided by county using unique identification codes and then the most common forest type was selected for each county. Because there are many sub-classes of forest type, we used the more general forest type groups, such as the loblolly/shortleaf pine group, specified in Appendix D of Burrill et al. (2024). Next, the number of forested plots for each Virginia county was calculated using the plot level forest condition information. These were totalled by county. By FIA definition, a forested plot is at least one acre in size, 120 feet wide, and 10% canopy cover of qualifying tree species *(USDA 2024)*. Lastly, the NAIP 2021 metadata was used to map the date of each flight line over the state resulting in 154 images in the study area. Because we only needed a sample of the study area to train and the model on, and to limit the storage and processing requirements, 10% (15) of the NAIP images were randomly selected to use as training data, and 5% of the images (7) were used for testing. There were approximately 311 acres of training and testing data within these 22 images.

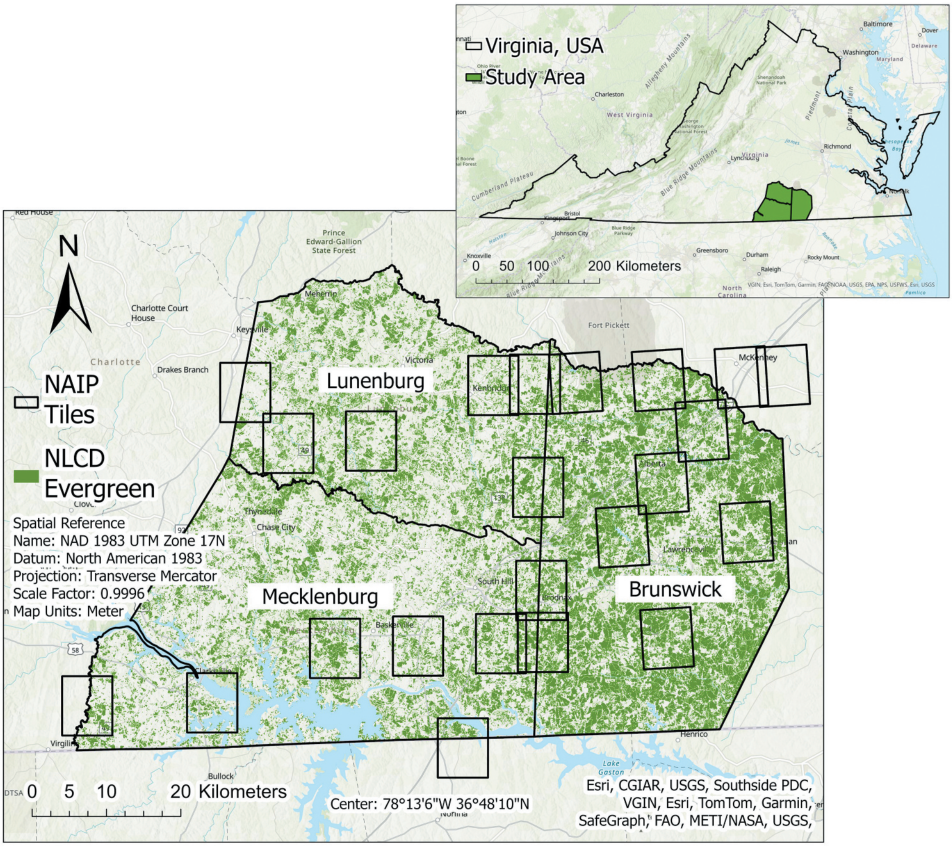

Figure 1. Lunenburg, Mecklenburg, and Brunswick counties in Virginia, U.S. The black outlined rectangles are the NAIP images used for this study. The green background layer is the locations of NLCD evergreen classifications.

Table 1. County and FIA information for the study area.

County Name	Area (km^2)	Total FIA Plots	FIA Forested Plots	Most Common Forest Type
Brunswick	943,786	724	565	Loblolly/Shortleaf
Lunenburg	716,707	480	439	Oak/Hickory
Mecklenburg	1,125,984	695	498	Loblolly/Shortleaf
Total	2,786,477	1,899	1,502	–

To compare the model across forest types, pine plantations and mixed/deciduous forests were separated into two secondary study areas within the main study area. The evergreen forests land clover class from the NLCD was used to distinguish pine plantations from surrounding mixed/deciduous forests in the study area. There were approximately 83 acres of pine plantation data and 228 acres of mixed/deciduous forest data. The mixed/deciduous forest study area was made up of all the training and testing areas that were not included in the pine plantation study area.

The study area for the pine plantations was further broken down by crown size/age. The LCMS Fast Loss dataset was used to roughly determine plantation age across the

Table 2. Pine plantation age distribution that was used in model training and testing.

Ages	% Area Included
0–3	22
4–7	3
8–11	15
12–15	13
16–19	17
20^+	30

study area to ensure a variety of stand ages in the training imagery. Ages were calculated by subtracting the year of LCMS Fast Loss from 2021 with the assumption that the plantation was replanted the same year as harvest (if harvested in 2021, the age is 0, and if harvested in 1985, the age is 36). Ages were verified using Google Earth Pro. The ages were then aggregated into groups of 4 years to account for any remaining error in age estimation. For example: Group 1 ranges from 0 to 3 years, Group 2 ranges from 4 to 7 years, and so on with Group 6 including those 20 years old and up. A percentage of the canopy cover for each age class included in the training can be found in Table 2. It should be noted that the 4–7-year-old plantations were difficult to locate and therefore had a smaller percentage of the overall training and testing data.

2.2. NAIP data

The NAIP imagery for Virginia was acquired with the Leica ADS-100 sensor with flight and sensor control management system (FCMS) firmware. The collection was done with a combination of twin-engine aircraft, flying at 27,100 ft above mean terrain in summer 2021. The cameras are calibrated radiometrically and geometrically by the manufacturer with 27% side overlap and a 0.6-m (60 cm) ground sampling distance (USDA 2013). NAIP imagery was formatted to the UTM coordinate zone and may have up to 10% cloud cover. The imagery has 8-bit pixel depth and includes bands in the red (600–700 nm), green (500–600 nm), blue (400–500 nm), and near-infrared (800–900 nm) spectrum, referred to as bands 1–4, respectively (USDA 2017). The collection was organized by the USDA-FSA Aerial Photography Field Office (Surdex Corporation 2021).

2.3. Convolutional Neural network architecture and code

We used the U-Net-id CNN architecture described in Wagner (2020) to detect individual crowns with NAIP imagery in Virginia. This model structure was based off the classical U-Net CNN from Ronneberger, Fischer, and Brox (2015) but includes a backend that allows for instance segmentation of the U-Net outputs. The U-Net-id model begins with the architecture of the classical U-Net replicated three times as separate paths to segment the imagery into tree and non-tree labelled pixels for the object border, the object segment, and then the object inner segment. The model then concatenates these three layers through two convolutional layers and an activation layer to then predict and segment the individual objects in each layer. In post processing of these layers, the model separates the features by extracting the inner segments (which do not overlap with one another) and adds a buffer of a defined number of pixels to make the final individualized object segment.

The model input is a 128 × 128 pixels wide NAIP RGB image at 0.6 m spatial resolution (Figure 2). The image randomly underwent data augmentation, which included rotating (0–360°) or flipping (horizontal/vertical) the images and adjusting the current values of the brightness (30–100%) or hue (0–200%). No changes were made to the saturation. This was to help with overfitting by the model (Shorten and Khoshgoftaar 2019). Each image is then input into the three U-Net-id segments which outputs a mask of 128 × 128 × 1 for each segment (Figure 3). The first segment is the full crown object segment. The second segment is the crown border which extends for 4 pixels outside and 3 pixels inside the

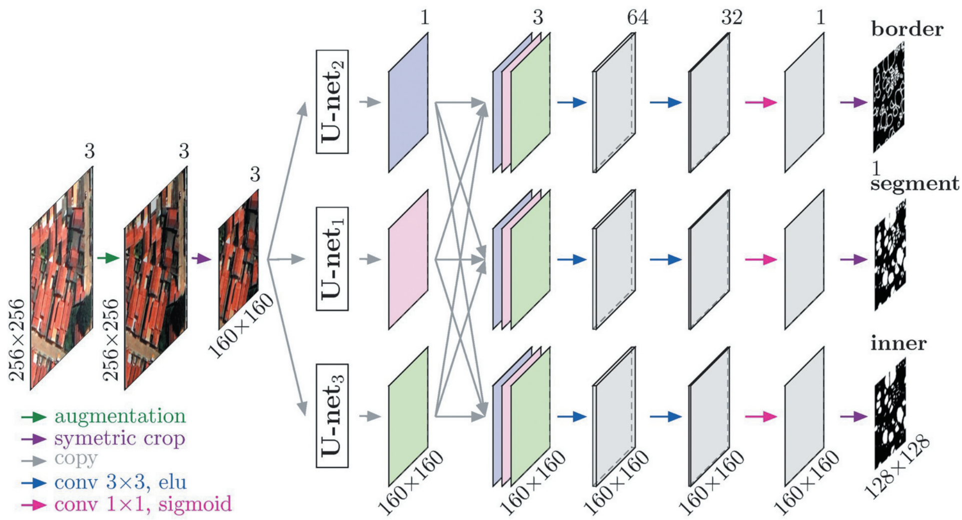

Figure 2. U-Net-id model architecture, adapted from Wagner et al. (2020). The three units in the model are identical but each unit results in one of the three predictions: crown boarder, crown segment, and crown inner segment.

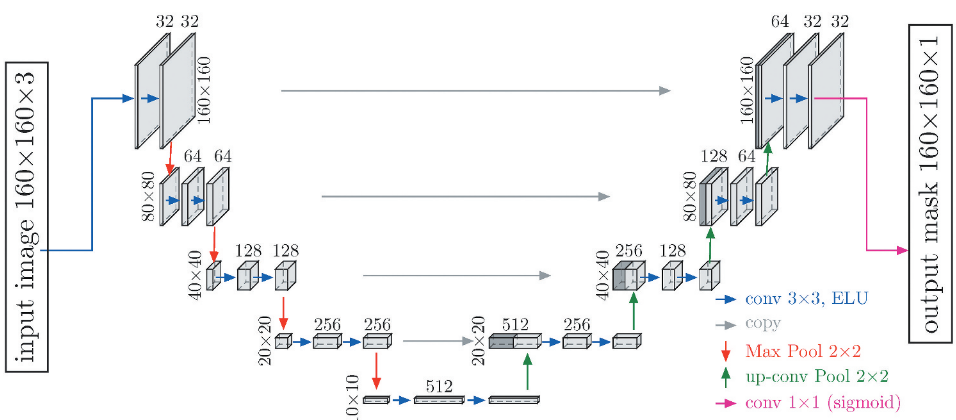

Figure 3. The U-Net model with layer and filter size characteristics, adapted from Wagner et al. (2020). The three U-Nets from Figure 2 are identical in this architecture.

object segment. The third segment is the crown inner segment which is the full object segment reduced by 2 pixels. This allows for an overlap of all three masks of 1–3 pixels. The U-net activation layers for each segment are then concatenated into each of the three paths, making a new $128 \times 128 \times 3$ mask. The model then iterates each mask over two convolutional layers, the first with 64 filters and the second with 32 filters. Finally, the model makes the three separate predictions for the three segment types with the last convolutional layer using a sigmoid activation function. The instances are made in post processing where the inner segments get a 2-pixel buffer to make a mask instance for each crown. These masks are then converted to individual polygon features. The model had a total of 25,994,598 features, of which 25,973,478 were trainable. 80% of the images were used for training and 20% for validation in the model.

We used a standard stochastic gradient descent optimization for the network training. We used a custom loss function that uses the sum of the binary crows-entropy and Dice coefficient-related loss for each of the predicted masks (Allaire and Chollet 2016; Dice 1945). Each model was trained for 10,000 epochs with a batch size of 50 images, a decay rate of $1e^{-15}$, and learning rate of $1e^{-5}$. It used mixed precision with a custom dice coefficient and a binary cross entropy loss function. Code for this can be found at Wagner (2020). The weighted model with the lowest validation loss for each training set was kept. After each model was trained, it was applied to the entire study area to create an instance segment for each of the crowns identified by the model. The model application also had a learning rate of $1e^{-5}$ with the same custom dice coefficient and binary cross entropy loss functions as well as a root mean square propagation (RMSprop) optimizer function. Although many studies might assess the model output and continue to train in the areas that did not perform well, we decided to apply the model only once to evaluate the initial training impact. This allowed us to inspect the original training data for its quality and model impact without bias towards areas that require more training so that the model may perform better. Continuous training is time-consuming and would require more storage space and processing time since, in theory, there is always room for improvement.

The models were trained and implemented in R using an interface to Keras and a TensorFlow backend (Abadi et al. 2015; Allaire and Chollet 2016; R Core Team 2016). This took place using the Graphics Processing Unit (GPU) on an NVIDIA GeForce RTX 2060 with an 11th Gen Intel(R) Core(TM) i7 at 3.60 GHz with 32 GB of installed RAM. Training and implementing the three models over the study area took approximately 74 h.

2.4. Creating the training datasets

Before training began, the interpreter applied an un-trained U-Net-id model over the study area. This helped the interpreter to identify any false positive areas in the NAIP imagery for training. The interpreter then used a 128×128 pixel grid created to match the size of the NAIP imagery (with some buffer) to section out the NAIP image and un-trained model results (Figure 4). The interpreter then used the shapefile and grid layer to randomly select a column in each image and label every 128×128 pixel patch from that column as either: (*a*) the model predicted there were crowns when there were none, or (*b*) the crowns were individually distinguishable and able to be hand drawn. An example of each correction type can be seen in Figure 5. If the trees in the patch were

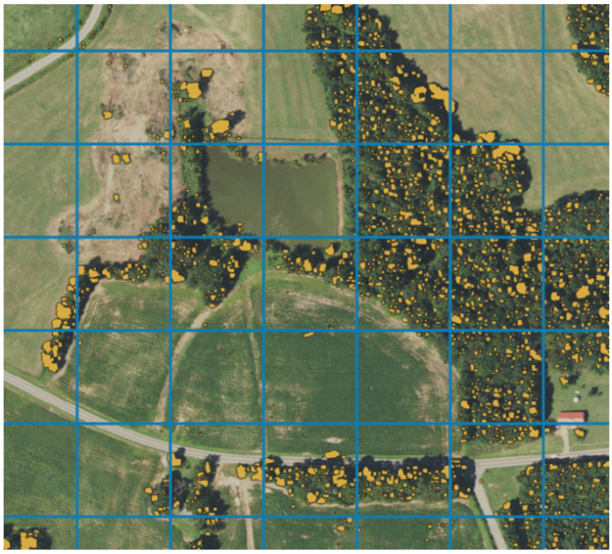

Figure 4. An example of the initial, un-trained U-Net-id crown segmentation (yellow polygons) over the study area NAIP imagery. The blue lines are the grid layer that segments the image into 128 × 128 pixel patches.

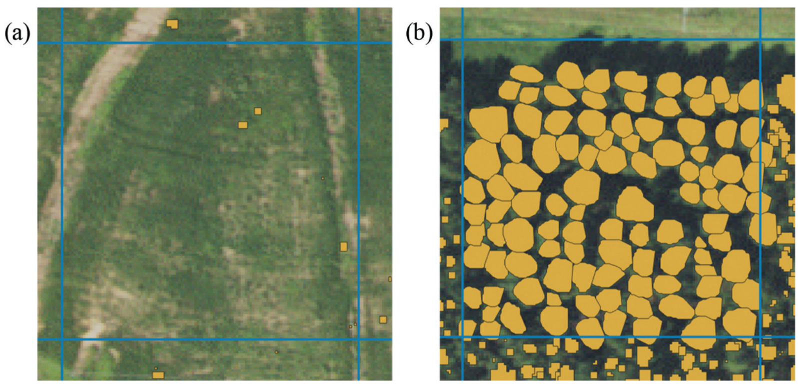

Figure 5. An example of each correction type within the study area imagery. (a) demonstrates where there are no crowns, but the model indicated that there were. (b) demonstrates where the crowns were separable and hand-drawn for training.

too dense to delineate by hand that patch was skipped. Then, the interpreter went back to each 128 × 128 pixel patch that was labelled as (*b*) and hand-delineated each tree crown in that patch. Hand-delineations were done in QGIS using the *Add Polygon* feature in the *Toggle stream Digitizing Mode*. Each hand-delineated polygon was then converted back into a binary mask and both patch types were used to train the new model.

There was a total of 214 patches coded as correction type (b) a total of 14,137 hand-drawn individual tree crowns. Within the 214 patches, 156 were labelled as mixed/deciduous forests and 57 were labelled as pine plantations. The mixed/deciduous forest had a total of 3,457 hand-drawn trees and the pine plantations had a total of 10,680 hand-drawn trees. The hand-drawn training dataset was comprised only crowns visible from a top-down view of the canopy. All tree crown delineations were assumed to include only the dominant and co-dominant canopy in the forested areas. The hand-drawn tree crowns used for training were all drawn by the same person. That individual used their best judgement to delineate the visible crown in each 128 × 128 pixel training image. Trees were delineated even with shadows, and the interpreter made their best judgement for the crown edge. Many of the mixed/deciduous forests exhibited tightly clustered crowns in an irregular spacing and shape, making them a challenge to separate. Although the pine plantations also displayed a tight clustering of their crowns, they were easier to delineate due to their uniform spacing from the row planting.

2.5. Model development

2.5.1. Combined model

The Combined model used all 14,137 hand-delineated trees to train the model. This model was then applied to the entire study area and assessed across all forest cover types.

2.5.2. Mixed/Deciduous model

The Mixed/Deciduous model used only the 3,457 hand-delineated trees within its study area to train the model. This model was then applied to the entire study area but only assessed in other mixed/deciduous forests.

2.5.3. Plantation model

The Plantation model used only the 10,680 hand-delineated trees within its study area to train the model. This model was then applied to the entire study area but only assessed in other pine plantations.

2.6. Accuracy assessment

Two accuracy assessments for each model were performed. First, an assessment of the tree crown counts and crown area calculations for each model over the entire study area (or respective study area) as well as for each 128 × 128 pixel patch used to train the respective model. Second, an intersection over union threshold was used to calculate precision, recall, and *F*1 scores for each model.

2.6.1. Crown area and crown count statistics

For the tree counts and crown area assessments, the ArcGIS Pro *select by location (ArcGIS Pro: Selection)* tool was used to locate the polygons for each model that were part of the training and validation study area. This resulted in the Count metric. Then, the *create field statistics (ArcGIS Pro: Field toolset)* tool in ArcGIS Pro was used to calculate the total polygon area in square metres (m^2) for the area of interest. This was done for the Combined, Mixed/Deciduous, and Plantation models.

Using R, we then separated the hand-delineated and model results for each 128 × 128 pixel patch. This was done by using the *st_join* function from the sf package (Pebesma et al. 2024) in R to find the intersection of the polygons for each of the 214 128 × 128 pixel

Table 3. Descriptions of the four types of model result outputs used to calculate the F1 score.

Model Result	Description
True Positive	The model predicted that there was a tree, there was a tree
False Positive	The model predicted that there was a tree, there was no tree
False Negative	The model predicted that there was no tree, there was a tree
True Negative	The model predicted that there was no tree, there was no tree

patches. A shapefile was exported for each file in the 128×128 pixel patch. We then created scatter plots of the hand-delineated outputs versus the model outputs for each of the 128×128 pixel patches.

2.6.2. Pixel level assessment of object detection

Intersection over union (IoU) has been used heavily in the literature for CNN image segmentation accuracy assessment (Martins et al. 2021; Rahman and Wang 2016; Wagner et al. 2020; Zhao et al. 2023). IoU detects the overlap between validation and the predicted segmentation of an object and gives the overlap a score from 0 to 1. The score, where 1 is considered 100% overlap, helps determine whether a predicted tree crown is a true positive or a false positive (Table 3). The *Compute Accuracy for Object Detection* tool in ArcGIS Pro *(ArcGIS Pro: Deep Learning toolset)* uses an IoU threshold of 0.25 to produce an accuracy report for all polygon intersections between the ground truth polygons and the model predictions. This accuracy report included a Precision, Recall, and *F*1 score *(ArcGIS Pro: Deep Learning toolset concepts)*.

An *F*1 score (Equation 3) was calculated using Precision (ratio of true positives to total positives, Equation 1) and Recall (ratio of true positives to the total number of objects, Equation 2), where:

$$Precision = (True\ Positive)/(True\ Positive + False\ Positive) \tag{1}$$

$$Recall = (True\ Positive)/(True\ Positive + False\ Negative) \tag{2}$$

$$F1\ Score = (Precision\ x\ Recall)/[(Precision + Recall)/2] \tag{3}$$

3. Results

The percentage for individual stem count and crown area for all three models can be found in Figure 6. This figure shows the over and under predictions from the various models as well as for the plantation ages and forest cover types, when applicable. The Combined model has several instances of over-predicting tree counts (meaning the model predicted more trees than were hand-delineated), while the Plantation model had a few instances of over-predicting the total crown area. Figure 6 also demonstrates the poor performance for both the Combined model and Plantation model for the 4–7-year-old age group. There is also a conflict of model counts and crown area from the Combined model in the 16–19-year-old plantations. Further inspection shows that the model detected most of the trees but not the entirety of the crown

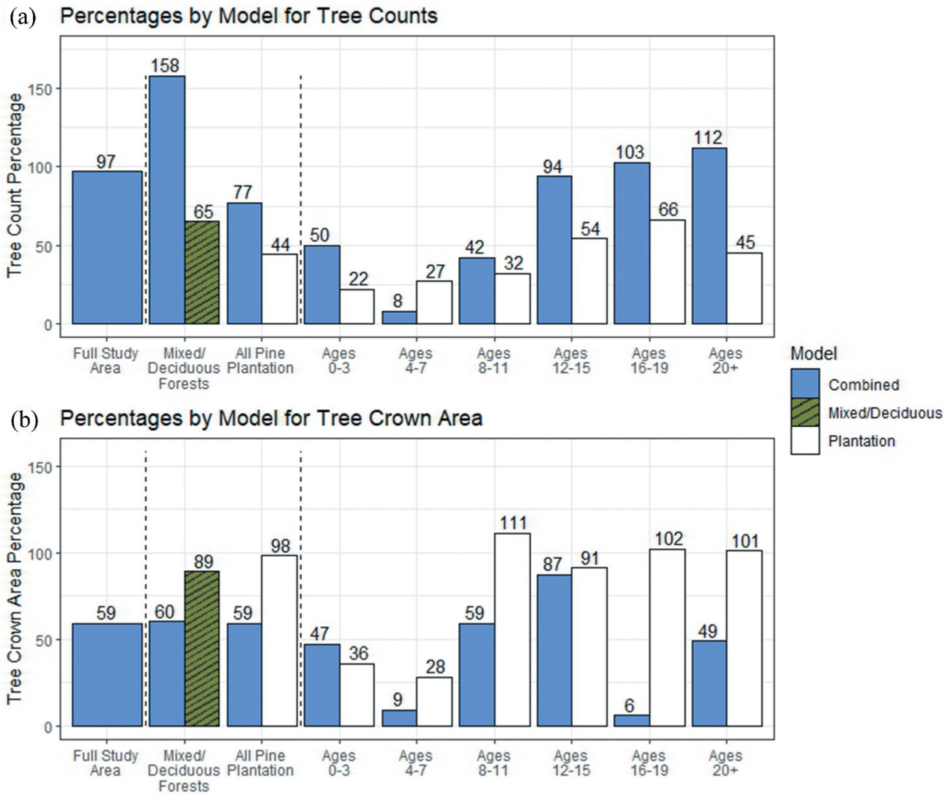

Figure 6. Counts (a) and crown area (b) percentages for the model predictions by forest types/age group for each model. Count refers to the number of individual segments identified. The percentage for the crown counts and area calculations were calculated using the following equation (model / hand-delineated) *100. The full study area percentage includes all the hand-delineated crowns and all the Combined model outputs. The percentages within the dotted lines were separated by their respective training areas. The percentages to the right of the dotted lines were separated by the respective plantation age group.

which severely diminished the total predicted crown area. Specific counts and area calculations can be found in Appendix.

Scatter plots of crown counts and crown area measurements for each of the 128×128 pixel patch versus the hand-delineations can be found in Figures 7, 8, and 9 for the Combined model, the Mixed/Deciduous model, and the Plantation model, respectively. In Figure 7 we see that the model tends to over predict in crown counts for the mixed/deciduous forests and under predict in crown area for both forest types. Although the Mixed/Deciduous model shows an underprediction of counts for most of the 128×128 patches, Figure 8(b) shows that the model does well at calculating area for all patches as the scatter plot follows the 1-to-1 line. The Plantation model shows an interesting result for the counts in Figure 9(a) where only one plot over predicted tree counts and the rest were underpredicted, especially in the younger age groups. The Plantation model area calculations do

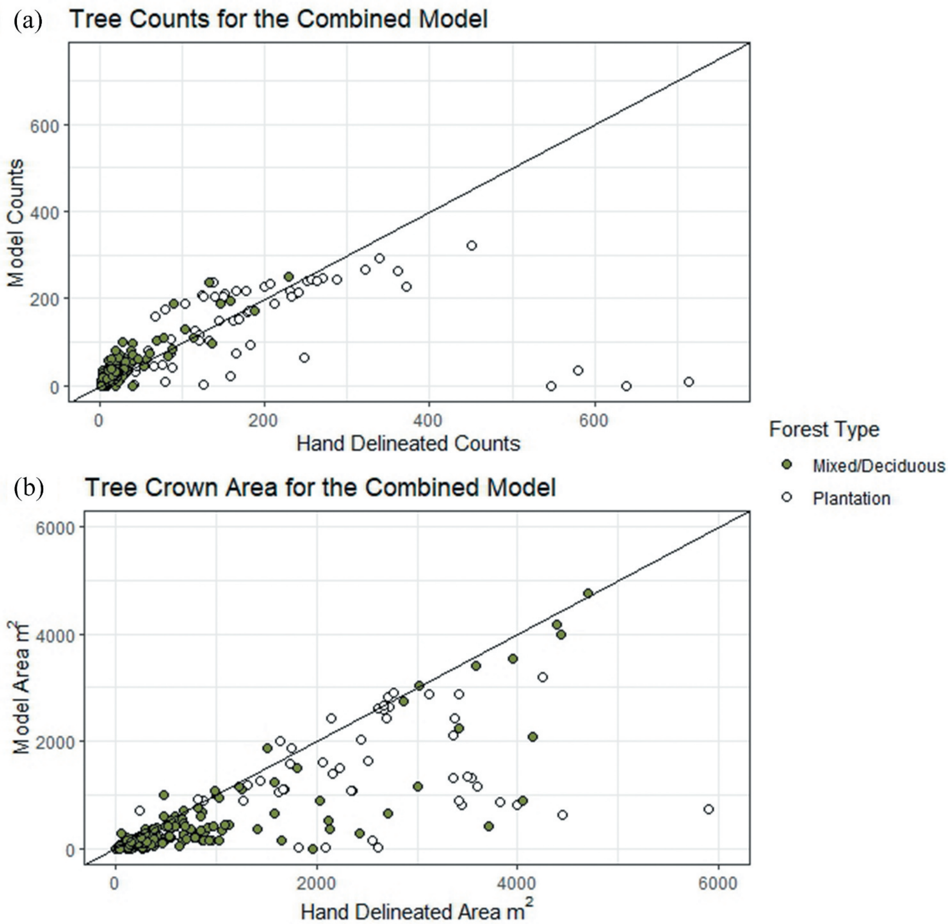

Figure 7. Counts (a) and crown area (m^2 (b) predictions for each 128 × 128 pixel patch for the Combined model. All 214 128 × 128 pixel patches were included. The 1-to-1 line is shown in black.

appear to be closely related to the 1-to-1 line but with a larger spread than the Mixed/Deciduous model for area.

3.1. Combined model

The Combined model did not perform well over the full study area after one run of training data. Although the stem count and crown area percentages sound promising (97% and 61% respectively), the model's *F*1 score was only 0.40 (Table 4). The model only correctly predicted the presence of a tree 45% of the time (precision) and only detected 36% of the trees in the image (recall). However, the model performance varied greatly over the secondary study areas with continued low performance in the mixed/deciduous forests but increased performance by age in the pine plantations. Precision, recall, and *F*1 scores for the overall model across the different study areas can be found in Table 4.

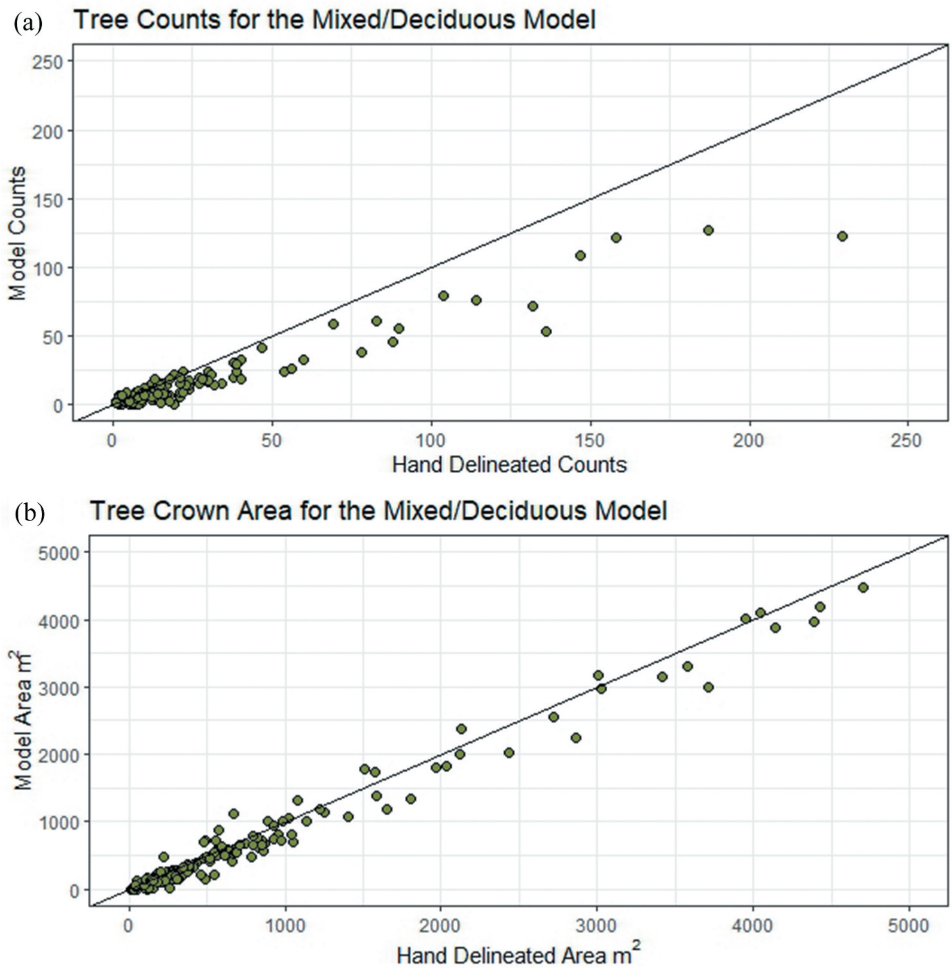

Figure 8. Counts (a) and crown area (m^2) (b) predictions for each 128 × 128 pixel patch for the mixed/deciduous model. Only the 156 mixed/deciduous forest cover patches were included. The 1-to-1 line is shown in black.

In the mixed/deciduous forest study area, the Combined model had varied success across different forest cover densities and crown size (Figure 10). In Figure 10(a,b), there is some visual evidence of crown detection but very poor crown area segmentation. Then in Figure 10(c), the model captured almost all the crowns and crown area in the image, but many of the tree crowns were not individually segmented from one another. The crown counts for this study area show an overestimation of trees by the model but only half of the amount of crown area. While recalling 50% of the trees in the study area is better than the overall model performance, only 30% of the predictions were correct (precision) which resulted in poorer model performance with an *F*1 score of 0.37 for the mixed/deciduous forests (Table 4).

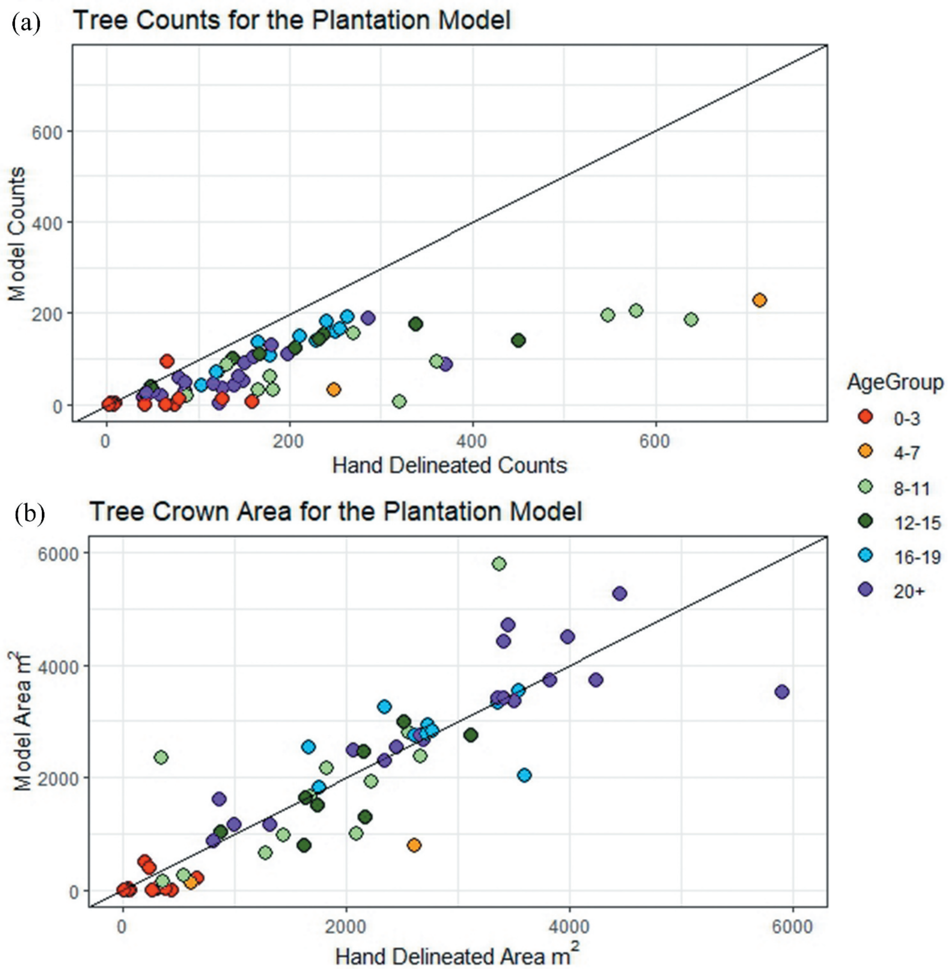

Figure 9. Counts (a) and crown area (m^2) (b) predictions for each 128 × 128 pixel patch for the Plantation model. Only the 57 plantation patches were included. The 1-to-1 line is shown in black.

Table 4. Precision, recall, and F1 scores for the Combined model by study area along with the counts for true Positive, false Positive, and false Negative results.

Study Area Type/Age Groups	Precision	Recall	*F*1 score	# True Positive	# False Positive	# False Negative
Full Study Area	0.45	0.36	0.40	1,682	2,072	3,090
Mixed/Deciduous Forests	0.30	0.50	0.37	1,642	3,852	1,751
All Pine Plantation	0.49	0.50	0.49	6,701	6,954	7,108
Pine Plantation 0–3	0.45	0.22	0.30	154	189	532
Pine Plantation 4–7	0.46	0.04	0.07	34	40	928
Pine Plantation 8–11	0.79	0.28	0.42	886	240	2,407
Pine Plantation 12–15	0.75	0.74	0.74	1,286	421	480
Pine Plantation 16–19	0.69	0.73	0.71	1,428	653	545
Pine Plantation 20$^+$	0.45	0.53	0.48	1,303	1,605	1,231

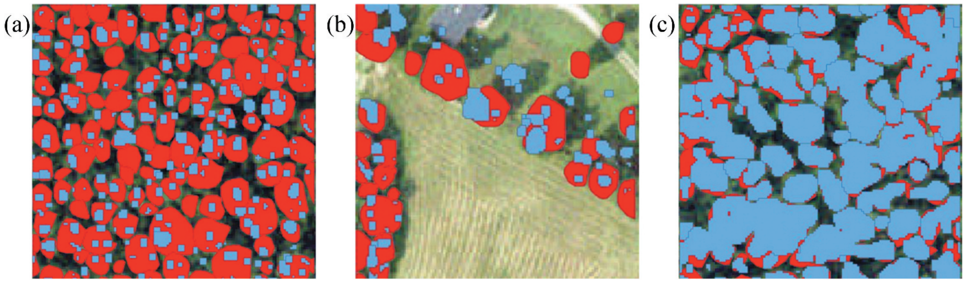

Figure 10. Examples of the Combined model's performance at different mixed/deciduous forests locations across the study area. (a) A dense, smaller crown forest canopy; (b) a border of trees around a field; (c) a dense, larger crown forest canopy. The blue polygons are the model output, and the red are the hand-drawn validation polygons.

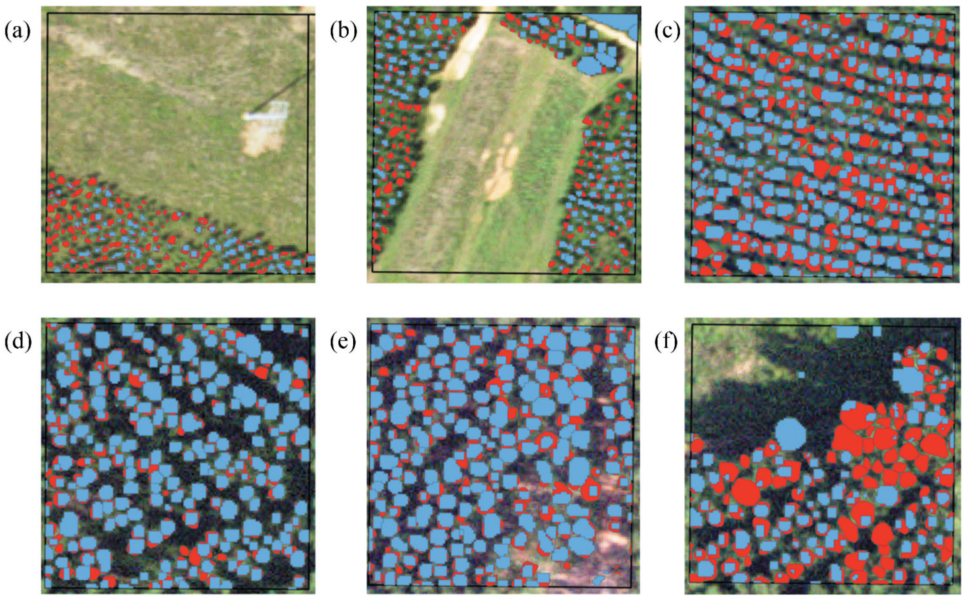

Figure 11. The Combined model performance for one sample area in each pine plantation age group. (a) 0–3-years-old; (b) 4–7-years-old; (c) 8–11-years-old; (d) 12–15-years-old; (e) 16–19-years-old; (f) 20 + years-old. The red polygons are the hand-delineated crowns, and the blue polygons are the model output.

Within the plantation secondary study area, there were some unexpected results from the Combined model's performances across the age groups. The model predicted crowns from the 12–15-year-old age group with 95% counted and 91% crown area captured, but the 16–19 and 20^{+}-year-old age groups were over-predicted (Figure 6). The model also only counted 8% of the 4–7-year-old crowns which is reflected by the poor precision, recall, and *F*1 scores. The model performed best in the 12–15-year-old stands with a precision, recall, and *F*1 score of 0.75, 0.74, and 0.74, respectively. These accuracy calculations can be visualized in Figure 11 where less of the hand-delineated polygons

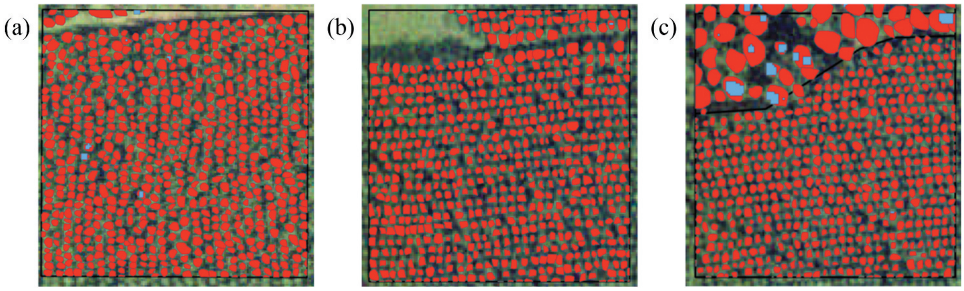

Figure 12. The three young pine plantations that had very poor Combined model performance. (a) 5-years-old; (b) and (c) 9-years-old. The red polygons are the hand-delineated crowns, and the blue polygons are the Combined model output.

(red) are visible as the pine plantation age increases, with almost all the trees located, and area calculated, in the 12–15-year-old plantation (Figure 11(d)).

Of the fifty-eight 128×128 pixel pine plantation patches, there were three that performed very poorly and stood out compared to the others in their age group: one 5-year-old stand and two 9-year-old stands (Figure 12). There were 1,901 hand-delineated crowns across these three locations with a total crown area of 6,426 m^2. The Combined model produced 13 polygon segments in this same area with a total crown area of only 25 m^2. Although these three patches do impact the crown count and crown area metrics for the whole study area and their respective age classes, they do not have an impact on the Precision, Recall, and $F1$ scores because of the IoU threshold. Based on the model outputs in the imagery and the results from Figure 4 and Table 4, there is a clear distinction between the younger and older age groups. Specifically, the worst performing age group being ages 4–7 which had the lowest $F1$ score. It is important to note that the 4–7-year-old age group had fewer samples than the other five age groups. These plantations were harder to locate, but issues with model performance are also present in the 0–3 and 8–11-year-old pine plantations so we believe that the small sample size may only partially impact the model training. Ages 12–15 and 16–19 performed the best in the model with $F1$ scores of 0.74 and 0.71, respectively.

3.2. Mixed/Deciduous model

The Mixed/Deciduous model performed much better at delineating mixed/deciduous forest tree crowns compared to the Combined model with an $F1$ score of 0.64 (Table 5). This model did undercount trees compared to the hand delineations; however, the crown area calculations were significantly better than the Combined model (Figure 6). This model was able to correctly predict the presence of trees 76% of the time (precision) and detected over half of the trees in the images (recall).

Visual inspection of the Mixed/Deciduous model output also shows its stronger performance (Figure 13) compared to the Combined model results in Figure 10. In general, the Mixed/Deciduous model significantly improved the U-Net-id's ability to identify and capture the mixed/deciduous forests canopy, especially in crown area calculations.

Table 5. Precision, recall, and F1 scores for the mixed/deciduous model along with the counts for true Positive, false Positive, and false Negative results.

Study Area Type	Precision	Recall	*F*1 score	# True Positive	# False Positive	# False Negative
Mixed/Deciduous Forests	0.76	0.55	0.64	1,702	540	1,570

Figure 13. Mixed/Deciduous model application to the same sample of trees from Figure 7 where red is the hand delineated crowns and green is the mixed/deciduous model segmentations.

3.3. Plantation model

The Plantation model performed better than the Combined model in some areas and worse in others, especially in the younger pine plantations (Table 6). This model did get better precision, recall, and *F*1 scores for the 4–7-year-old plantations, which were the worst performing for the Combined model, but it did worse in the 12–15-year-old, which were the best performing in the Combined model. Overall, the precision scores increased for all the age groups except for the 0–3-year-olds, but the recall and *F*1 scores had some age groups increase, some decrease, and one stay the same with the Plantation model. Hence, this model does well at detecting a full tree crown but is often still missing many of the crowns in the image and in some cases missing more than the Combined model did.

Visual inspection of the Plantation model over the same areas as in Figure 11 demonstrates how many more crowns are missing than before (Figure 14). Ages 0–3 (*a*) and 4–7 (*b*) show very little detection compared to the results from the Combined model. Overall, the Plantation model improved in some areas but diminished in others, especially the younger plantations which were already not performing well in the Combined model.

Table 6. Precision, recall, and F1 scores for the Plantation model along with the counts for true Positive, false Positive, and false Negative results.

Study Area Age Group	Precision	Recall	*F*1 score	# True Positive	# False Positive	# False Negative
All Pine Plantation	0.81	0.39	0.52	3,820	889	6,489
Pine Plantation 0–3	0.20	0.04	0.07	30	121	652
Pine Plantation 4–7	0.63	0.18	0.28	164	97	789
Pine Plantation 8–11	0.83	0.28	0.42	703	143	1,908
Pine Plantation 12–15	0.82	0.47	0.59	805	180	970
Pine Plantation 16–19	0.88	0.64	0.74	1,168	155	714
Pine Plantation 20^+	0.82	0.43	0.56	960	206	1,477

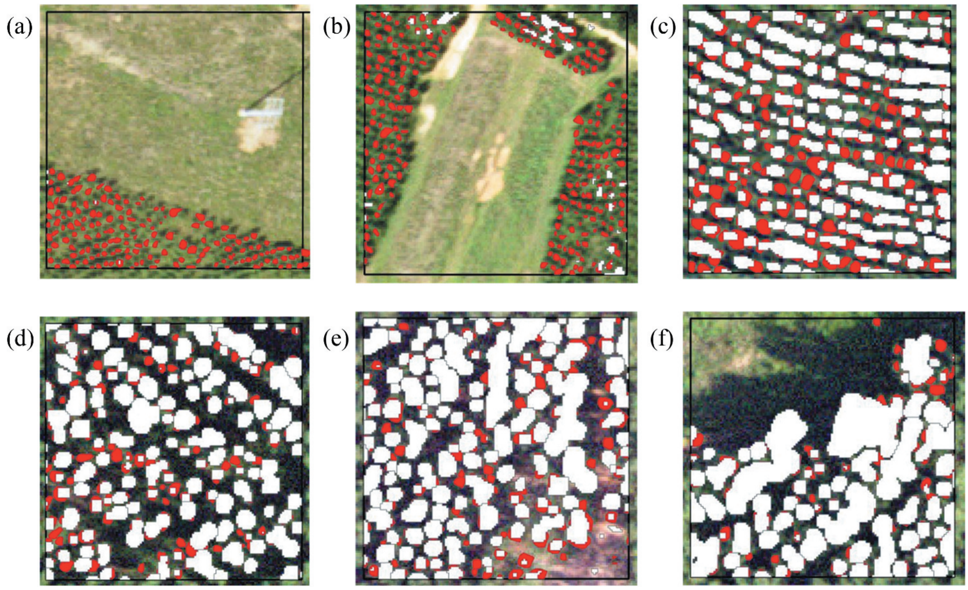

Figure 14. The Plantation model performance for the same sample areas as in Figure 6. (a) 0–3-years-old; (b) 4–7-years-old; (c) 8–11-years-old; (d) 12–15-years-old; (e) 16–19-years-old; (f) 20+ years-old. The red polygons are the hand-delineated crowns, and the white polygons are the model output.

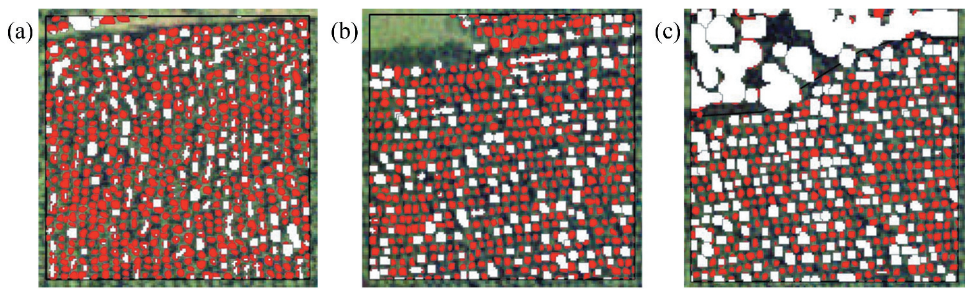

Figure 15. The three young pine plantations with the Plantation model, which had better model performance compared to the Combined model from Figure 9. (a) 5-years-old; (b) and (c) 9-years-old. The red polygons are the hand-delineated crowns, and the white polygons are the model output.

One improvement from the Plantation model was the enhanced detection of the three outlier plantations from the Combined model (Figure 15). This figure demonstrates where the Planation model improved by capturing some of the plantations that were almost completely missed by the Combined model in Figure 7.

One new outlier that appeared with the Plantation model occurred in the 8–11-year-old age group where almost the entire patch was segmented as tree crowns (Figure 16). This is the only occurrence of a model result like this in the study area. This patch had 321 hand-delineated crowns and 3,371 m^2 crown area calculated by the delineator. The Plantation model segmented five crowns and 5,795 m^2 crown area here.

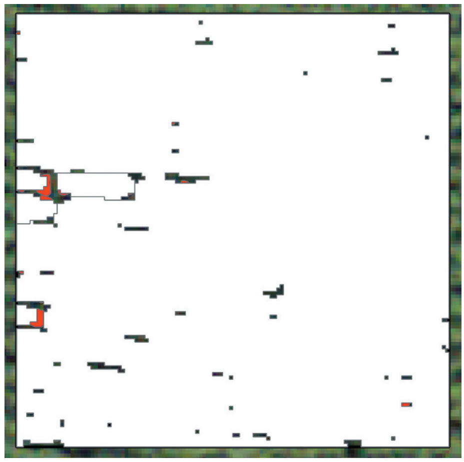

Figure 16. An outlier patch from the Plantation model in the 8–11-year-old age group. The slightly visible red polygons are the hand-delineated crowns, the white is the model result.

4. Discussion

Mapping and measuring individual trees using imagery and machine learning allows researchers to monitor the health of the world's forests. These measurements can provide information such as tree count and crown area which can inform researchers about tree growth. However, there are two challenges often associated with this information: data can be costly, large, or even unavailable to the public, and many of the delineation models used have demonstrated issues with detecting stand-alone trees, dense tree canopies, and tree cover less than 10% per acre (USDA definition). The data and methods used in this work tackled both challenges.

First, the NAIP imagery used in this study, although not a global coverage dataset, is freely available to the public and is routinely collected across the continental U.S. (USDA-FPAC-BC-GEO 2023). Imagery such as this provides researchers with a means to study the earth's surface at high resolution over multiple years which can be difficult to collect. Many satellite data products are freely available such as Sentinel and Landsat imagery, however, their resolution is too coarse for individual tree crown detection. There are groups like Planet Labs that have satellites that collect imagery at higher resolutions, however, these are not publicly available for use unless released by the company. The NAIP imagery is free, publicly available, and is routinely collected across the U.S. at sub metre resolution.

Second, the CNN model used in this study has shown to not only capture tree crowns in forest canopies but also capture those trees outside of forests that often are not accounted for in larger forest cover assessments. Within machine and deep learning, there are multiple segmentation options including semantic segmentation, class-specific segmentation, and instance segmentation (Ghosh et al. 2019). The U-Net-id model is a combination of the U-Net semantic segmentation model and an instance segmentation architecture. The U-Net architecture that our model is based off has been able to delineate individual tree crowns across forests and non-forest areas, quantifying canopy coverage that has been previously missed by others (Brandt et al. 2020). The U-Net model was designed in 2015 by Ronneberger, Fischer, and Brox (2015) to adapt to the large number of training samples required in deep learning by augmenting training data so that it may be used more efficiently. The augmentation of the training data has allowed researchers to use this model with fewer training samples than other deep learning models with positive results, especially in individual tree crown detection (Freudenberg et al. 2019; Jodas et al. 2023; Korznikov et al. 2021; Martins et al. 2021; Rhinane et al. 2021).

An important factor in deep learning and neural network models is the large quantity of data expected to train the model (Munappy et al. 2019). The more categories or area covered typically correlate with the number of training data required. However, in this study, the Combined model performed poorly in the mixed/deciduous forest when that model was trained over the entire study area with a larger distribution of canopy cover types. This result is particularly striking when compared with the Mixed/Deciduous model, which, despite having only a quarter of the training samples as the Combined model, performed significantly better over the mixed/deciduous forest study area, with the F1 score nearly doubling. This demonstrates that the quantity of training samples does not necessarily have as significant of an impact on the model as does the training data quality. Based on these results, researchers should consider focusing on localized and representative training data for model application. Lowering the number of training samples will also allow researchers to limit their data storage and computing power, which is an increasing problem as deep learning grows (Thompson et al. 2020).

Numerous research articles utilize machine learning and high-resolution imagery for forest monitoring, however, not many have attempted these techniques in tightly packed mixed/deciduous forest canopies. Previous work has utilized the U-Net CNN model with high-resolution imagery in arid regions for individual tree crown delineation (Brandt et al. 2020). Mugabowindekwe et al. (2023) mapped crowns across a variety of forest cover types in Africa using high-resolution imagery and the U-Net model, but the authors acknowledge that the tightly packed forest canopies were challenging to hand delineate for training. Other studies like Basu et al. (2015) and Weinstein et al. (2020) have used different forms of machine learning with high-resolution imagery to successfully map tree cover in closed canopies. S. Li et al. (2023) was able to use a U-Net adapted model with 20 cm resolution to map trees across Denmark, achieving an F1 score of 0.80 for thick deciduous canopies, but also acknowledges the difficulty in training on these forests. These previous results are reflected in our study where we had a more difficult time training on the closed mixed forest canopies as well as the model performing poorly in these locations. For a more homogeneous landscape, Rhinane et al. (2021) achieved a detection accuracy of 97% using a U-Net architecture over an agricultural palm tree Oasis in Morocco but noted that trees with overlapping crowns were not properly segmented. Various ages of pine plantations were counted and segmented in South America using two other types of machine learning models, but tree counts were overestimated in both tested models and younger pine trees proved to be more difficult to segment than older pine trees where the crowns were overlapping (Pérez-Carrasco et al. 2022). Again, our results reflected those of earlier work, where younger pine trees were not properly captured by the models. This review of literature covers a large span of forest types and model applications, but difficulties are present in training and mapping in tightly packed forest canopies and small stand-alone trees. Therefore, there is a need for exploration of high-resolution imagery combined with the U-Net architecture for individual tree crown segmentation in dense mixed/deciduous forests and pine plantations like in the eastern United States.

One surprising result of this study is the dramatically low scores from the 0–3 nad 4–7-year-old pine plantations. In both the Combined model and Plantation model predictions these age groups exhibited very low *F*1 and Recall scores. The 0–3-year-olds decreased in performance with the more specific model, dropping in *F*1 score from 0.30 to 0.07. This was more so surprising considering every other age group either stayed the same or saw an increase in performance with the Plantation model. When revierwing the imagery and model outputs, one possible reason for this is the resolution of the imagery. Although the tree crowns are visible in the imagery and not overlapping like many of the older plantation groups, they are often only one or two pixels in size for these age groups. It is possible that in a higher resolution image the younger trees will be better detected. Another possible reason for these low scores could be the high visibility of shadows around the trees. Previous work from Ritz et al. (2022) found that pine plantation rows that displayed a lot of shadowing in NAIP imagery performed poorly with a canopy height model. The shadows could also be impacting the models ability to detect trees in our study.

Sampling for this study was difficult due to the closed crown canopies and shadows that made it challenging to discern one crown from another. The interpreter did their best to separate individual trees, but it is anticipated the edges of some crowns were missed in shadows. This is not surprising since the United States' natural eastern coast forests are made up of thick, closed canopies with relatively uniform height and colour. This makes trees tough to distinguish from one another and limits the capabilities of visual crown segmentation for training. This is especially true in the pine plantations. Marconi et al. (2019) summarize a data science challenge where one of the tasks was using airborne data for crown segmentation and report that visually separating the tree crowns for training was difficult for most of the teams. This data science challenge also found that the crown segmentation models worked best on larger crowns, which reflects the results of Pérez-Carrasco et al. (2022) and our study results where some of the younger tree crowns in the pine plantations were almost entirely missed by all three models.

The U-Net-id model has shown here that it could work in a tightly packed mixed/deciduous forest canopy for crown area estimations, but not for crown count. More intense and specific training is required than one would normally expect for a machine learning model to achieve this measurement in these canopies. Some drawbacks to consider with these results include the bias associated with the sampling dataset location and its creation. For the dataset, we used hand-delineated crowns over areas of the NAIP imagery that were selected to train and test the model. These hand delineations were only made in areas where the interpreter felt comfortable with their ability to draw the individual crowns. The hand-drawn crowns only include those crowns that could be visually identified and separated. Therefore, there is bias associated with where the sample data came from and what tree crowns were included in the dataset. However, even though both types of bias do impact the model, the bias is consistent across all the training data. The areas that were used for training did well when compared to the validation data, but the models did not transfer this performance to the untrained sites, like the results of Graves et al. (2023).

The results of this study highlight the need to precise, local training for a model such as this for individual tree detection. The dense, hetergenous canopy of the eastern U.S. requires a more precise set of training data for the various canopy covers of the region. This work showed that the smaller dataset made up of only mixed/deciduous canopies perfromed twice as well as it did with the model that combined canopy types. Our work also demonstrated the successful use of freely available imagery for a task such as this and that opens the door for others to utilize this methodolgy for their own needs. Although this model may not have performed as well as some other studies have for individual tree detection, there is still utility in the crown area performance which can provide insight into many other forest metrics.

5. Conclusion

We investigated the influence of localized training for the U-Net-id CNN over different forest cover types in Virgina, U.S. The results of three differently trained U-Net-id models were compared over mixed/deciduous and pine plantation forests in Virginia, U.S. This work demonstrates the need for intense localized training of different forest cover types when working with U-Net CNN's. Our approach used a training dataset comprised only mixed/deciduous forests, one of only pine plantation forests, and one that combined the training datasets of the mixed/deciduous and pine plantation crowns. Our goal was to segment individual tree crowns across tightly packed mixed/deciduous forest canopies and pine plantations so that we may calculate both overstory crown counts and crown areas for the study region. We found that localized training generally improved both the Mixed/Deciduous and Plantation model performance for locating and segmenting tree crowns than the Combined model that had more training data. Of the two type/age group study areas, the Mixed/Deciduous model performed better with its localized training compared to the Plantation model, which saw both improved and diminished results across the age groups. Pine plantations make up a large portion of the southeastern U.S. and the ability to map individual tree crowns across this region would allow plantation managers to have an in-depth understanding of forest resources at the individual tree level. Given that the model does have varying degrees of success across the pine plantations, researchers should continue to investigate U-Net-id model parameters that may impact the model application, such as higher resolution or introducing other spectral bands.

Acknowledgment

We thank the Jet Propulsion Laboratory team for their assistance with the code and processing, especially F.H.W. Portion of this manuscript were prepared by U.S. Government employees on official time, it is therefore considered in the public domain and not subject to copyright.

Author contributions

Conceptualization, A.L.R., R.H.W., V.A.T., S.S., P.C.G., T.A.S., and F.H.W.; methodology, A.L.R., R.H.W., V.A.T., S.S., P.C.G., T.A.S., F.H.W.; software, A.L.R., S.S., F.H.W.; validation, A.L.R.; writing-original draft preparation, A.L.R.; writing-review and editing, A.L.R., R.H.W., V.A.T., S.S., P.C.G., T.A.S., F.H.W.; visualization, A.L.R.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Virginia Polytechnic Institute and State University.

ORCID

Alison L. Ritz http://orcid.org/0000-0002-1112-7424

Data availability statement

Data will be made available upon request.

References

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, et al. 2015. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems [Online]." Accessed January 19, 2024. https://www.tensorflow.org/.

Allaire, J., and F. Chollet. 2016. Keras: R Interface to 'Keras'. R Package Version 2.1.4. Accessed January 19, 2024. https://keras.rstudio.com.

ArcGIS Pro: Deep Learning toolset. "ArcGIS Pro 3.2. Compute Accuracy for Object Detection (Image Analyst)." *ESRI*. Accessed 08 03 2023. https://pro.arcgis.com/en/pro-app/latest/tool-reference/image-analyst/compute-accuracy-for-object-detection.htm .

ArcGIS Pro: Deep Learning toolset concepts. "ArcGIS Pro 3.2. How Compute Accuracy for Object Detection Works." *ESRI*. Accessed 09 04 2023. https://pro.arcgis.com/en/pro-app/latest/tool-reference/image-analyst/how-compute-accuracy-for-object-detection-works.htm.

ArcGIS Pro: Fields toolset. "ArcGIS Pro 3.2. Field Statistics to Table (Data Management)." *ESRI*. Accessed 18 04 2023. https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/field-statistics-to-table.htm.

ArcGIS Pro: Selection. "ArcGIS Pro 3.2. Select Features by Location." *ESRI*. Accessed 16 02 2023. https://pro.arcgis.com/en/pro-app/latest/help/mapping/navigation/select-features-by-location.htm.

Bai, M., and R. Urtasun. 2017. "Deep Watershed Transform for Instance Segmentation." *Conference on Computer Vision and Pattern Recognition,* Honolulu, HI, USA, 5221–5229. IEEE. https://doi.org/10.48550/arXiv.1611.08303

Basu, S., S. Ganguly, R. R. Nemani, S. Mukhopadhyay, G. Zhang, C. Milesi, A. Michaelis, et al. 2015. "A Semiautomated Probabilistic Framework for Tree-Cover Delineation from 1-M NAIP Imagery Using a High-Performance Computing Architecture." *IEEE Transactions on Geoscience and Remote Sensing* 53 (10): 5690–5708. https://doi.org/10.1109/TGRS.2015.2428197

Bishop, B. D., B. C. Dietterick, R. A. White, and T. B. Mastin. 2014. "Classification of Plot-Level Fire-Caused Tree Mortality in a Redwood Forest Using Digital Orthophotography and LIDAR." *Remote Sensing* 6 (3): 1954–1972. https://doi.org/10.3390/rs6031954

Brandt, M., C. J. Tucker, A. Kariryaa, K. Rasmussen, C. Abel, J. Small, J. Chave, et al. 2020. "An Unexpectedly Large Count of Trees in the West African Sahara and Sahel." *Nature* 587 (7832): 78–82. https://doi.org/10.1038/s41586-020-2824-5

Burrill, E. A., A. M. DiTommaso, J. A. Turner, S. A. Pugh, G. Christensen, C. J. Perry, L. C. Lepine, D. M. Walker, and B. L. Conkling. 2024. "The Forest Inventory and Analysis Database, FIADB user guides, volume database description (version 9.2), nationwide forest inventory (NFI)." *U.S. Department of Agriculture, Forest Service*. p.1042. https://www.fs.usda.gov/research/products/dataandtools/datasets/fia-datamart .

Cheng, Y., S. Oehmcke, M. Brandt, L. Rosenthal, A. Das, A. Vrieling, S. Saatchi, et al. 2024. "Scattered Tree Death Contributes to Substantial Forest Loss in California." *Nature Communications* 15 (1): 641. https://doi.org/10.1038/s41467-024-44991-z

Dalponte, M., L. Frizzera, H. O. Ørka, T. Gobakken, E. Næsset, and D. Gianelle. 2018. "Predicting Stem Diameters and Aboveground Biomass of Individual Trees Using Remote Sensing Data." *Ecological Indicators* 85:367–376. https://doi.org/10.1016/j.ecolind.2017.10.066
Dice, L. R. 1945. "Measures of the Amount of Ecologic Association Between Species." *Ecology* 26 (3): 297–302. https://doi.org/10.2307/1932409
EROS Center (CenterEarth Resources Observation and Science Center) Center. 2018. "National Land Cover Database." *U.S. Geological Survey*. https://www.usgs.gov/centers/eros/science/national-land-cover-database.
FIA DataMart 2.0. 2023. *USDA Forest Inventory and Analysis*. https://apps.fs.usda.gov/fia/datamart/datamart.html.
Freudenberg, M., P. Magdon, and N. Nölke. 2022. "Individual Tree Crown Delineation in High-Resolution Remote Sensing Images Based on U-Net." *Neural Computing and Applications* 34 (24): 22197–22207. https://doi.org/10.1007/s00521-022-07640-4
Freudenberg, M., N. Nölke, A. Agostini, K. Urban, F. Wörgötter, and C. Kleinn. 2019. "Large Scale Palm Tree Detection in High Resolution Satellite Images Using U-Net." *Remote Sensing* 11 (3): 312. https://doi.org/10.3390/rs11030312
Gartner, M. H., T. T. Veblen, S. Leyk, and C. A. Wessman. 2015. "Detection of Mountain Pine Beetle-Killed Ponderosa Pine in a Heterogeneous Landscape Using High-Resolution Aerial Imagery." *International Journal of Remote Sensing* 36 (21): 5353–5372. https://doi.org/10.1080/01431161.2015.1095369
Ghosh, S., N. Das, I. Das, and U. Maulik. 2019. "Understanding Deep Learning Techniques for Image Segmentation." *ACM Computing Surveys (CSUR)* 52 (4): 1–35. https://doi.org/10.1145/3329784
Graves, S. J., T. T. Caughlin, G. P. Asner, and S. A. Bohlman. 2018. "A Tree-Based Approach to Biomass Estimation from Remote Sensing Data in a Tropical Agricultural Landscape." *Remote Sensing of Environment* 218:32–43. https://doi.org/10.1016/j.rse.2018.09.009
Graves, S. J., S. Marconi, D. Stewart, I. Harmon, B. Weinstein, Y. Kanazawa, V. M. Scholl, et al. 2023. "Data Science Competition for Cross-Site Individual Tree Species Identification from Airborne Remote Sensing Data." *PeerJ* 11:e16578. https://doi.org/10.7717/peerj.16578
Han, K., Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu. 2020. "Ghostnet: More Features from Cheap Operations." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* Seattle, WA, 1580–1589. IEEE. https://doi.org/10.48550/arXiv.1911.11907
He, K., G. Gkioxari, P. Dollár, and R. Girshick. 2017 "Mask R-CNN." In *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 2961–2969. IEEE. https://doi.org/10.48550/arXiv.1703.06870
He, K., X. Zhang, S. Ren, and J. Sun. 2016. "Deep Residual Learning for Image Recognition." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 770–778. IEEE. https://doi.org/10.48550/arXiv.1512.03385
Huang, G., Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. 2017. "Densely Connected Convolutional Networks." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 4700–4708. IEEE. https://doi.org/10.48550/arXiv.1608.06993
Hulet, A., B. A. Roundy, S. L. Petersen, S. C. Bunting, R. R. Jensen, and D. B. Roundy. 2014. "Utilizing National Agriculture Imagery Program Data to Estimate Tree Cover and Biomass of Piñon and Juniper Woodlands." *Rangeland Ecology & Management* 67 (5): 563–572. https://doi.org/10.2111/REM-D-13-00044.1
Jodas, D. S., G. D. N. Velasco, R. A. de Lima, A. R. Machado, and J. P. Papa. 2023. "Deep Learning Semantic Segmentation Models for Detecting the Tree Crown Foliage." *Visigrapp (4: Visapp)*: 143–150. https://doi.org/10.5220/0011604600003417
Kattenborn, T., J. Leitloff, F. Schiefer, and S. Hinz. 2021. "Review on Convolutional Neural Networks (CNN) in Vegetation Remote Sensing." *ISPRS Journal of Photogrammetry and Remote Sensing* 173:24–49. https://doi.org/10.1016/j.isprsjprs.2020.12.010
Keefe, R. F., E. G. Zimbelman, and G. Picchi. 2022. "Use of Individual Tree and Product Level Data to Improve Operational Forestry." *Current Forestry Reports* 8 (2): 148–165. https://doi.org/10.1007/s40725-022-00160-3

Korznikov, K. A., D. E. Kislov, J. Altman, J. Doležal, A. S. Vozmishcheva, and P. V. Krestov. 2021. "Using U-Net-Like Deep Convolutional Neural Networks for Precise Tree Recognition in Very High Resolution RGB (Red, Green, Blue) Satellite Images." *Forests* 12 (1): 66. https://doi.org/10.3390/f12010066

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2017. "ImageNet Classification with Deep Convolutional Neural Networks." *Communications of the ACM* 60 (6): 84–90. https://doi.org/10.1145/3065386

Li, S., M. Brandt, R. Fensholt, A. Kariryaa, C. Igel, F. Gieseke, T. Nord-Larsen, et al. 2023. "Deep Learning Enables Image-Based Tree Counting, Crown Segmentation, and Height Prediction at National Scale." *PNAS Nexus* 2 (4): 076. https://doi.org/10.1093/pnasnexus/pgad076

Li, Z., F. Liu, W. Yang, S. Peng, and J. Zhou. 2021. "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects." *IEEE Transactions on Neural Networks and Learning Systems* 33 (12): 6999–7019. https://doi.org/10.1109/TNNLS.2021.3084827

Marconi, S., S. J. Graves, D. Gong, M. S. Nia, M. Le Bras, B. J. Dorr, P. Fontana, et al. 2019. "A Data Science Challenge for Converting Airborne Remote Sensing Data into Ecological Information." *PeerJ* 6:e5843. https://doi.org/10.7717/peerj.5843

Martins, J. A. C., K. Nogueira, L. P. Osco, F. D. G. Gomes, D. E. G. Furuya, W. N. Gonçalves, D. A. Sant'ana, et al. 2021. "Semantic Segmentation of Tree-Canopy in Urban Environment with Pixel-Wise Deep Learning." *Remote Sensing* 13 (16): 3054. https://doi.org/10.3390/rs13163054

Meneguzzo, D. M., G. C. Liknes, and M. D. Nelson. 2013. "Mapping Trees Outside Forests Using High-Resolution Aerial Imagery: A Comparison of Pixel-and Object-Based Classification Approaches." *Environmental Monitoring and Assessment* 185 (8): 6261–6275. https://doi.org/10.1007/s10661-012-3022-1

Mugabowindekwe, M., M. Brandt, J. Chave, F. Reiner, D. L. Skole, A. Kariryaa, C. Igel, et al. 2023. "Nation-Wide Mapping of Tree-Level Aboveground Carbon Stocks in Rwanda." *Nature Climate Change* 13 (1): 91–97. https://doi.org/10.1038/s41558-022-01544-w

Munappy, A., J. Bosch, H. H. Olsson, A. Arpteg, and B. Brinne. 2019. "Data Management Challenges for Deep Learning." In *2019 45th Euromicro Conference on Software Engineering and Advanced Applications*, Kallithea, Greece, 140–147. IEEE. https://doi.org/10.1109/SEAA.2019.00030

Pebesma, E., R. Bivand, E. Racine, M. Sumner, I. Cook, T. Keitt, R. Lovelace, et al. 2024. Simple Featured for R. CRAN. Version 1.0-18. https://github.com/r-spatial/sf.

Pérez-Carrasco, M., B. Karelovic, R. Molina, R. Saavedra, P. Cerulo, and G. Cabrera-Vives. 2022. "Precision Silviculture: Use of UAVs and Comparison of Deep Learning Models for the Identification and Segmentation of Tree Crowns in Pine Crops." *International Journal of Digital Earth* 15 (1): 2223–2238. https://doi.org/10.1080/17538947.2022.2152882

Pont, D., M. O. Kimberley, R. K. Brownlie, C. O. Sabatia, and M. S. Watt. 2015. "Calibrated Tree Counting on Remotely Sensed Images of Planted Forests." *International Journal of Remote Sensing* 36 (15): 3819–3836. https://doi.org/10.1080/01431161.2015.1054048/

Rahman, M. A., and Y. Wang. 2016. "Optimizing Intersection-over-Union in Deep Neural Networks for Image Segmentation." In *Advances in Visual Computing,* Las Vegas, NV, USA,234–244. https://doi.org/10.1007/978-3-319-50835-1_22

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Redmon, J., and Farhadi A. 2018. "YOLOv3: An Incremental Improvement" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 7794–7803. IEEE.

Reiner, F., M. Brandt, X. Tong, D. Skole, A. Kariryaa, P. Ciais, A. Davies, et al. 2023. "More Than One Quarter of Africa's Tree Cover Is Found Outside Areas Previously Classified as Forest." *Nature Communications* 14 (1): 2258. https://doi.org/10.1038/s41467-023-37880-4

Rhinane, H., A. Bannari, M. Maanan, and N. Aderdour. 2021. "Palm Trees Crown Detection and Delineation from Very High Spatial Resolution Images Using Deep Neural Network (U-Net)." In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Brussels, Belgium , 6516–6519. IEEE. https://doi.org/10.1080/01431161.2015.1054048

Ritz, A. L., V. A. Thomas, R. H. Wynne, P. C. Green, T. A. Schroeder, T. J. Albaugh, H. E. Burkhart, et al. 2022. "Assessing the Utility of NAIP Digital Aerial Photogrammetric Point Clouds for Estimating Canopy Height of Managed Loblolly Pine Plantations in the Southeastern United States."

International Journal of Applied Earth Observation and Geoinformation 113:103012. https://doi.org/10.1016/j.jag.2022.103012

Ronneberger, O., P. Fischer, and T. Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, 234–241. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

Schiefer, F., T. Kattenborn, A. Frick, J. Frey, P. Schall, B. Koch, and S. Schmidtlein. 2020. "Mapping Forest Tree Species in High Resolution UAV-Based RGB-Imagery by Means of Convolutional Neural Networks." *ISPRS Journal of Photogrammetry and Remote Sensing* 170:205–215. https://doi.org/10.1016/j.isprsjprs.2020.10.015

Shorten, C., and T. M. Khoshgoftaar. 2019. "A Survey on Image Data Augmentation for Deep Learning." *Journal of Big Data* 6 (1): 1–48. https://doi.org/10.1186/s40537-019-0197-0

Surdex Corporation. 2021. "Virginia_UTM_17_DSM/Virginia_UTM_18_DSM. [metadata] *Surdex Corporation*, *Chesterfield*, *MO*."

Thompson, N. C., K. Greenewald, K. Lee, and G. F. Manso. 2020. "The Computational Limits of Deep Learning." *ArXiv Preprint ArXiv:2007.05558*. https://doi.org/10.48550/arXiv.2007.05558

Trumbore, S., P. Brando, and H. Hartmann. 2015. "Forest Health and Global Change." *Science* 349 (6250): 814–818. https://doi.org/10.1126/science.aac6759

USDA. 2013. "National Agricultural Imagery Program (NAIP): Information Sheet." *USGS EROS Archive-Aerial Photography-National Agricultural Imagery Program (NAIP)*. https://www.fsa.usda.gov/Internet/FSA_File/naip_info_sheet_2013.pdf .

USDA. 2017. "Four Band Digital Imagery: Information Sheet." *USGS EROS Archive-Aerial Photography-National Agricultural Imagery Program* (NAIP). https://www.fsa.usda.gov/Assets/USDA-FSA-Public/usdafiles/APFO/support-documents/pdfs/fourband_infosheet_2017-1.pdf.

USDA. 2024. "Forest Definitions Applied in Reporting: Forest Definitions." *Forest Service: U.S. Department of Agriculture*. Accessed July 8, 2024. https://research.fs.usda.gov/programs/fia/forestdefinitions .

USDA Ag Data Commons. 2022. "Landscape Change Monitoring System Conterminous United States Year of Highest Probability of Fast Loss (Image Service)." *U.S. Forest Service*. Accessed November 1, 2023. https://apps.fs.usda.gov/lcms-viewer/.

USDA-FPAC-BC-GEO. 2023. "National Agriculture Imagery Program (NAIP) Information Sheet 2023." *United States Department of Agriculture*. https://www.arcgis.com/home/item.html?id=f2c24dbe58ac4c58a9eadaba89a5b765.

Vuola, A. O., S. U. Akram, and J. Kannala. 2019 "Mask-RCNN and U-Net Ensembled for Nuclei Segmentation." In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Venice, Italy, 208–212. IEEE. https://doi.org/10.1109/ISBI.2019.8759574

Wagner, F. H. 2020. *Model Package from "U-Net-Id, an Instance Segmentation Model for Building Extraction from Satellite Images"*, 2.0. Zenodo. https://doi.org/10.5281/zenodo.3926831

Wagner, F. H., R. Dalagnol, Y. Tarabalka, T. Y. Segantine, R. Thomé, and M. C. Hirye. 2020. "U-Net-Id, an Instance Segmentation Model for Building Extraction from Satellite Images—Case Study in the Joanópolis City, Brazil." *Remote Sensing* 12 (10): 1544. https://doi.org/10.3390/rs12101544

Weinstein, B. G., S. Marconi, S. A. Bohlman, A. Zare, and E. P. White. 2020. "Cross-Site Learning in Deep Learning RGB Tree Crown Detection." *Ecological Informatics* 56:101061. https://doi.org/10.1016/j.ecoinf.2020.101061

Weinstein, B. G., S. Marconi, S. Bohlman, A. Zare, and E. White. 2019. "Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks." *Remote Sensing* 11 (11): 1309. https://doi.org/10.3390/rs11111309

Zhao, H., J. Morgenroth, G. Pearse, and J. Schindler. 2023. "A Systematic Review of Individual Tree Crown Detection and Delineation with Convolutional Neural Networks (CNN)." *Current Forestry Reports* 9 (3): 149–170. https://doi.org/10.1007/s40725-023-00184-3

Appendix

Table A1 provides the crown count and crown area statistics for the hand-delineations and the model outputs for all three models over the whole study area and by each forest cover type and age when appropriate.

Table A1. Counts and crown area (m^2) results for the models and hand-delineations by forest types/age groups. Count refers to the number of individual segments identified.

Study Area Type/Age Group	Measurement	Hand-Delineation	Combined Model	Mixed/Deciduous Model	Plantation Model
Full Study Area	Count	14,137	13,655	–	–
	Crown Area	234,327	138,571	–	–
Mixed/Deciduous Forests	Count	3,457	5,474	2,242	–
	Crown area	114,376	68,258	102,197	–
All Pine Plantation	Count	10,680	8,181	–	4,732
	Crown Area	119,951	70,313	–	117,182
Pine Plantation 0–3	Count	682	343	–	151
	Crown Area	3,175	1,483	–	1,152
Pine Plantation 4–7	Count	962	74	–	261
	Crown Area	3,222	290	–	901
Pine Plantation 8–11	Count	2,643	1,117	–	846
	Crown Area	15,606	9,255	–	17,334
Pine Plantation 12–15	Count	1,822	1,707	–	985
	Crown Area	15,832	13,833	–	14,402
Pine Plantation 16–19	Count	1,999	2,056	–	1,323
	Crown Area	26,854	18,635	–	27,388
Pine Plantation 20^+	Count	2,572	2,884	–	1,166
	Crown Area	55,262	26,817	–	56,005