remote sensing

MDPI

Article

Individual Tree Crown Delineation Using Airborne LiDAR Data and Aerial Imagery in the Taiga–Tundra Ecotone

Yuanyuan Lin [1,2], Hui Li [1,3,*], Linhai Jing [1,3], Haifeng Ding [1,3] and Shufang Tian [4]

[1] International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China; 2001210192@email.cugb.edu.cn (Y.L.); jinglh@radi.ac.cn (L.J.); hfding@cbas.ac.cn (H.D.)
[2] China Siwei Surveying and Mapping Technology Co., Ltd., Beijing 100190, China
[3] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
[4] School of Earth Sciences and Resources, China University of Geosciences, Beijing 100083, China; sftian@cugb.edu.cn
* Correspondence: lihui@radi.ac.cn

Abstract: The circumpolar Taiga–Tundra Ecotone significantly influences the feedback mechanism of global climate change. Achieving large-scale individual tree crown (ITC) extraction in the transition zone is crucial for estimating vegetation biomass in the transition zone and studying plants' response to climate change. This study employed aerial images and airborne LiDAR data covering several typical transitional zone regions in northern Finland to explore the ITC delineation method based on deep learning. First, this study developed an improved multi-scale ITC delineation method to enable the semi-automatic assembly of the ITC sample collection. This approach led to the creation of an individual tree dataset containing over 20,000 trees in the transitional zone. Then, this study explored the ITC delineation method using the Mask R-CNN model. The accuracies of the Mask R-CNN model were compared with two traditional ITC delineation methods: the improved multi-scale ITC delineation method and the local maxima clustering method based on point cloud distribution. For trees with a height greater than 1.3 m, the Mask R-CNN model achieved an overall recall rate (A_r) of 96.60%. Compared to the two conventional ITC delineation methods, the A_r of Mask R-CNN showed an increase of 1.99 and 5.52 points in percentage, respectively, indicating that the Mask R-CNN model can significantly improve the accuracy of ITC delineation. These results highlight the potential of Mask R-CNN in extracting low trees with relatively small crowns in transitional zones using high-resolution aerial imagery and low-density airborne point cloud data for the first time.

Keywords: individual tree crown delineation; the taiga–tundra ecotone; Mask R-CNN; deep learning

check for updates

Citation: Lin, Y.; Li, H.; Jing, L.; Ding, H.; Tian, S. Individual Tree Crown Delineation Using Airborne LiDAR Data and Aerial Imagery in the Taiga–Tundra Ecotone. *Remote Sens.* **2024**, *16*, 3920. https://doi.org/10.3390/rs16213920

Academic Editor: Henning Buddenbaum

Received: 6 August 2024
Revised: 17 September 2024
Accepted: 14 October 2024
Published: 22 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The Taiga–Tundra Ecotone stretches 13,000 km in the Northern Hemisphere, spanning three major continents [1]. This critical transitional zone, marking the boundary between the southern Taiga forests and the northern treeless tundra, plays a pivotal role in sustaining water cycles, soil conservation, carbon storage, and habitat preservation [2,3]. The vegetation growth in the ecotone is susceptible to external factors such as climate change [4,5], making it a significant indicator of climatic change [6,7]. Existing research indicates that rapid climate warming in the Arctic and Alpine regions is triggering a range of responses in the ecosystem of the Taiga–Tundra Ecotone [8,9]. These include changes in vegetation phenology, prolonged growing periods of the forest, and an increase in the height and coverage of a majority of the grasses and shrubs in tundra areas, as well as the pushing of trees and shrubs towards higher latitudes and altitudes in the permafrost zone [10–13]. Concurrently, alterations in the carbon cycle and energy balance of the transitional zone ecosystem have a reciprocal impact on regional and global climate [14,15]. Monitoring the

vegetation growth within the transitional zone is crucial for quantitatively assessing the impact of climate change on the transitional zone and developing sustainable environmental management and conservation strategies.

The circumpolar Taiga–Tundra Ecotone covers a vast area at high latitudes, where vegetation types and densities exhibit transitional characteristics, with an increase in low plants towards the north. The coniferous trees that grow in this region have relatively low heights and small crowns compared to the Taiga, and forests are less dense. It is a challenging task to identify individual trees in the ecotone on a large scale using traditional field inventory methods. Currently, aerial imagery and high-resolution satellite remote sensing images allow for the observation of ITCs and show significant differences between trees and grasslands regarding spatial structure and texture. Airborne Light Detection and Ranging (LiDAR) data provide elevation information and 3D structures of tree canopies, effectively distinguishing between vegetation types such as trees, shrubs, and grasslands, thereby offering a more effective data source for the delineation of ITCs [16]. Undertaking ITC delineation in transitional zones using high-resolution remote sensing data such as aerial imagery and airborne LiDAR data not only allows for the monitoring of individual trees within forested areas but also detects many trees outside of forests.

Over the past three decades, scholars both domestically and internationally proposed various ITC delineation algorithms, such as the local maximum method, multi-scale analysis method, valley tracking method, template matching method, local ray method, region growth method, watershed segmentation method, and object-oriented techniques [17–21]. Among these methods, the methods for airborne point cloud data can be divided into two categories: methods based on canopy height models (CHMs) and point cloud-based methods. The two commonly used CHM-based ITC delineation methods developed based on CHMs are the multi-scale analysis and the watershed segmentation methods. The multi-scale segmentation method can continuously represent the diverse scales of tree crowns, reducing crown merging and under-segmentation caused by sparse scales [22,23]. For example, Chen et al. [24] utilized a marker-controlled watershed segmentation method based on CHMs for ITC delineation in sparse savanna woodlands, achieving an accuracy of 64.1%. Jing et al. [25] introduced the Multi-Scale Adaptive Segmentation (MSAS) method based on CHMs to accurately delineate approximately 69%, 65%, and 73% of tree crowns in coniferous, deciduous, and mixed forests. The MSAS method primarily relies on the brightness information of a single spectral band. Incorporating more multispectral information will be a focus of future research [26]. The point cloud-based methods for ITC delineation include clustering algorithms and region-growing methods. For example, Wu et al. [27] utilized a region-growing method based on point cloud distribution for artificial forest ITC delineation, achieving an accuracy of 83%. Zhang et al. [28] used an improved clustering algorithm with direct point cloud data for ITC delineation in an urban forest, achieving an accuracy of 93.5%. Li et al. [29] proposed a local maximum clustering approach based on point cloud distribution, achieving a recall of 86% in the ITC delineation of mixed coniferous forests in the Nevada Mountains of California. Li's method avoided information loss and errors due to data interpolation and smoothing during the CHM creation. However, this approach cannot delineate tree crown boundaries, significantly limiting the accuracy. A large amount of point cloud data increases processing time, which is not conducive to large-scale ITC delineation. Although the traditional ITC delineation methods have the advantages of a simple algorithm and low calculation amount, they need to adjust parameters manually, the parameter settings depend on expert knowledge, and the robustness of traditional methods is relatively poor, significantly reducing the efficiency and automaticity of ITC delineation [30].

The advent of deep learning (DL) technology has propelled the interpretation of remote sensing imagery towards automation and intelligence [31,32]. Convolutional Neural Networks (CNNs) represent a significant algorithm in deep learning and have been extensively applied for object detection, boasting advantages such as fast recognition speed, high accuracy, minimal feature requirements, and self-training capabilities [33]. In the archi-

tecture of CNNs, the Mask Region-based Convolutional Neural Network (Mask R-CNN) stands out [34]. Building on the traditional object detection algorithm Fast Region-based Convolutional Network (Fast R-CNN), Mask R-CNN integrates the semantic segmentation algorithm of Fully Convolutional Networks to generate a corresponding mask prediction branch. Further, Mask R-CNN introduced the Region of Interest (ROI) Align strategy, thus rectifying the pixel misalignment problem present in ROI Pooling, achieving higher accuracy while maintaining speed [35]. In recent years, deep learning has shown great potential in ITC delineation research [36–38]. Brandt et al. [39] employed U-Net for ITC delineation using sub-meter resolution satellite imagery and mapped individual trees with a crown size over 3 m^2 for a study area of 1.3 million square kilometers in West Africa. Braga et al. [40] conducted an ITC delineation study in a region of Brazil using Mask R-CNN on 0.5 m resolution WorldView-2 satellite images, achieving an accuracy of 91% and a recall of 81%. Erdem et al. [41] used Mask R-CNN to detect apricot trees from aerial imagery, achieving an accuracy of 98.7% and a recall of 99.7%, outperforming U-Net. Perez et al. [42] achieved an accuracy of 82% in the delineation of individual pine tree crowns using Mask R-CNN, which was about 10 points in percentage higher compared to the YOLO model. These studies indicated that Mask R-CNN demonstrates exceptional performance in ITC delineation.

Some studies have utilized airborne laser scanning data to detect small pioneer trees in the Taiga–Tundra Ecotone [43,44]. Most existing ITC delineation research in this ecotone utilizes high-density point cloud data [45–52]. Næsset and Nelson [45] monitored tree migration in northern Alpine transitional zones using point cloud data with a density of 7.7 pts/m^2. Thieme et al. [46] used point cloud data with densities of 6.8 and 8.5 pts/m^2 for ITC delineation in the Taiga–Tundra Ecotone, achieving an accuracy of 90% for trees higher than 1 m. Hauglin et al. [47] monitored small individual trees in the Taiga–Tundra Ecotone using multi-temporal point cloud data with densities of 8~15 pts/m^2. As point cloud density increases, ITC delineation accuracy significantly improves [48–50]. However, due to random errors such as sampling density, system errors, and forest environmental interference, the point cloud density of LiDAR data obtainable in some forest areas is limited. Methods for ITC delineation in transitional zones using airborne LiDAR data with a relatively low density still require further research. On the other hand, a single data source contains only limited feature information, which restricts the accuracy of ITC delineation and individual tree species classification. More in-depth research shows that combining high-resolution aerial imagery with airborne LiDAR data can improve accuracy [51–53]. This is mainly due to the complementary advantages of the high spatial resolution of aerial imagery and the 3D structural information provided by LiDAR data. Existing studies have primarily used traditional ITC delineation methods, such as watershed segmentation algorithms based on CHMs and clustering algorithms based on point clouds [51,53], which are difficult to apply on a large scale. There is an urgent need to explore ITC delineation methods and models based on deep learning to meet the demand for large-scale ITC extraction in the transition zone. Additionally, one of the main challenges currently facing CNNs (including Mask R-CNN) is acquiring a sufficient number of high-quality sample training sets for the training of CNNs [54,55]. Obtaining the sample sets was also a primary issue that needs to be resolved for large-scale individual tree delineation in transitional zones.

This study combined the advantages of high-resolution aerial imagery and low-density airborne LiDAR data to explore the potential of mapping ITCs based on deep learning on a large scale in the Taiga–Tundra Ecotone. The main contributions of this study are as follows: First, the feasibility of using deep learning and low-density point cloud data (point cloud density of 0.8~1.1 pts/m^2) for large-scale individual tree extraction in the transition zone was verified for the first time. Second, an improved multi-scale ITC delineation method was proposed in this work to facilitate the semi-automatic acquisition of ITC labels. Third, an individual tree sample set containing over 20,000 trees in the transitional zone was constructed and the trained Mask R-CNN model obtained by training the sample can provide data and technical support for large-scale ITC delineation within the transitional

zone. The subsequent sections of this study are organized as follows: Section 2 describes the background of the study area, data collection, and preprocessing. Section 3 details the semi-automatic construction process of the sample set, ITC delineation methods, and accuracy evaluation metrics. Section 4 presents the results and accuracy of different ITC delineation algorithms. Section 5 compares the methods and results of this study with other similar research and discusses factors affecting accuracy. Finally, Section 6 presents this study's main conclusions and explores the application prospects of ITC delineation research in transitional zones.

2. Study Area and Data

2.1. Study Area

The study area is located in the northern part of Lapland Province, Finland, within the circumpolar Taiga–Tundra Ecotone. The latitude of the study area ranges from 69°23′27″ to 70°5′45″N, and the longitude ranges from 25°51′25″ to 29°20′10″E (Figure 1). This region experiences cold winters and relatively warm summers. The snowy period persists from October to May or June of the following year [56], with a vegetation growing period of less than four months and annual precipitation ranging between 400 and 600 mm [57]. The predominant vegetation types in the study area include coniferous trees, deciduous trees, cold-resistant low shrubs, mosses, and lichens. The main tree species are Scots pine (*Pinus sylvestris* L.), Norway spruce [*Picea abies* (L.) Karst.], and the mountain birch (Betula pubescens subsp. czerepanovii Ehrh.), distributed from south to north, respectively [11,58]. The soil in the area is relatively infertile. Trees grow slowly, with relatively low heights and small crowns. The stands are sparse, and there is limited vegetation diversity.

For this study, five plots and nine training areas were selected for testing and training, respectively. The tree species predominantly include Scots pine and Norway spruce. Figure 1a shows the locations of these areas. Visual interpretation results were used as reference samples for accuracy evaluation and comparative analysis. The false color aerial orthophotos of the five plots are shown in Figure 1b–f.

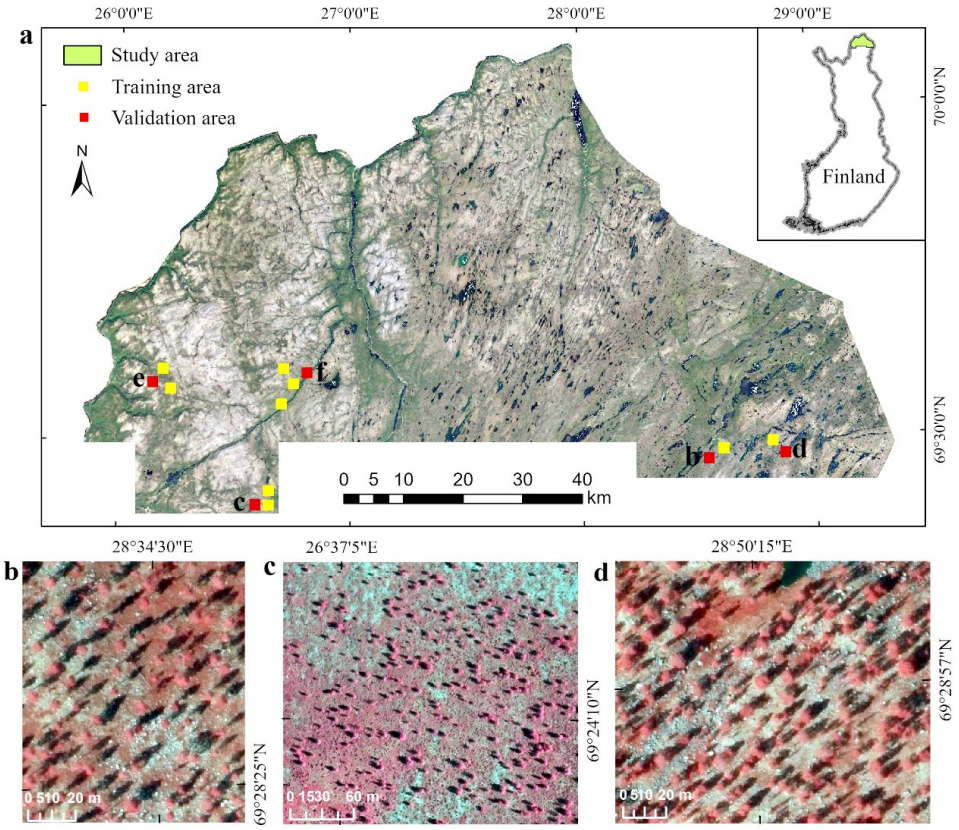


Figure 1. *Cont.*

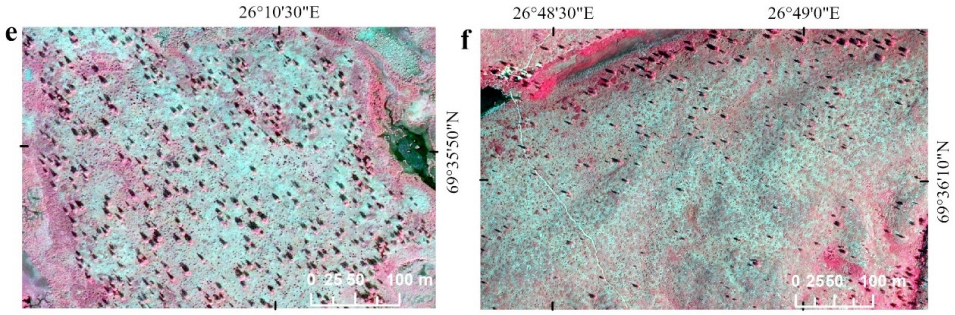


Figure 1. The location of the study area and the distribution of images used for training and validation. (**a**) The location of the study area; (**b**–**f**) the false color aerial orthophotos of the five validation areas of plot 1, plot 2, plot 3, plot 4, and plot 5, respectively (shown in a combination of near-infrared, red, and green bands).

2.2. Data Collection

2.2.1. Image Collection and Preprocessing

The aerial orthophotos and airborne LiDAR data used in this work were provided by the Finnish National Land Survey Authority (URL: https://asiointi.maanmittauslaitos.fi/karttapaikka/tiedostopalvelu (accessed on 26 September 2022)). The aerial images had 189 image slices with a spatial resolution of 0.5 m and a size of 6000 × 6000 pixels. The airborne LiDAR data corresponding to an image slice of the aerial image were recorded in four parts. The point cloud density varied from 0.8 to 1.1 pts/m^2, which is relatively low. The acquired date of the aerial images and LiDAR data ranged from 2014 to 2022. The details of the training and testing areas are listed in Table 1, which also shows the temporal variation between an aerial image and the corresponding LiDAR data. The aerial images had a positional accuracy ranging from 0.5 to 2 m, whereas the elevation accuracy of the LiDAR data was 0.15 m. The aerial images comprised four spectral bands: near-infrared, red, green, and blue. The airborne LiDAR data were acquired using a Leica ALS70-HP scanner equipped on an aircraft with flight altitudes between 2500 and 2874 m.

Table 1. Size and acquisition time of aerial imagery and airborne LiDAR data.

Area ID		Image Size (Pixels)	Acquisition Time	
			Aerial Imagery	Airborne LiDAR Data
Validation areas	Plot 1	192 × 226	2022	2015.08
	Plot 2	560 × 500	2021	2018.09
	Plot 3	300 × 250	2022	2015.08
	Plot 4	935 × 635	2014	2018.07
	Plot 5	1165 × 725	2021	2018.07
Training areas	T1	865 × 493	2022	2015.08
	T2	1806 × 1555	2022	2015.08
	T3	1715 × 1115	2022	2015.08
	T4	1389 × 1383	2022	2015.08
	T5	1650 × 1140	2014	2018.07
	T6	2100 × 2025	2014	2018.07
	T7	1695 × 1130	2021	2018.07
	T8	665 × 660	2021	2018.07
	T9	810 × 660	2021	2018.07

The CHM of the study area was derived from the airborne LiDAR data using the ENVI LiDAR module in ENVI 5.3 software. First, by preprocessing the airborne LiDAR data, ground points and non-ground points were separated. Then, the Triangulated Irregular Network (TIN) algorithm was used to interpolate the separated ground and non-ground points to generate Digital Elevation Models (DEMs) and Digital Surface Models (DSMs). Finally, a 1 m resolution CHM was produced by performing raster differencing operations

(subtracting the DTM from the DSM). Figure 2 shows the CHM of plot 2 and the corresponding individual tree 3D imagery. The initially generated CHM exhibited "holes", as shown in Figure 3a. The "holes" destroyed the integrity of the morphological structure of tree crowns (Figure 3b), severely impacting the precision of ITC delineation [59]. Therefore, smoothing filters were commonly used to optimize the CHM before ITC delineation. This study compared the effects of three filters, which are the low-pass, median, and Gaussian low-pass filters. In contrast, as shown in Figure 3c,d, the low-pass filter effectively removed "holes" while preserving canopy surface structures and tree crown edge information, maintaining high consistency with the original point cloud data. Consequently, the low-pass filter was selected for CHM optimization in this study.

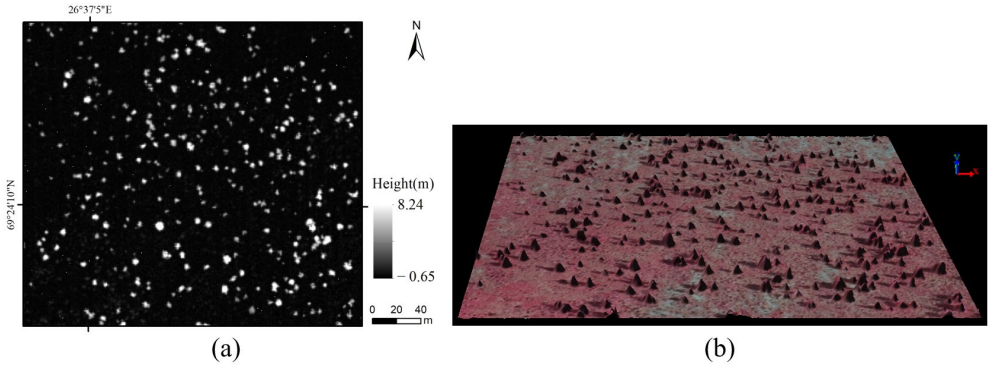


Figure 2. The CHM of plot 2 (**a**) and the corresponding false color aerial orthophoto showing individual trees in 3D (**b**).

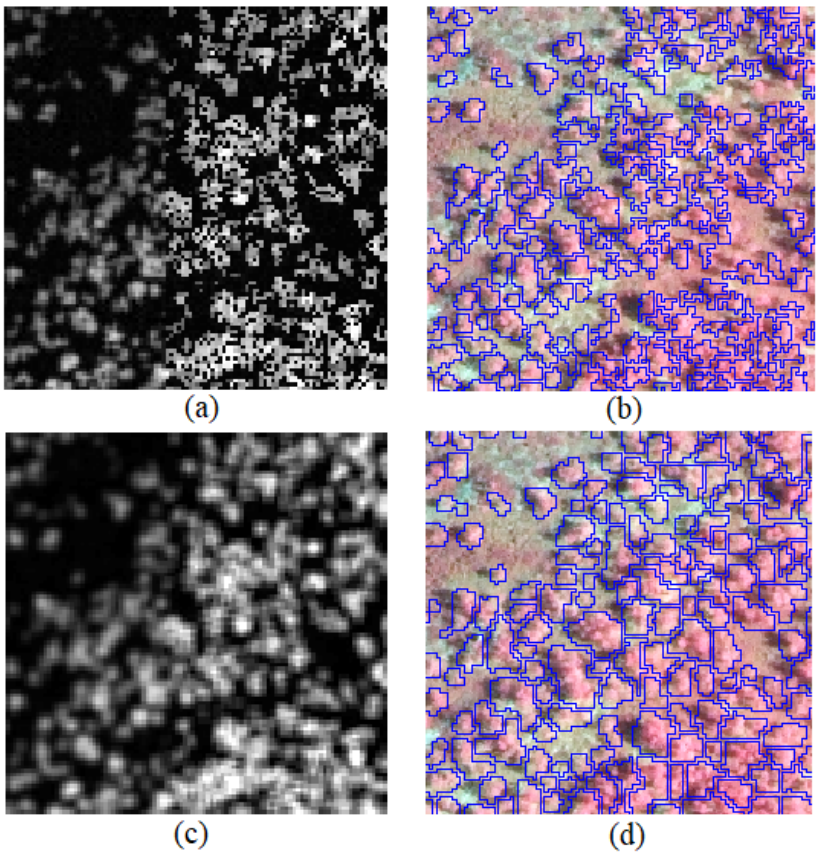


Figure 3. The CHM optimization results and ITC delineation results using different filtering methods. (**a**) The original CHM; (**b**) ITC delineation results (the polygons in blue) using the original CHM; (**c**) the CHM obtained by low-pass filtering; and (**d**) ITC delineation results (the polygons in blue) using the low-pass filtered CHM.

2.2.2. Validation Data

Using aerial imagery and CHMs, ITCs of the five plots were manually delineated to serve as reference data for assessing the accuracies of ITC delineation methods. In this study, only trees higher than 1.3 m were delineated, and only crowns with clear boundaries and accompanying shadows were outlined. A few small trees were under the shadow of adjacent tall trees and thus excluded from visual interpretation due to uncertainties. A total of 1558 tree crowns were delineated from the five plots, which have a total area of 459,186 m^2. The area of these tree crowns ranged from 3.50 to 90 m^2, and the average area was 21.94 m^2. The tree heights varied from 1.30 to 13.83 m, and the average height was 4.47 m. Table 2 shows the area, number of tree crowns, and the minimum, maximum, and average values of tree crown area and height of each plot. Figure 4 shows some of the manually delineated tree crowns.

Table 2. The area, number of tree crowns, crown area, and tree height of the five plots used for validation.

Area ID	Area (m^2)	Number of ITCs	Area of ITCs (m^2)			Height (m)		
			Minimum Value	Maximum Value	Average Value	Minimum Value	Maximum Value	Average Value
Plot 1	10,848	93	5.50	50.25	21.02	1.54	13.83	8.99
Plot 2	70,000	346	4.00	90.00	18.94	1.41	8.24	3.94
Plot 3	18,750	226	5.25	81.00	24.43	1.40	12.49	6.17
Plot 4	148,431	562	3.50	80.00	23.25	1.39	8.12	4.17
Plot 5	211,156	331	3.75	72.25	21.38	1.30	7.75	3.09
Total	459,186	1558	3.50	90.00	21.94	1.30	13.83	4.47

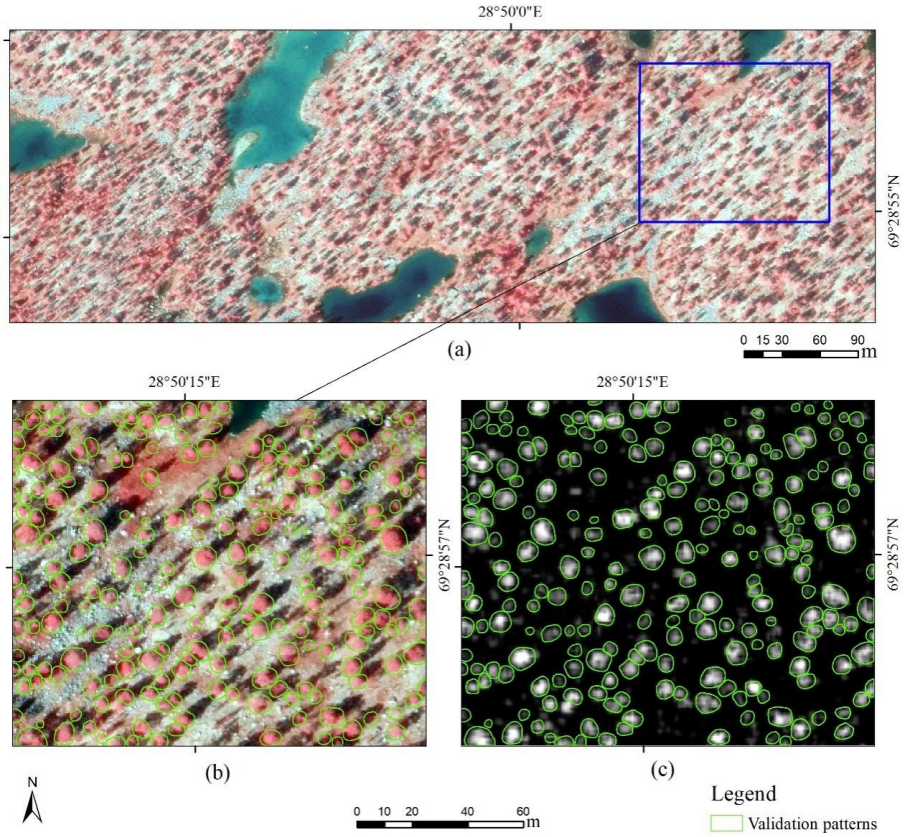


Figure 4. An example of manually delineated ITCs. (**a**) A sub-area of the study region, in which the yellow rectangular frame represents the location of plot 3; (**b**) the manually delineated tree crowns of plot 3 (overlapping the aerial imagery); and (**c**) the manually delineated tree crowns of plot 3 (overlapping the CHM).

3. Methodology

3.1. Semi-Automatic Construction of Sample Sets

3.1.1. Improved Multi-Scale Method for ITC Delineation

This study employed an improved multi-scale ITC delineation method for the semi-automatic acquisition of ITC labels using the combination of CHMs and NDVI. As an improved version of the MSAS method proposed by Jing et al. [25], the method is henceforth referred to as the improved MSAS method. The improved MSAS method made several revisions to adapt to the background and available data of the study area. First, non-vegetation areas were eliminated by applying a threshold on NDVI. Further, the Gaussian filter employed by the original method was omitted for the following reasons: (i) The tree crowns of the five plots are relatively small and low, and the main tree species are conifers with regular crown shapes (round or oval); therefore, the branches have little influence on delineation. (ii) The point cloud data used in this study have a relatively low density.

The workflow of the improved MSAS method is presented in Figure 5. The main processing steps include scale analysis, removal of non-vegetation areas, multi-scale ITC delineation algorithm, and optimization of ITC delineation results. The details of these steps are introduced in the following paragraphs.

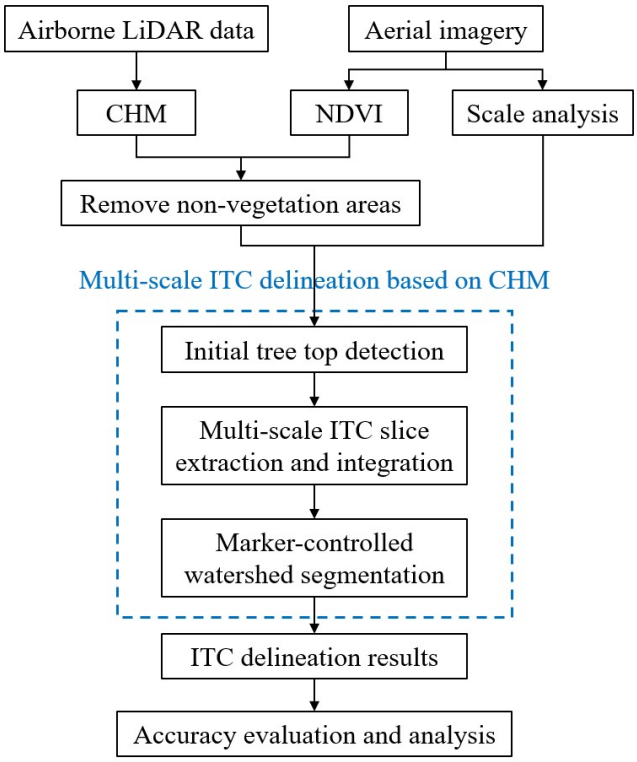


Figure 5. The flowchart of the improved MSAS method.

(1) Scale analysis

In this study, the scale analysis technique [60] was utilized to determine the dominant size of target tree crowns. First, the open morphological operation was applied to the CHM of training areas (Figure 6) [61,62]. An example of a CHM image is shown in Figure 6a, with the corresponding aerial image of the area presented in Figure 6b. A series of morphologically opened images were generated using a series of disk structural elements (SEs) with diameters ranging from 3 to 23 pixels in an increment of 2 pixels. Then, the difference image was computed for any two consecutive opened images, and the mean value of each difference image was calculated. The curve of all the mean values of the different images derived from the series of opened images is shown in Figure 6c. A local minimum occurs when there are significant differences in object sizes between two adjacent opened images.

Therefore, the curve of these mean values reveals the dominant sizes of the objects in the image. Due to the relatively small size of the tree crowns in the study area, the diameters 3, 5, and 7 pixels (1.5 m, 2.5 m, and 3.5 m) were considered as dominant tree crown sizes, representing small, medium, and large tree crowns, respectively, which were then selected as scales employed in the multi-scale ITC delineation.

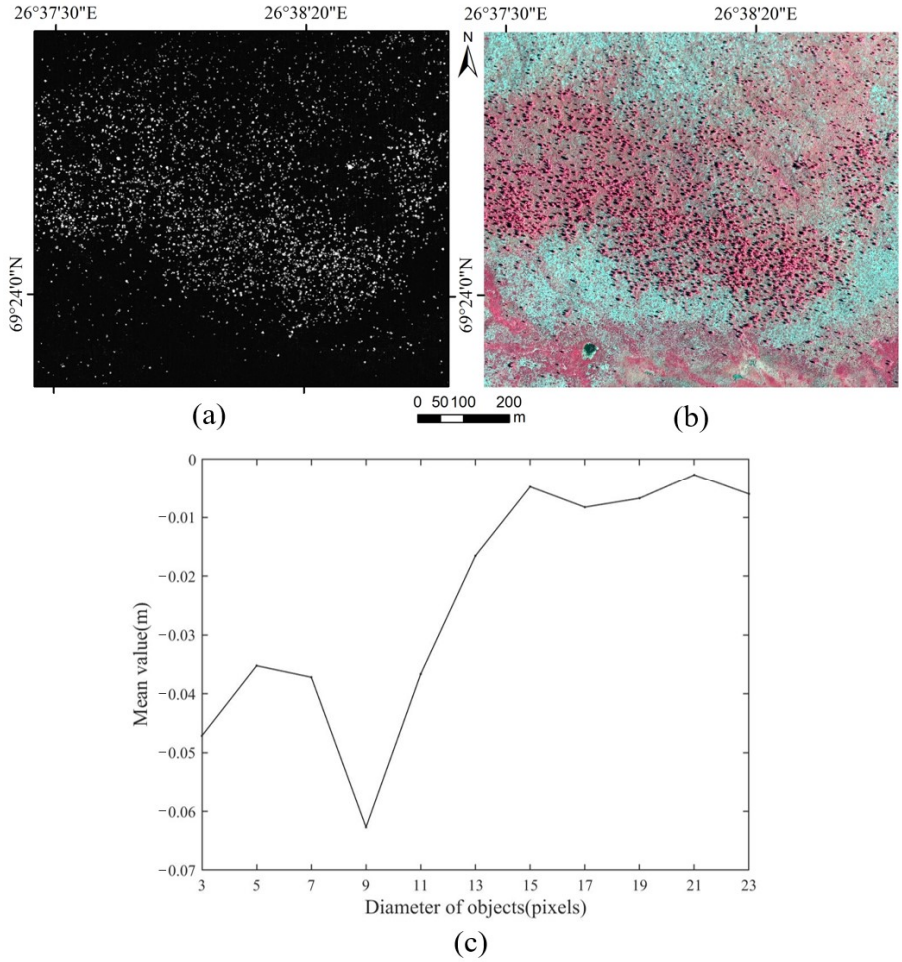


Figure 6. An example of scale analysis and the mean value of different images derived from a series of morphologically opened images of the CHM. (**a**) A sub-area of the CHM; (**b**) the corresponding false color aerial imagery of (**a**); and (**c**) the mean values of the different images derived from a series of opened images obtained from the CHM using a disk SE with a gradually increasing diameter.

(2) Remove non-vegetation areas

NDVI, CHM height values, and shadow areas were employed to reduce the impact of other image objects in this study. First, a reasonable NDVI threshold was established by analyzing the difference in NDVI values between tree crowns and non-tree crown areas such as water, artificial buildings, bare soil, grassland, and shrubs, effectively reducing the influence of non-vegetation objects on the accuracy of ITC delineation. After that, according to the height distribution of trees in the study area, pixels with CHM values lower than 0.7 m were eliminated, as they were more likely to be shrubs or ground objects. Then, the threshold of the shadow area was used to accentuate gaps between tree crowns, facilitating the segmentation of adjacent crowns. At the same time, the objects smaller than a quarter of the shadow area threshold were removed from the imagery.

(3) Multi-scale ITC delineation method based on CHMs

First, branches and tree crowns with diameters less than 3 pixels were removed. Then, the local maxima method was used to detect the initial tree tops, and the multi-scale ITC

slices were extracted using disk SEs, according to the scale analysis results. The initial tree tops were then refined using shape parameters (height, roundness, and aspect ratio), leading to the integration of multi-scale tree crown slices. Finally, the integrated tree crowns were used as markers for marker-controlled watershed segmentation.

(4) Optimization of ITC delineation results

Due to uneven and discontinuous tree crowns, the segmentation results may include isolated points, fragmented patches, and internal holes. The optimization steps in this study include eliminating polygons with a diameter area of less than 3 pixels (1.5 m), filling in internal holes, and finally outputting an optimized binary tree crown segmentation image to ensure complete closure of the tree crown.

3.1.2. Construction of Sample Set

This study selected regions with typical tree distribution as the validation areas, and 9 sub-areas were selected near the validation areas as the training areas, covering single trees of different heights and crown areas, providing sufficient training samples for the deep learning model. The improved MSAS method used in this study was implemented using Matlab software (version R2021b). First, preliminary ITC delineation results for the nine training areas were obtained using the improved MSAS method. Subsequently, manual optimization of tree crowns was performed with the assistance of high-resolution aerial imagery and CHMs, significantly improving the efficiency of sample delineation. Some research indicates that the combination of multispectral images and elevation data (DSM or CHM) as the input of CNNs can improve accuracy [63,64]. Based on the above research conclusions, this study employed a combination of NIR, R, and G bands, and CHMs as the input image of Mask R-CNN.

A number of 20,753 tree crowns were delineated from the nine selected sub-areas covering a total area of 8 km^2, which were then used to generate training samples. One of the nine training images and a sub-image are shown in Figure 7. The five plots covering a total area of 459,186 m^2 with 1558 manually delineated tree crowns were used as the testing images. Thus, a total number of 22,687 tree crowns were included in the training and testing images. The training images were clipped with a size of 128 × 128 pixels to generate the sample images. To address the issue of split-edge tree crowns, a stride (overlap degree) of 50% was set to ensure complete capture of each tree crown. Data augmentation was performed on the original sample images to mitigate the risk of overfitting through horizontal flipping, vertical flipping, and rotations of 90°, 180°, and 270°, as shown in Figure 8. The augmented sample set consisted of 31,007 sample images and their corresponding labels. These were randomly divided into training and validation sets with a ratio of 7:3.

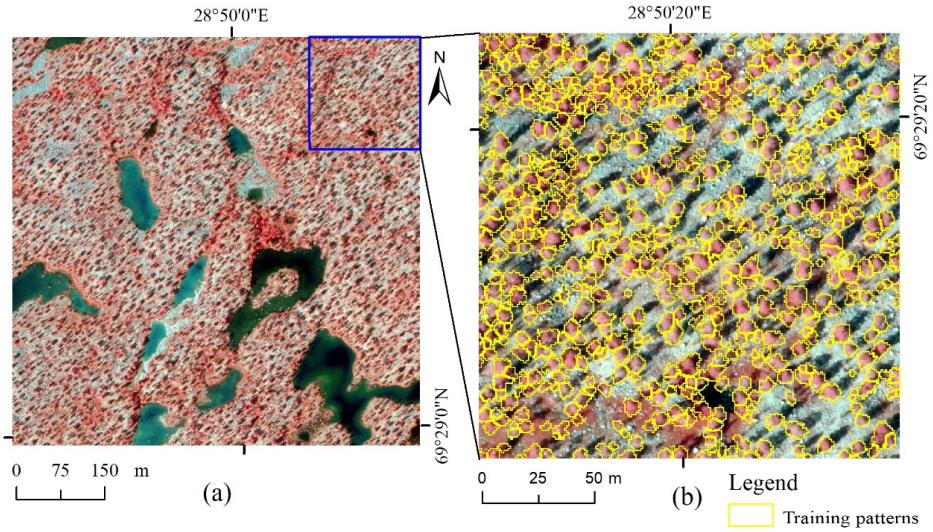


Figure 7. *Cont.*

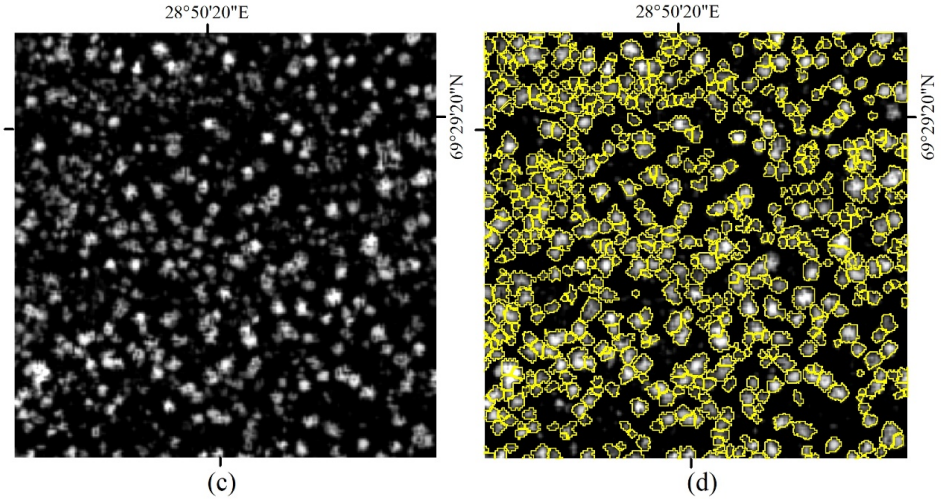


Figure 7. One of the nine training images and a sub-image of the training image. (**a**) One of the nine training images; (**b**) a sub-image of (**a**) and training patterns; (**c**) CHM of the sub-image; and (**d**) CHM of the sub-image and training patterns.

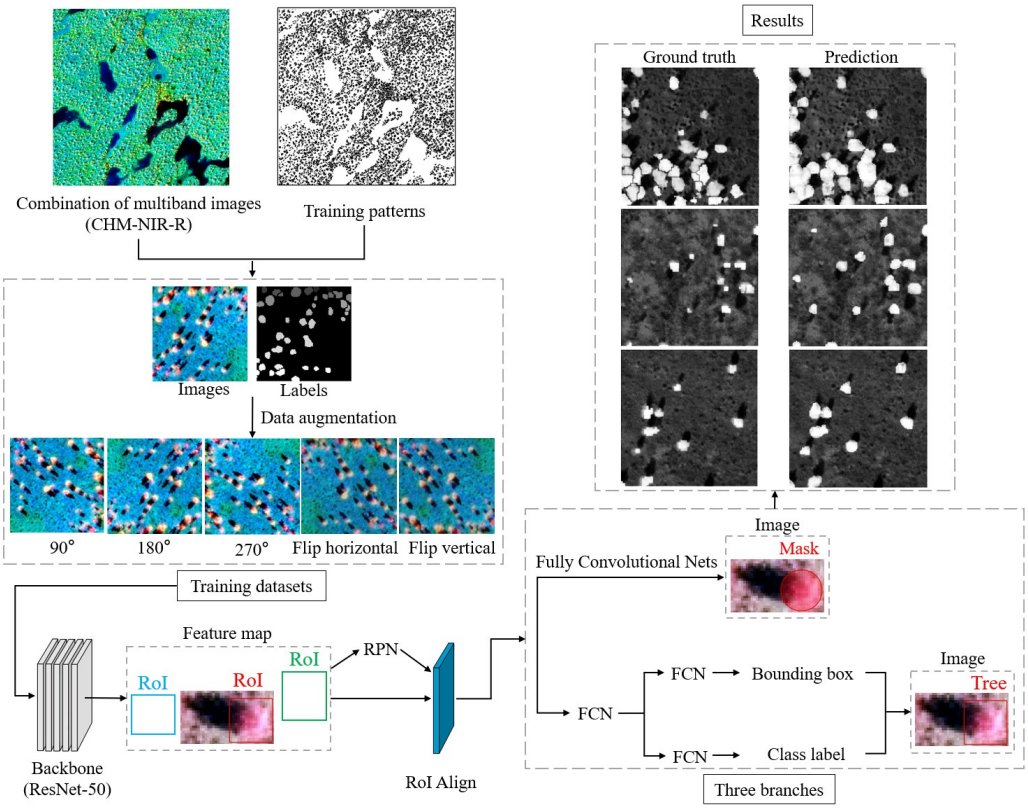


Figure 8. The workflow for ITC delineation using Mask R-CNN. The FCN generates labels and bounding boxes for each object, and the Fully Convolutional Nets produce masks including the pixels of each object.

3.2. Training and Application of the Mask R-CNN Model

3.2.1. Instance Segmentation with Mask R-CNN

Mask R-CNN is a multi-task model integrating object detection and semantic segmentation, which can realize end-to-end training. It assigns specific numbers for identifying each object and precisely delineates the boundaries of target objects at the pixel level [35]. Figure 8 shows the workflow of ITC delineation using Mask R-CNN, with the following main processing steps: (i) The preprocessed images were input into the pre-trained neural

network ResNet-50 to extract feature maps. (ii) These feature maps were fed into the Region Proposal Network (RPN) to select the Regions of Interest (ROIs). (iii) The ROI Align operation was applied to these ROIs, transforming all ROIs into a uniform shape. (iv) A Fully Connected Network (FCN) was employed to determine the class (label) of objects and refine the position and size of bounding boxes to encapsulate the objects. (v) Fully Convolutional Nets generated masks by identifying pixels belonging to each tree crown.

3.2.2. Training and Prediction of Mask R-CNN Model

The Mask R-CNN model was trained using the above samples (Section 3.1.2). After evaluating the performance of different backbone models [55], ResNet-50 was chosen as the pre-trained model for this study. Training stopped when the model ceased showing improvement (no validation loss reduction within five epochs). The model's learning rate and number of iterations were set at 2.2909×10^{-5} and 39, respectively. The trained model was then applied to the five plots with the following prediction parameters: (i) To reduce information loss caused by the contraction of effective edge pixels, a Padding parameter was set, adding extra rows or columns of pixels around the input image's outer edges. (ii) The confidence threshold was adjusted to determine whether objects within the bounding boxes were positive or negative samples. In this study, objects with a confidence score higher than 0.2 were considered positive samples, while those below 0.2 were considered negative samples or backgrounds, thereby preserving the best candidate boxes. (iii) Perform non-maximal suppression to remove overlapping tree crowns with lower confidence scores; the maximum overlap ratio was set to 0 [65]. The model was implemented in the ArcGIS API for Python, with the following hardware configuration used for model training and application: (i) Operational System: Windows 10; (ii) CPU: Intel(R) Xeon(R) W-2245 CPU @ 3.90 GHz (16 CPUs), ~3.91 GHz; (iii) GPU: NVIDIA GeForce RTX 3080, 10 GB; and (iv) RAM: 64.0 GB.

3.3. Comparison Methods and Accuracy Evaluation Indicators

3.3.1. Comparison with Traditional ITC Delineation Methods

This study considered two traditional ITC delineation methods compared to the Mask R-CNN approach. The first method is the improved MSAS method (Section 3.1.1), whereas the second is a local maximum clustering approach based on point cloud distribution proposed by Li et al. [29].

Li's method performs ITC delineation directly based on point cloud distribution, utilizing the principle of local maximum clustering. Figure 9 shows the workflow of ITC delineation using Li's method. First, the point cloud data were preprocessed to separate ground points from non-ground points. Kriging is then used to interpolate the ground points, generating a 1 m resolution DEM. After that, the non-ground points were normalized using the DEM to remove the influence of topographic relief on the elevation values of the point cloud data. The normalized point cloud data were iteratively classified from the highest to the lowest points into two classes: target and adjacent trees. Points with spacing smaller than the threshold were classified based on a minimum spacing rule. The process was iterated until all point clouds were classified and iteration ceased.

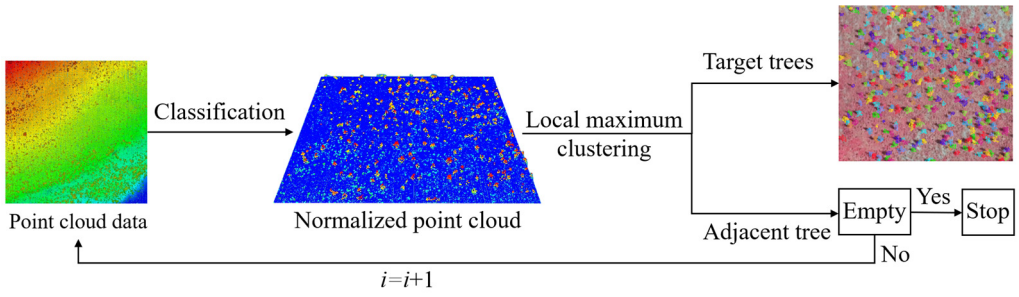


Figure 9. The workflow for ITC delineation using Li's method; "*i*" represents the iteration count.

3.3.2. Accuracy Indices

To quantitatively assess accuracy, delineated tree crowns were classified into five categories according to their spatial relationship with reference tree crowns (manually delineated tree crowns) as shown in Figure 10 [25,60,66]: (i) Nearly matched: Only one of the reference or target tree crowns has an overlap area greater than 50% with the other. (ii) Matched: Both the reference and target tree crowns have a mutual overlap area greater than 50%. (iii) Merged: Multiple reference tree crowns are covered by a single target tree crown with an overlap area exceeding 50%, implying that multiple reference tree crowns are merged into a single target tree crown. (iv) Split: Multiple target tree crowns are covered by a single reference tree crown with an overlap area exceeding 50%, signifying that a single tree crown is split into multiple target tree crowns. (v) Missing: The reference tree crown overlaps less than 50% with any target tree crown and more than 50% with the background, indicating that the crown is omitted. Based on these category definitions, the sum of nearly matched and matched tree crowns is considered correctly delineated and classified as True Positives (*TPs*). False Positives (*FPs*): The model predicted that there is a tree, and it is incorrect. False Negatives (*FNs*): The model predicted that there is no tree, and it is incorrect.

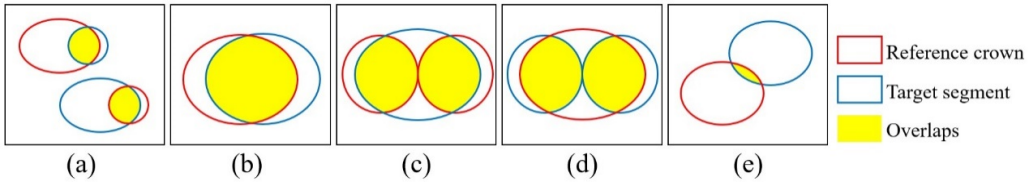


Figure 10. The diagram of classifying target tree crowns into five categories according to their spatial relationship with reference tree crowns: (**a**) nearly matched; (**b**) matched; (**c**) merged; (**d**) split; and (**e**) missing.

The recall (A_r), delineation accuracy (A_d), and F-measure (F) were used to evaluate the accuracy of ITC delineation. For reference and target tree crowns, the Intersection over Union (IoU), the ratio of the intersection to the union of their areas, is calculated to assess the accuracy of the tree crown polygons. The IoU was considered acceptable when it was higher than 0.5. The specific formulas for these evaluation metrics are as follows:

$$A_r = \frac{TP}{N_r} \times 100\% \quad (1)$$

$$A_d = \frac{TP}{N_t} \times 100\% \quad (2)$$

$$F = \frac{2 \times A_r \times A_d}{A_r + A_d} \quad (3)$$

$$IoU = \frac{TP}{FP + TP + FN} \quad (4)$$

where TP is True Positive, indicating the number of correctly delineated tree crowns in the target canopy; N_r is the total number of tree crowns in the reference canopy; N_t is the total number of tree crowns in the target canopy; FP is False Positive; and FN is False Negative.

4. Results

The trained Mask R-CNN model was applied to predict the ITCs of the five plots. We assessed accuracy using 1558 manually delineated tree crowns. The accuracies of the three ITC delineation methods are shown in Tables 3 and 4. Figures 11–14 present the reference tree crowns and the delineation results of the three methods.

Table 3. The accuracy metric of Mask R-CNN for individual trees higher than 1.3 m in each plot.

Area ID	Target Tree Crowns	Reference Tree Crowns	Nearly Matched	Matched	Merged	Missing	Split	A_d (%)	A_r (%)	F (%)	IoU (%)
Plot 1	92	93	3	89	0	1	0	100.00	98.93	99.46	98.92
Plot 2	339	346	78	247	2	17	2	95.87	93.93	94.89	90.28
Plot 3	239	226	5	211	4	6	0	90.38	95.58	92.90	86.75
Plot 4	564	562	15	530	3	10	4	96.63	96.98	96.80	93.80
Plot 5	334	331	4	323	1	2	1	97.90	98.79	98.35	96.75

Table 4. Overall ITC delineation accuracy of different algorithms.

Algorithms	Target Tree Crowns	Reference Tree Crowns	Correctly Delineated Tree Crowns	A_d (%)	A_r (%)	F (%)	IoU (%)
Mask R-CNN	1568	1558	1505	95.98	96.60	96.29	92.84
The improved MSAS method	1744	1558	1474	84.52	94.61	89.28	80.63
Li's method	1635	1558	1419	86.79	91.08	88.88	79.99

As shown in Table 3, the A_r of Mask R-CNN at five plots ranged from 93.93% to 98.93%. Specifically, 92 tree crowns of the 93 reference tree crowns were accurately delineated in plot 1, achieving the highest A_d, A_r, F, and IoU of 100.00%, 98.93%, 99.46%, and 98.92%, respectively. In contrast, 325 tree crowns were correctly delineated for plot 2, offering the lowest A_r of 93.93%. Twenty-one crowns were misidentified, of which 17 crowns were missing due to the relatively small crown area. As shown in Table 2, plot 2 had an average tree crown area of 18.94 m^2, which is the smallest among the five plots. The relatively smaller tree crown sizes contribute to the lower accuracy of plot 2. The A_d of Mask R-CNN at five plots ranged from 90.38% to 100.00%. Specifically, plot 1 achieved the highest A_d of 100.00%, with 92 tree crowns accurately delineated and one tree crown missing. In contrast, plot 3 yielded the lowest A_d of 90.38%. As shown in Figures 12–14, the tree density and complexity of plot 3 were significantly higher than plot 5, with an A_r of 95.58% for plot 3 and 98.79% for plot 5, which was 3.21 percentage points higher than plot 3. The tree density of plot 4 was higher than plot 3 but lower than plot 5, with an A_r of 96.98%, which was 1.4 percentage points higher than plot 3 and 1.81 percentage points lower than plot 5. As shown in Table 2, all of the reference tree crowns exceed an area of 3 m^2. The area of 3 m^2 corresponds to an area covering 12 pixels in 0.5 m resolution aerial imagery and 3 pixels in the 1 m resolution CHM. These conclusions indicate that the high-resolution aerial imagery combined with low-density point cloud data used in this experiment achieved high accuracy for tree crowns larger than 3 m^2.

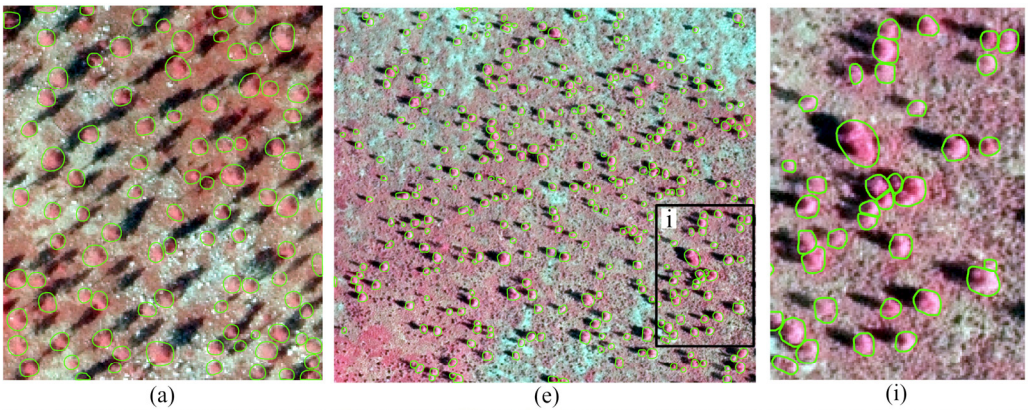


Figure 11. *Cont.*

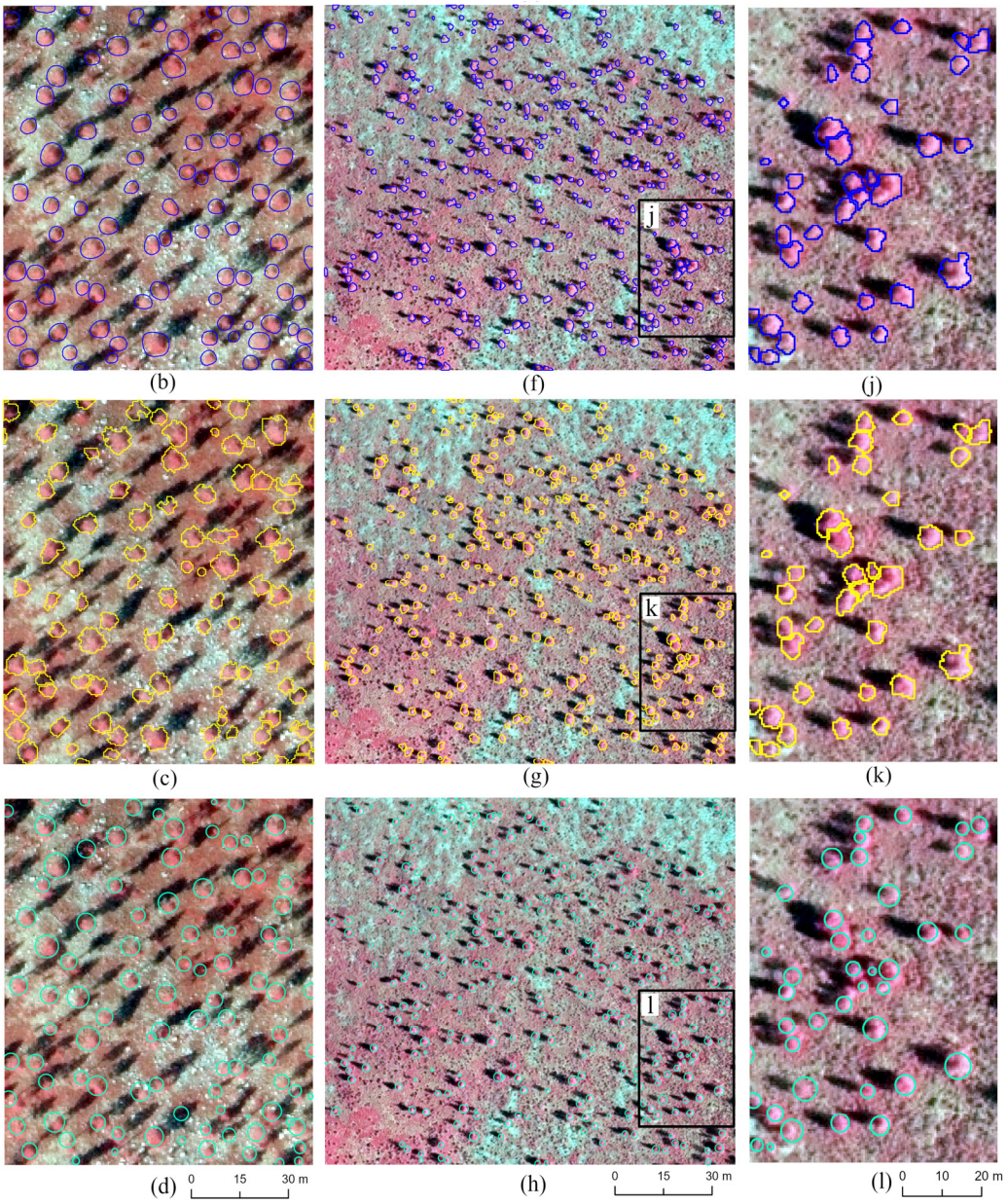


Figure 11. The reference tree crowns (**a**) and the ITC delineation results of plot 1 using Mask R-CNN (**b**), the improved MSAS method (**c**), and Li's method (**d**). The tree crowns are shown in polygons with different colors. The reference tree crowns (**e**) and the ITC delineation results of plot 2 using Mask R-CNN (**f**), the improved MSAS method (**g**), and Li's method (**h**); (**i**–**l**) are the local detail maps of (**e**–**h**), respectively.

The results in Table 4 show that the overall A_r of Mask R-CNN reaches 96.60%, indicating that the model achieves high ITC delineation accuracy. The improved MSAS method delineated 1474 ITCs out of the 1558 reference tree crowns of the five plots, achieving A_r, A_d, F, and IoU values of 94.61%, 84.52%, 89.28%, and 80.63%, respectively. Compared to the improved MSAS method, the overall A_r, A_d, F, and IoU of Mask R-CNN improved by 1.99, 11.46, 7.01, and 12.21 points in percentage, respectively. The number of delineated tree crowns decreased by 176, and the number of correctly delineated tree crowns increased by 31. As shown in Figure 12f,g, Mask R-CNN outperformed the improved MSAS method in distinguishing between tree crowns and shrubs, resulting in fewer incorrectly delineated tree crowns. Li's method delineated 1419 ITCs, achieving A_r, A_d, F, and IoU values of 91.08%, 86.79%, 88.88%, and 79.99%, respectively. Compared to Li's method, Mask R-CNN

showed improvements of 5.52, 9.19, 7.41, and 12.85 points in percentage in A_r, A_d, F, and IoU, respectively. The number of delineated tree crowns decreased by 67, and the number of correctly delineated tree crowns increased by 86. It can be seen from the local details shown in Figure 11j,l that the contours of the delineated tree crowns obtained by Mask R-CNN were more accurate than those of Li's method. The results demonstrated that Mask R-CNN significantly improves ITC delineation precision for trees higher than 1.3 m compared to traditional methods, which is consistent with the results of Yu et al. [67]. Yu et al. [67] also showed that the ITC delineation accuracy of Mask R-CNN outperformed those of the local maxima algorithm and the marker-controlled watershed segmentation method. The accuracy of ITC delineation was influenced by various factors, the sources of error or uncertainty in ITC results include factors such as shape and density of tree crowns, occlusion between trees, and image quality.

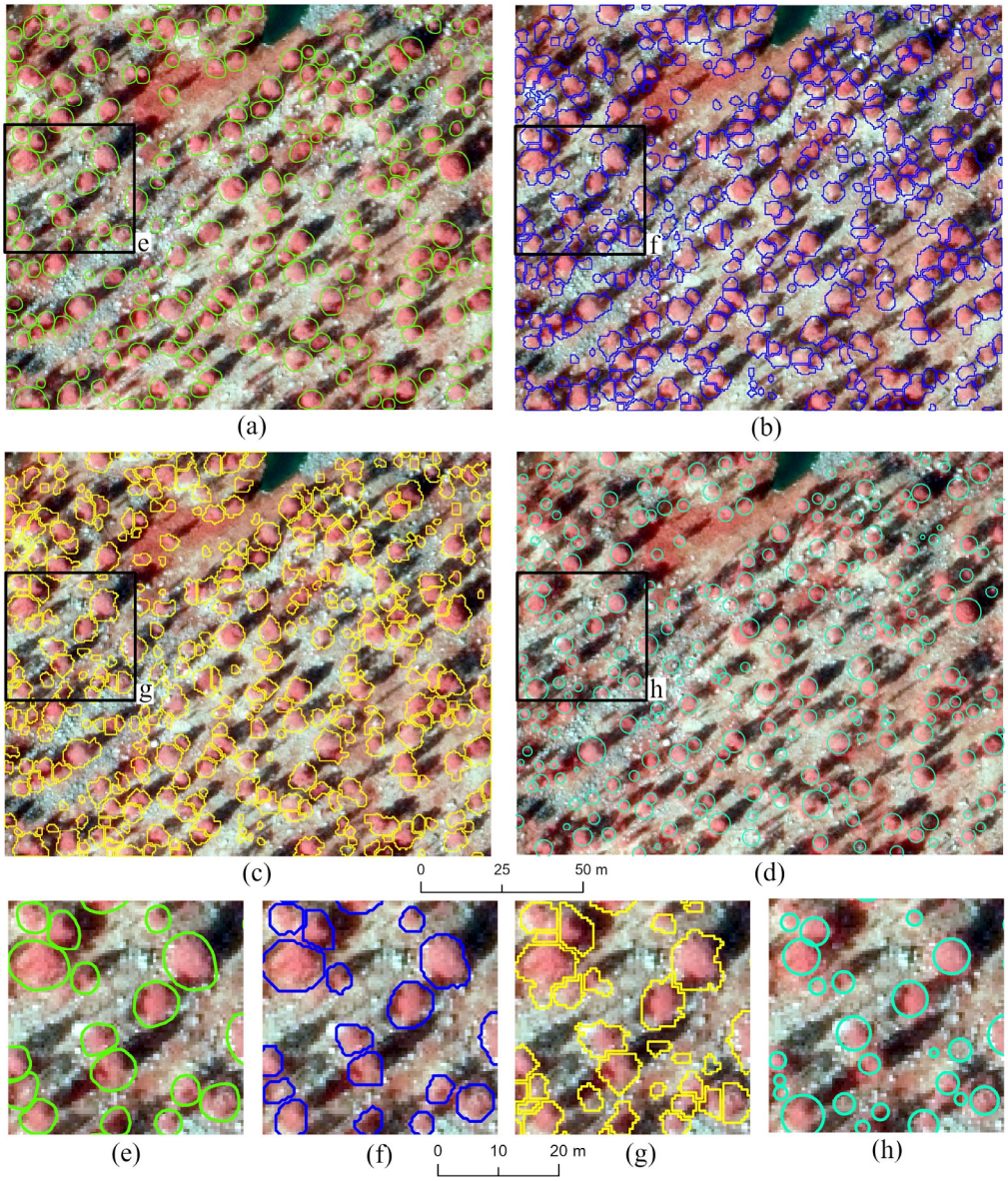


Figure 12. The reference tree crowns (**a**) and the ITC delineation results of plot 3 using Mask R-CNN (**b**), the improved MSAS method (**c**), and Li's method (**d**). The tree crowns are shown in polygons with different colors. (**e**–**h**) are the local detail maps of (**a**–**d**), respectively.

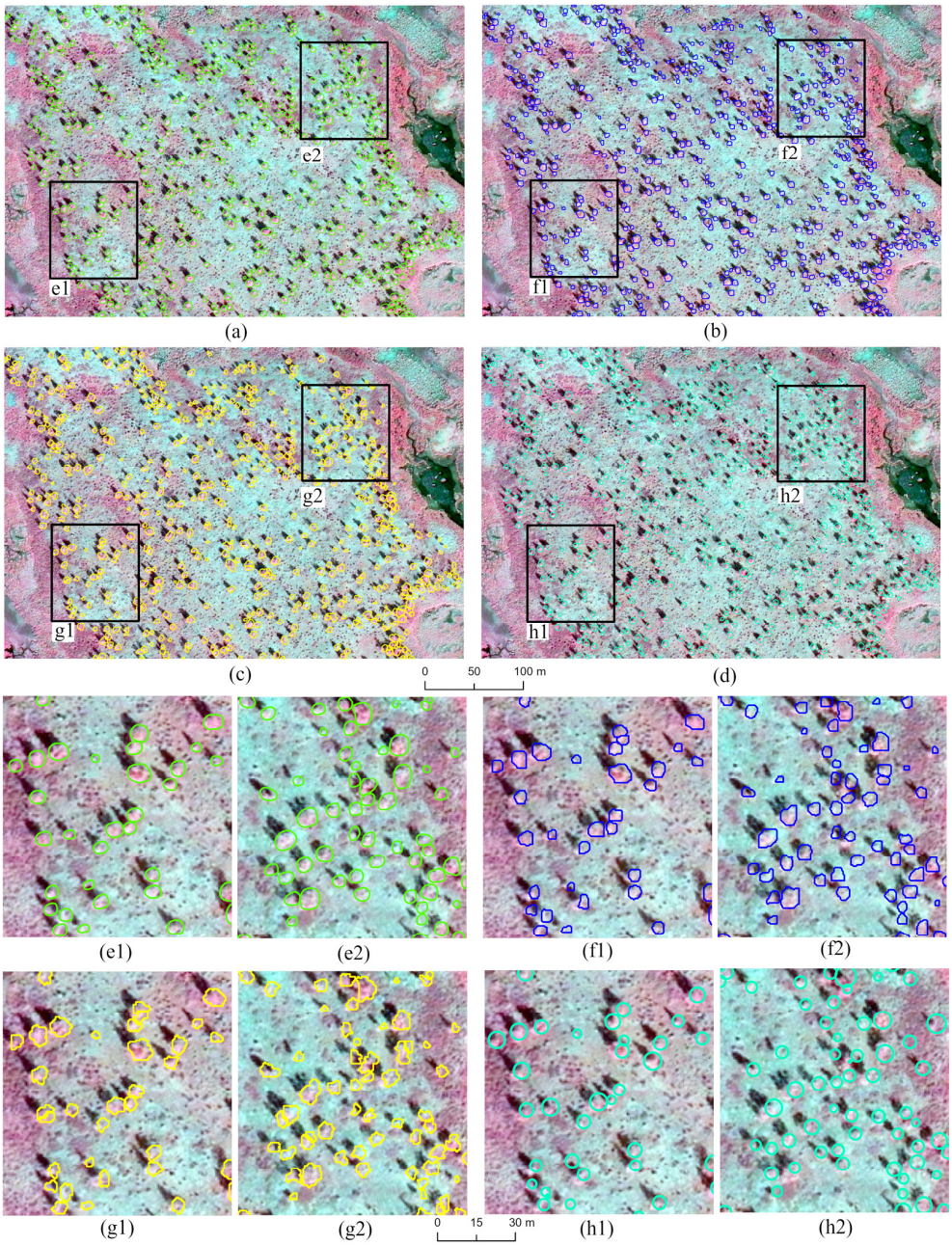


Figure 13. The reference tree crowns (**a**) and the ITC delineation results of plot 4 using Mask R-CNN (**b**), the improved MSAS method (**c**), and Li's method (**d**) The tree crowns are shown in polygons with different colors. (**e1**–**h2**) are the local detail maps of (**a**–**d**), respectively.

Deep learning exploits end-to-end learning to extract information from multiple image layers, avoiding the subjectivity of traditional feature extraction methods. It boasts rapid identification, high accuracy, minimal feature requirements, and self-training capabilities. In experimental comparisons with other instance segmentation architectures, Mask R-CNN has demonstrated superior performance [35]. It integrates the tasks of object detection and semantic segmentation. However, deep learning methods require additional training sets for model training, which requires the computer to have solid computational power in addition to a large number of training samples. Although the application of deep learning in forestry remote sensing is still in the initial stage, it has a broad prospect in forest surveys with significant advantages.

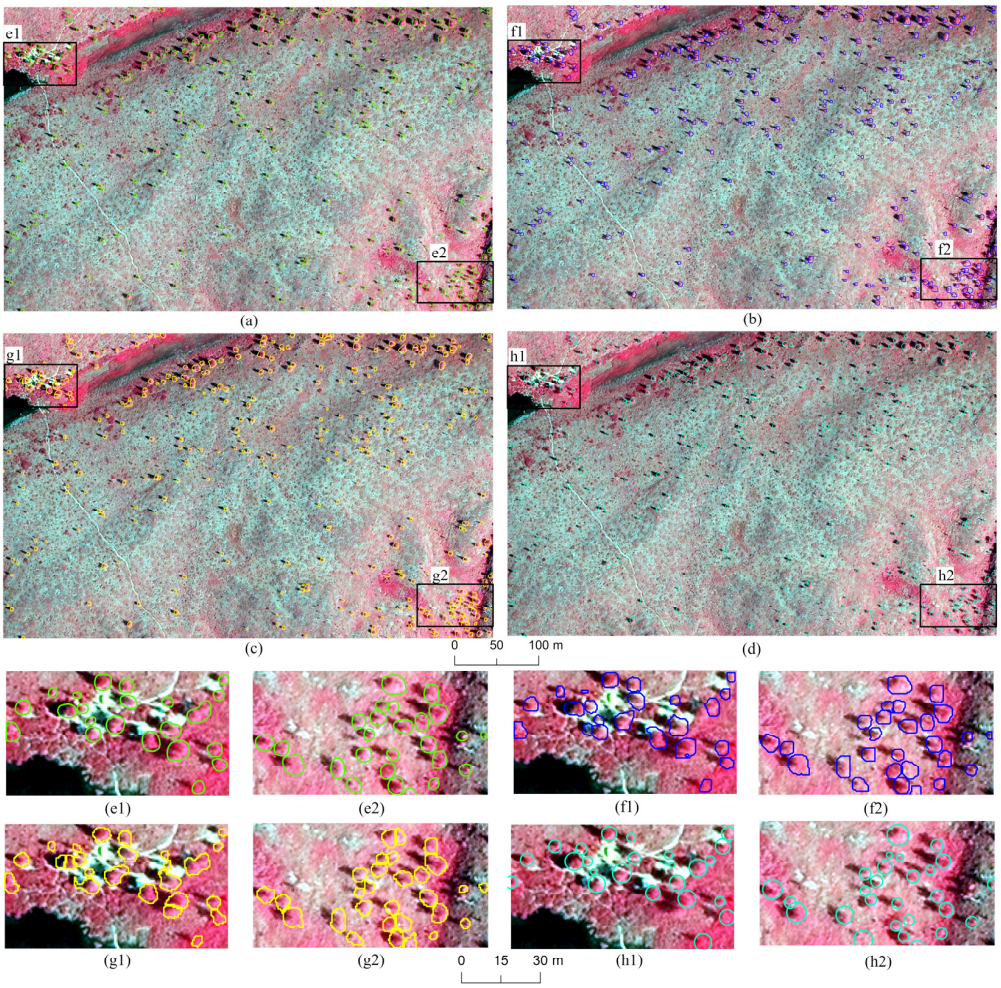


Figure 14. The reference tree crowns (**a**) and the ITC delineation results of plot 5 using Mask R-CNN (**b**), the improved MSAS method (**c**), and Li's method (**d**) The tree crowns are shown in polygons with different colors. (**e1**–**h2**) are the local detail maps of (**a**–**d**), respectively.

5. Discussion

5.1. Comparison with Other Similar Studies

This study achieved outstanding accuracies compared to existing studies on ITC delineation using deep learning, although it showed significant differences in the study area background and data.

(1) First, there are differences in forest land types. Hao et al. [63] utilized Mask R-CNN for detecting young Chinese fir (1~4 m) in a young plantation forest, achieving an F1 score of 84.68% and an *IoU* of 91.27%. Specifically, the plantation included only a single tree species (Chinese fir) with a relatively low height, the distribution and shape of which were relatively regular. In contrast, the plots in this study primarily included natural coniferous trees (Scots pine and Norway spruce), with a more irregular distribution of individual trees and crown shapes, making it more difficult to delineate.

(2) In addition, there is a difference in the height of individual trees. Dersch et al. [68] employed Mask R-CNN for ITC delineation in a coniferous forest with Norway spruce as the dominant tree species, achieving an F1 score of 92% and an accuracy of 85%. The height of the trees in the test area ranged from 19 to 35 m, and this study excluded point cloud data below 10 m to focus just on trees taller than 10 m, thereby avoiding confusion between trees and shrubs. The heights of the individual trees in this study area ranged from 1.30 to 13.83 m and an average height of 4.47 m (Table 2). Although

the trees in this study area were relatively low, the A_r for trees higher than 1.3 m still reached 96.60%, providing an F1 score of 96.29%.

(3) Differences in density and complexity. Weinstein et al. [69] employed LiDAR data (with a density of 4–6 points/m^2) and RGB images (at a resolution of 0.5 m), leveraging deep learning techniques to identify individual trees within four distinct forest types: Oak Woodland, Mixed Pine, Alpine, and Eastern Deciduous. The Alpine forest is close to the treeline, and the major tree species in the Alpine forest are subalpine fir and Englemann spruce. The study area of this work was located in the transitional zone, similar to the Alpine forest, and the trees in this region were small and low. The study of Weinstein et al. [69] indicated that the best performance was achieved in open-canopy forests with large, well-spaced trees, such as the Oak Woodland forest. In contrast, the Alpine forest showed the greatest challenge, where short and small trees posed a significant obstacle compared to the complex, interconnected tree boundaries observed across other forest types. The smaller tree sizes in the Alpine forest and the transitional zone forests make them more susceptible to variations in the spatial resolution of the data used. As the major dataset of this work, the CHM derived from the low-density LiDAR data was crucial. However, the 1 m spatial resolution CHM and the 0.5 m resolution aerial imagery were adequate for large trees, but smaller trees may be missed.

(4) The aerial imagery utilized in this study was recorded in 8 bit, with pixel values ranging from 0 to 255. The imagery might be a radiometrically stretched version rather than the original raw imagery. Further, there was a temporal discrepancy between the LiDAR data and the aerial imagery used in this study (Table 1). Although there were few human activities in the study area where the vegetation showed minor changes, there were still some differences in the land cover and canopy shadows between the data recorded in different time phrases, which might have a certain degree of impact on the accuracy.

Despite these limitations in data quality, the Mask R-CNN model employed in this study still achieved outstanding performances. Compared to other ITC delineation studies conducted in the transitional zone using point cloud data, this research obtained a higher accuracy [43,44,46]. The results demonstrated the potential of applying the Mask R-CNN model and low-density point cloud data for ITC delineation in the Taiga–Tundra Ecotone.

5.2. Semi-Automatic Acquisition of Training Sample Set

In this study, trees in the Taiga–Tundra Ecotone were relatively low and their crowns were small. Manual delineation of sufficient samples within the study area is time-consuming. However, the construction of high-accuracy deep learning models heavily depends on a large number of training samples. This study combined automated ITC delineation methods with manual visual inspection and optimization to construct the ITC sample set, significantly improving the efficiency of sample delineation. The improved MSAS method was employed to obtain preliminary results of sample tree crowns, followed by visual inspection and manual optimization to ensure the accuracy of sample delineation. The results showed that the improved MSAS method achieved an overall A_r of 94.61% of the five plots (Table 4). Compared with other studies using traditional ITC delineation methods in the transitional zone, the improved MSAS method reached higher accuracy, ensuring more accurate preliminary ITC delineation results. For example, Stumberg et al. [43] used an unsupervised classification method on point cloud data for ITC delineation in the Taiga–Tundra Ecotone, achieving a maximum accuracy of 72.9%. Hauglin and Næsset [44] applied an ITC delineation method using an allometric growth model on a point cloud data of Norway in the Taiga–Tundra Ecotone, achieving an accuracy of 80% for trees higher than 3 m. The proposed sample set construction method can also be used for sparse forests and arid and semi-arid regions with sparse trees.

This study constructed a sample set including 22,687 individual tree crowns covering a total area of 8 km^2 in the transitional zone. The sample number was approximately

11 times that used by Zhang et al. [38] and 15 times that of Braga et al. [40]. The area of the corresponding training images was approximately 43 times that of Zhang et al. [38]. The individual tree sample set constructed in this study can provide a valuable resource for subsequent research on ITC delineation methods in the Taiga–Tundra Ecotone, assisting in large-scale ITC delineation within these areas. Further improvements can be made by supplementing the samples from the central and northern parts of the study area to increase the number of samples and the significance of the sample set further. The ITC delineation of the transition zone and the further extracting information such as tree heights and crown diameter of these trees is helpful for studying the response of individual tree growth in transitional zones to climate change and dynamically inverting ecological indicators such as vegetation biomass and carbon sequestration in the ecotone, which can be used for biodiversity assessments, ecosystem services, and conservation research.

5.3. Influence of Model Input of Mask R-CNN-Based ITC Delineation

By constructing complex multilayer neural networks, deep learning models learn and extract deep feature information from multi-band images, significantly improving the automation of ITC delineation. However, using too many feature bands as input may overload redundant information and potentially reduce the accuracy of ITC delineation. Several studies have explored the impact of different inputs of Mask R-CNN on the ITC delineation accuracy. For example, Yu et al. [67] showed that the Mask R-CNN model using RGB images as input achieved higher accuracy (with an F1 score of 94.68%) than models using single-band images. Additionally, some research indicates that combining multispectral images and elevation data (DSM or CHM) as the input of CNNs can improve accuracy. Hao et al. [63] explored ITC delineation using six band combinations from aerial imagery and CHMs as the input of Mask R-CNN. They found that the combination of NDVI and CHMs achieved the best performance, with an F1 score of 84.68%. Lucena et al. [64] compared the trained Mask R-CNN models using only RGB images and the combinations of RGB images and CHMs as input, discovering that the latter obtained an overall accuracy of 97.01%, which is higher than the accuracy of 90.42% provided by the former. This study employed a combination of NIR, R, G bands, and CHMs as the input image of Mask R-CNN, achieving an overall A_r of 96.60%. However, the optimal input for deep learning models still needs to be explored for different study areas and data situations to ensure sufficient feature information while reducing redundancy. For example, ablation experiments can be further performed to quantify the critical role of spectral bands and CHM data on ITC delineation.

5.4. Influence of the Ratio of Tree Crown Diameter to Spatial Resolution

The ratio of tree crown diameter to spatial resolution is a crucial factor influencing the accuracy of ITC detection. Brandt et al. [39] and Braga et al. [40] showed that image resolution and tree crown size can influence the accuracy of ITC delineation using deep learning models such as U-Net and Mask R-CNN. Larger tree crown diameters and higher image resolutions generally lead to higher accuracy for deep learning models.

High-density LiDAR data can describe the terrain more accurately, which is essential for generating high-precision CHMs, enabling the detection of small targets [70,71]. Studies on ITC delineation using LiDAR data indicated that an increase in the point cloud density significantly improved the accuracy of ITC delineation. Reitberger et al. [72] and Yao et al. [73] showed that point cloud data with a density of approximately 10 pts/m^2 is optimal for ITC delineation. The aerial imagery and CHM used by Hao et al. [63] had a spatial resolution of 0.02 m. Dersch et al. [68] used 0.05 m resolution multispectral imagery and a 0.05 m CHM derived from high-density point cloud data (560 pts/m^2). Existing ITC delineation research in transitional zones employed LiDAR data with point cloud densities ranging from 6.8 to 15 pts/m^2 [43–47]. In contrast, the airborne LiDAR data used in this study had a spatial resolution of 0.5 m, and the CHM had a spatial resolution of 1 m (the density of point cloud data is only 0.8~1.1 pts/m^2), which is lower than other works con-

ducted in the transitional zone. The diameter range of the 1558 tree crowns of the five plots of this study is 2.11~10.70 m, with an average diameter of 5.28 m. A number of 1224 tree crowns, accounting for 79% of the total, had diameters greater than 4 m. According to the empirical rule proposed by Yin and Wang [74] that the spatial resolution of a CHM should be at least one-quarter of the crown diameter, this dataset enabled the delineation of the trees with a crown diameter greater than 4 m. Only 79% of the tree crowns of the five plots met the standard. Overall, the resolution of the data used in this work was lower, and the tree crown in the transition zone was relatively small, which is lower than the delineation standard. However, the Mask R-CNN model used in this study still achieved outstanding performance, demonstrating an overall accuracy of 96.60%. If higher-density point cloud data can be obtained, it may be possible to further improve the precision of ITC delineation in the study area. According to the empirical rule of Yin and Wang [74], using a CHM with a spatial resolution of 0.33–0.5 m in further work would improve the accuracy while enabling the delineation of trees with a crown diameter of less than 2 m.

Additionally, the quality of the CHM derived from LiDAR data with relatively low density is another factor that needs attention. The optimization problem of the CHM also needs to be considered. The original CHM obtained from the airborne LiDAR data used in this work exhibited "holes", which disrupted the integrity of the morphological structure of tree crowns. These anomalies were typically attributed to three factors. First, it may be the gap between tree crowns, which was formed due to the low height of vegetation and bare land around the tree crown. The second was the invalid values related to the scanning mode of the laser sensor [59]. The third may be affected by low point density, especially near slice boundaries. Currently, smoothing filters are commonly used to optimize CHMs. This study compared the effects of three filters, which are the low-pass, median, and Gaussian low-pass filters, and ultimately selected the low-pass filter with a better optimization effect. The optimization methods for CHMs can be discussed in depth to improve the accuracy of ITC delineation.

6. Conclusions

The Taiga–Tundra Ecotone covers a large area of mixed coniferous forests. However, the trees in this zone have relatively low heights and small crowns. There is an urgent need to explore ITC delineation methods and models using deep learning, especially for large-scale applications in the transition zone. For the first time, this study explores the feasibility of the Mask R-CNN model for the large-scale delineation of ITCs in the transition zone using low-density point cloud data and aerial imagery. The accuracies of the Mask R-CNN model were compared with two traditional ITC delineation methods (the improved MSAS method and Li's method) for trees with different heights.

The results show the following: (i) For individual trees higher than 1.3 m, the improved MSAS method and Li's method achieved an overall recall rate (A_r) of 94.61% and 91.08%, respectively; Mask R-CNN achieved an overall recall rate (A_r) of 96.60% (with an A_d of 95.98%, an F of 96.29%, and an IoU of 92.84%), which is 1.99 and 5.52 points in a percentage higher than the two traditional ITC delineation methods, respectively, indicating higher delineation accuracy than the traditional methods. (ii) The data and algorithms used in this study accurately delineated individual trees with crown areas greater than 3 m^2. (iii) These results highlight the potential of the Mask R-CNN model to delineate small individual trees in the transition zone using high-resolution aerial imagery and low-density point cloud data. Additionally, this study proposes an improved multi-scale ITC delineation method for the semi-automatic delineation of ITC samples, ultimately constructing a sample set of over 20,000 trees in the transition zone, available for open sharing (URL: https://github.com/linyy000/ITC-samples/tree/main). This study's individual tree dataset and deep learning model provide data and technical support for large-scale ITC delineation planning and implementation in the transition zone. The results of this work promote the accurate estimation of carbon storage and biomass at the individual tree scale and research into the response of individual trees to climate change within the transition zone.

Author Contributions: Conceptualization, L.J.; methodology, Y.L.; software, Y.L.; data curation, H.D. and S.T.; writing—original draft preparation, Y.L.; writing—review and editing, H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China under grant number 2023YFB3907702; the International Research Center of Big Data for Sustainable Development Goals under grant number CBAS2022RP03; and the Key Project of "Prospering Mongolia with Science and Technology" under grant number KJX-MEEDS-2020006. We wish to thank the anonymous reviewers and editors for their detailed comments and suggestions.

Data Availability Statement: The authors confirm that the data supporting the findings of this study are available within the article. The aerial orthophotos and airborne LiDAR data used in this work were provided by the Finnish National Land Survey Authority (URL: https://asiointi.maanmittauslaitos.fi/karttapaikka/tiedostopalvelu (accessed on 26 September 2022)). This paper establishes a publicly archived dataset containing over 20,000 tree samples from the Taiga–Tundra Ecotone, which can be accessed at the following URL: https://github.com/linyy000/ITC-samples/tree/main.

Conflicts of Interest: Author Yuanyuan Lin was employed by the China Siwei Surveying and Mapping Technology Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Crawford, R.M.M. Arctic and subarctic treelines and the tundra-taiga interface. In *Plants at the Margin: Ecological Limits and Climate Change*; Cambridge University Press: Cambridge, UK, 2008; pp. 161–196.
2. Callaghan, T.V.; Werkman, B.R.; Crawford, R.M.M. The tundra-taiga interface and its dynamics: Concepts and applications. *Ambio* **2002**, *12*, 6–14.
3. Montesano, P.M.; Neigh, C.S.R.; Macander, M.; Feng, M.; Noojipady, P. The bioclimatic extent and pattern of the cold edge of the boreal forest: The circumpolar taiga-tundra ecotone. *Environ. Res. Lett.* **2020**, *15*, 105019. [CrossRef]
4. Maguire, A.J.; Eitel, J.U.H.; Vierling, L.A.; Johnson, D.M.; Griffin, K.L.; Boelman, N.T.; Jensen, J.E.; Greaves, H.E.; Meddens, A.J.H. Terrestrial lidar scanning reveals fine-scale linkages between microstructure and photosynthetic functioning of small-stature spruce trees at the forest-tundra ecotone. *Agric. For. Meteorol.* **2019**, *269*, 157–168. [CrossRef]
5. Payette, S.; Fortin, M.J.; Gamache, I. The subarctic forest-tundra: The structure of a biome in a changing climate. *Bioscience* **2001**, *51*, 709–718.
6. Bjorkman, A.D.; Myers-Smith, I.H.; Elmendorf, S.C.; Normand, S.; Rüger, N.; Beck, P.S.A.; Blach-Overgaard, A.; Blok, D.; Cornelissen, J.H.C.; Forbes, B.C.; et al. Plant functional trait change across a warming tundra biome. *Nature* **2018**, *562*, 57–62. [CrossRef] [PubMed]
7. Harding, R.; Kuhry, P.; Christensen, T.R.; Sykes, M.T.; Dankers, R.; van der Linden, S. Climate feedbacks at the tundra-taiga interface. *Ambio* **2002**, *31*, 47–55.
8. Post, E.; Forchhammer, M.C.; Bret-Harte, M.S.; Callaghan, T.V.; Christensen, T.R.; Elberling, B.; Fox, A.D.; Gilg, O.; Hik, D.S.; Hoye, T.T. Ecological dynamics across the arctic associated with recent climate change. *Science* **2009**, *325*, 1355–1358. [CrossRef]
9. Skre, O.; Baxter, R.; Crawford, R.M.M.; Callaghan, T.V.; Fedorkov, A. How will the tundra-taiga interface respond to climate change? *Ambio* **2002**, *12*, 37–46.
10. Criado, M.G.; Myers-Smith, I.H.; Bjorkman, A.D.; Lehmann, C.E.R.; Stevens, N. Woody plant encroachment intensifies under climate change across tundra and savanna biomes. *Glob. Ecol. Biogeogr.* **2020**, *29*, 925–943. [CrossRef]
11. Franke, A.K.; Feilhauer, H.; Bräuning, A.; Rautio, P.; Braun, M. Remotely sensed estimation of vegetation shifts in the polar and alpine tree line ecotone in Finnish Lapland during the last three decades. *For. Ecol. Manag.* **2019**, *454*, 117668. [CrossRef]
12. Walker, M.D.; Wahren, C.H.; Hollister, R.D.; Henry, G.H.R.; Ahlquist, L.E.; Alatalo, J.M.; Bret-Harte, M.S.; Calef, M.P.; Callaghan, T.V.; Carroll, A.B. Plant community responses to experimental warming across the tundra biome. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 1342–1346. [CrossRef] [PubMed]
13. Wieser, G. Alpine and polar treelines in a changing environment. *Forests* **2020**, *11*, 254. [CrossRef]
14. Holtmeier, F.-K.; Broll, G. Altitudinal and polar treelines in the northern hemisphere causes and response to climate change. *Polarforschung* **2009**, *79*, 139–153.
15. Lokken, J.O.; Evju, M.; Söderström, L.; Hofgaard, A. Vegetation response to climate warming across the forest-tundra ecotone: Species-dependent upward movement. *J. Veg. Sci.* **2020**, *31*, 854–866.
16. Liu, H.; Zhang, X.; Zhang, Y.; Zhu, Y.; Liu, H.; Wang, L. Review on individual tree detection based on airborne LiDAR. *Laser Optoelectron. Prog.* **2018**, *55*, 82805.

17. Gougeon, F.A. A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. *Can. J. Remote Sens.* **1995**, *21*, 274–284. [CrossRef]
18. Pollock, R.J. The Automatic Recognition of Individual Trees in Aerial Images of Forests Based on A Synthetic Tree Crown Image Model. Doctoral Dissertation, University of British Columbia, Vancouver, BC, Canada, 1996.
19. Brandtberg, T.; Walter, F. Automated delineation of individual tree crowns in high spatial resolution aerial images by multiple-scale analysis. *Mach. Vis. Appl.* **1998**, *11*, 64–73.
20. Pouliot, D.A.; King, D.J.; Bell, F.W.; Pitt, D.G. Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration. *Remote Sens. Environ.* **2002**, *82*, 322–334.
21. Erikson, M. Structure-preserving segmentation of individual tree crowns by brownian motion. *Lect. Notes Comput. Sci.* **2003**, *2749*, 283–289.
22. Wang, L. A multi-scale approach for delineating individual tree crowns with very high resolution imagery. *Photogramm. Eng. Remote Sens.* **2010**, *76*, 371–378. [CrossRef]
23. Barnes, C.; Balzter, H.; Barrett, K.; Eddy, J.; Milner, S.; Suárez, J.C. Individual tree crown delineation from airborne laser scanning for diseased larch forest stands. *Remote Sens.* **2017**, *9*, 231. [CrossRef]
24. Chen, Q.; Baldocchi, D.; Gong, P.; Kelly, M. Isolating individual trees in a savanna woodland using small footprint lidar data. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 923–932. [CrossRef]
25. Jing, L.H.; Hu, B.X.; Li, J.L.; Noland, T. Automated delineation of individual tree crowns from LiDAR data by multi-scale analysis and segmentation. *Photogramm. Eng. Remote Sens.* **2012**, *78*, 1275–1284. [CrossRef]
26. Qiu, L.; Jing, L.H.; Hu, B.X.; Li, H.; Tang, Y.W. A new individual tree crown delineation method for high resolution multispectral imagery. *Remote Sens.* **2020**, *12*, 585. [CrossRef]
27. Wu, X.Q.; Shen, X.; Cao, L.; Wang, G.B.; Cao, F.L. Assessment of individual tree detection and canopy cover estimation using unmanned aerial vehicle based light detection and ranging (UAV-LiDAR) data in planted forests. *Remote Sens.* **2019**, *11*, 908. [CrossRef]
28. Zhang, C.Y.; Zhou, Y.H.; Qiu, F. Individual tree segmentation from LiDAR point clouds for urban forest inventory. *Remote Sens.* **2015**, *7*, 7892–7913. [CrossRef]
29. Li, W.K.; Guo, Q.H.; 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. [CrossRef]
30. Ke, Y.H.; Quackenbush, L.J. A review of methods for automatic individual tree-crown detection and delineation from passive remote sensing. *Int. J. Remote Sens.* **2011**, *32*, 4725–4747. [CrossRef]
31. Komati, P.R.; Komati, R.D. Review on deep learning in remote sensing image classification. *IJFGCN* **2020**, *13*, 1223–1227.
32. Adegun, A.A.; Viriri, S.; Tapamo, J.R. Review of deep learning methods for remote sensing satellite images classification: Experimental survey and comparative analysis. *J. Big Data* **2023**, *10*, 93. [CrossRef]
33. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on convolutional neural networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [CrossRef]
34. He, K.M.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV) 2017, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
35. He, K.M.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 386–397. [CrossRef] [PubMed]
36. Wang, J.M.; Chen, X.X.; Cao, L.; An, F.; Chen, B.Q.; Xue, L.F.; Yun, T. Individual rubber tree segmentation based on ground-based LiDAR data and Faster R-CNN of deep learning. *Forests* **2019**, *10*, 793. [CrossRef]
37. Wu, J.T.; Yang, G.J.; Yang, H.; Zhu, Y.H.; Li, Z.H.; Lei, L.; Zhao, C.J. Extracting apple tree crown information from remote imagery using deep learning. *Comput. Electron. Agric.* **2020**, *174*, 105504. [CrossRef]
38. Zhang, C.; Atkinson, P.M.; George, C.; Wen, Z.F.; Diazgranados, M.; Gerard, F. Identifying and mapping individual plants in a highly diverse high-elevation ecosystem using UAV imagery and deep learning. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 280–291. [CrossRef]
39. Brandt, M.; Tucker, C.J.; Kariryaa, A.; Rasmussen, K.; Abel, C.; Small, J.; Chave, J.; Rasmussen, L.V.; Hiernaux, P.; Diouf, A.A. An unexpectedly large count of trees in the West African Sahara and Sahel. *Nature* **2020**, *587*, 78–82. [CrossRef]
40. Braga, J.R.G.; Peripato, V.; Dalagnol, R.; Ferreira, M.P.; Tarabalka, Y.; Aragao, L.; Velho, H.E.D.; Shiguemori, E.H.; Wagner, F.H. Tree crown delineation algorithm based on a convolutional neural network. *Remote Sens.* **2020**, *12*, 1288. [CrossRef]
41. Erdem, F.; Ocer, N.E.; Matci, D.K.; Kaplan, G.; Avdan, U. Apricot tree detection from UAV-images using Mask R-CNN and U-Net. *Photogramm. Eng. Remote Sens.* **2023**, *89*, 89–96. [CrossRef]
42. Perez, M.I.; Karelovic, B.; Molina, R.; Saavedra, R.; Cerulo, P.; Cabrera, G. Precision silviculture: Use of UAVs and comparison of deep learning models for the identification and segmentation of tree crowns in pine crops. *Int. J. Digit. Earth* **2022**, *15*, 2223–2238. [CrossRef]
43. Stumberg, N.; Bollandsås, O.M.; Gobakken, T.; Næsset, E. Automatic detection of small single trees in the forest-tundra ecotone using airborne laser scanning. *Remote Sens.* **2014**, *6*, 10152–10170. [CrossRef]
44. Hauglin, M.; Næsset, E. Detection and segmentation of small trees in the forest-tundra ecotone using airborne laser scanning. *Remote Sens.* **2016**, *8*, 407. [CrossRef]
45. Næsset, E.; Nelson, R. Using airborne laser scanning to monitor tree migration in the boreal-alpine transition zone. *Remote Sens. Environ.* **2007**, *110*, 357–369. [CrossRef]

46. Thieme, N.; Bollandsås, O.M.; Gobakken, T.; Næsset, E. Detection of small single trees in the forest-tundra ecotone using height values from airborne laser scanning. *Can. J. Remote Sens.* **2011**, *37*, 264–274. [CrossRef]
47. Hauglin, M.; Bollandsås, O.M.; Gobakken, T.; Næsset, E. Monitoring small pioneer trees in the forest-tundra ecotone: Using multi-temporal airborne laser scanning data to model height growth. *Environ. Monit. Assess.* **2018**, *190*, 12. [CrossRef]
48. Kaartinen, H.; Hyyppä, J.; Yu, X.W.; Vastaranta, M.; Hyyppä, H.; Kukko, A.; Holopainen, M.; Heipke, C.; Hirschmugl, M.; Morsdorf, F.; et al. An international comparison of individual tree detection and extraction using airborne laser scanning. *Remote Sens.* **2012**, *4*, 950–974. [CrossRef]
49. Priedītis, G.; Šmits, I.; Daģis, S.; Dubrovskis, D. Individual tree identification using combined LiDAR data and optical imagery. *Latv. Univ. Agric.* **2012**, *2*, 7–13.
50. Mongus, D.; Zalik, B. An efficient approach to 3D single tree-crown delineation in LiDAR data. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 219–233. [CrossRef]
51. Qin, Y.C.; Ferraz, A.; Mallet, C.; Iovan, C. Individual tree segmentation over large areas using airborne LiDAR point cloud and very high resolution optical imagery. In Proceedings of the IEEE Joint International Geoscience and Remote Sensing Symposium (IGARSS), Quebec City, QC, Canada, 13–18 July 2014.
52. Qin, H.M.; Zhou, W.Q.; Yao, Y.; Wang, W.M. Individual tree segmentation and tree species classification in subtropical broadleaf forests using UAV-based LiDAR, hyperspectral, and ultrahigh-resolution RGB data. *Remote Sens. Environ.* **2022**, *280*, 113143. [CrossRef]
53. You, H.T.; Tang, X.; You, Q.X.; Liu, Y.; Chen, J.J.; Wang, F. Study on the differences between the extraction results of the structural parameters of individual trees for different tree species based on UAV LiDAR and high-resolution RGB images. *Drones* **2023**, *7*, 317. [CrossRef]
54. Pearse, G.D.; Tan, A.Y.S.; Watt, M.S.; Franz, M.O.; Dash, J.P. Detecting and mapping tree seedlings in UAV imagery using convolutional neural networks and field-verified data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 156–169. [CrossRef]
55. Hao, Z.B.; Post, C.J.; Mikhailova, E.A.; Lin, L.L.; Liu, J.; Yu, K.Y. How does sample labeling and distribution affect the accuracy and efficiency of a deep learning model for individual tree-crown detection and delineation. *Remote Sens.* **2022**, *14*, 1561. [CrossRef]
56. Autio, J.; Heikkinen, O. The climate of northern Finland. *Univ. Oulu Univ. Oulu* **2002**, *180*, 61–66.
57. Forbes, B.C.; Fauria, M.M.; Zetterberg, P. Russian Arctic warming and 'greening' are closely tracked by tundra shrub willows. *Glob. Change Biol.* **2010**, *16*, 1542–1554. [CrossRef]
58. Sutinen, R.; Kuoppamaa, M.; Hanninen, P.; Middleton, M.; Narhi, P.; Vartiainen, S.; Sutinen, M.L. Tree species distribution on mafic and felsic fells in Finnish Lapland. *Scand. J. For. Res.* **2011**, *26*, 11–20. [CrossRef]
59. Liu, Y.; Wang, C.; Xi, X.; Wang, J.; Zhang, H. A pits removal method for LiDAR CHM based on distance weighting and canopy control. *Sci. Surv. Mapp.* **2021**, *46*, 108–113.
60. Jing, L.H.; Hu, B.X.; Noland, T.; Li, J.L. An individual tree crown delineation method based on multi-scale segmentation of imagery. *ISPRS J. Photogramm. Remote Sens.* **2012**, *70*, 88–98. [CrossRef]
61. Serra, J.P. *Image Analysis and Mathematical Morphology*; Academic Press: New York, NY, USA, 1982; p. 610.
62. Soille; Pierre. *Morphological Image Analysis Principles and Applications*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2003; p. 392.
63. Hao, Z.B.; Lin, L.L.; Post, C.J.; Mikhailova, E.A.; Li, M.H.; Yu, K.Y.; Liu, J.; Chen, Y. Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (Mask R-CNN). *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 112–123. [CrossRef]
64. Lucena, F.; Breunig, F.M.; Kux, H. The combined use of UAV-based RGB and DEM images for the detection and delineation of orange tree crowns with Mask R-CNN: An approach of labeling and unified framework. *Future Internet* **2022**, *14*, 275. [CrossRef]
65. Plesoianu, A.I.; Stupariu, M.S.; Sandric, I.; Patru-Stupariu, I.; Dragut, L. Individual tree-crown detection and species classification in very high-resolution remote sensing imagery using a deep learning ensemble model. *Remote Sens.* **2020**, *12*, 2426. [CrossRef]
66. Hu, B.X.; Li, J.L.; Jing, L.H.; Judah, A. Improving the efficiency and accuracy of individual tree crown delineation from high-density LiDAR data. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *26*, 145–155. [CrossRef]
67. Yu, K.Y.; Hao, Z.B.; Post, C.J.; Mikhailova, E.A.; Lin, L.L.; Zhao, G.J.; Tian, S.F.; Liu, J. Comparison of classical methods and Mask R-CNN for automatic tree detection and mapping using UAV imagery. *Remote Sens.* **2022**, *14*, 295. [CrossRef]
68. Dersch, S.; Schöttl, A.; Krzystek, P.; Heurich, M. Towards complete tree crown delineation by instance segmentation with mask R-CNN and DETR using UAV-based multispectral imagery and lidar data. *ISPRS Open J. Photogramm. Remote Sens.* **2023**, *8*, 100037. [CrossRef]
69. Weinstein, B.G.; Marconi, S.; Bohlman, S.A.; Zare, A.; White, E.P. Cross-site learning in deep learning RGB tree crown detection. *Ecol. Inform.* **2020**, *56*, 101061. [CrossRef]
70. Anderson, E.S.; Thompson, J.A.; Austin, R.E. LIDAR density and linear interpolator effects on elevation estimates. *Int. J. Remote Sens.* **2005**, *26*, 3889–3900. [CrossRef]
71. Olsen, R.C.; Puetz, A.M.; Anderson, B. Effects of LiDAR point density on bare earth extraction and DEM creation. In Proceedings of the ASPRS Annual Conference, Baltimore, Maryland, 9–13 March 2009.
72. Reitberger, J.; Schnörr, C.; Krzystek, P.; Stilla, U. 3D segmentation of single trees exploiting full waveform LIDAR data. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 561–574. [CrossRef]

73. Yao, W.; Krull, J.; Krzystek, P.; Heurich, M. Sensitivity analysis of 3D individual tree detection from LiDAR point clouds of temperate forests. *Forests* **2014**, *5*, 1122–1142. [CrossRef]
74. Yin, D.M.; Wang, L. Individual mangrove tree measurement using UAV-based LiDAR data: Possibilities and challenges. *Remote Sens. Environ.* **2019**, *223*, 34–49. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.