An Advanced Terrain Vegetation Signal Detection Approach for Forest Structural Parameters Estimation Using ICESat-2 Data

Yifan Li, Xin Shen and Lin Cao *

Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China; meterwalon@njfu.edu.cn (Y.L.); xinshen@njfu.edu.cn (X.S.)

* Correspondence: lincao@njfu.edu.cn

Abstract: Accurate forest structural parameters (such as forest height and canopy cover) support forest carbon monitoring, sustainable forest management, and the implementation of silvicultural practices. The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2), which is a spaceborne Light Detection and Ranging (LiDAR) satellite, offers significant potential for acquiring precise and extensive information on forest structural parameters. However, the ICESat-2 ATL08 product is significantly influenced by the geographical environment and forest characteristics, maintaining considerable potential for enhancing the accuracy of forest height estimation. Meanwhile, it does not focus on providing canopy cover data. To acquire accurate forest structural parameters, the Terrain Signal Neural Network (TSNN) framework was proposed, integrating Computer Vision (CV), Ordering Points to Identify the Clustering Structure (OPTICS), and deep learning. It encompassed an advanced approach for detecting terrain vegetation signals and constructing deep learning models for estimating forest structural parameters using ICESat-2 ATL03 raw data. First, the ATL03 footprints were visualized as Profile Raster Images of Footprints (PRIF), implementing image binarization through adaptive thresholding and median filtering denoising to detect the terrain. Second, the rough denoising buffers were created based on the terrain, combining with the OPTICS clustering and Gaussian denoising algorithms to recognize the terrain vegetation signal footprints. Finally, deep learning models (convolutional neural network (CNN), ResNet50, and EfficientNetB3) were constructed, training standardized PRIF to estimate forest structural parameters (including forest height and canopy cover). The results indicated that the TSNN achieved high accuracy in terrain detection (coefficient of determination ($R^2$) = 0.97) and terrain vegetation signal recognition (F-score = 0.72). The EfficientNetB3 model achieved the highest accuracy in forest height estimation ($R^2$ = 0.88, relative Root Mean Squared Error (rRMSE) = 13.5%), while the CNN model achieved the highest accuracy in canopy cover estimation ($R^2$ = 0.80, rRMSE = 18.5%). Our results have significantly enhanced the accuracy of acquiring ICESat-2 forest structural parameters, while also proposing an original approach combining CV and deep learning for utilizing spaceborne LiDAR data.

Keywords: forest structural parameters; ICESat-2; deep learning; terrain vegetation signal detection

1. Introduction

The growing global attention on the repercussions of climate change on terrestrial ecosystems is driving governments and individuals worldwide to respond to and mitigate these impacts [1]. The international community has unanimously recognized the promotion of a green and low-carbon economy as a consensus strategy to respond to climate change [2]. Forests significantly influence global carbon cycling and climate change as they form a vital component of terrestrial ecosystems [3]. High-coverage, high-resolution forest structural parameter data play a critical role in estimating forest carbon storage, which is an essential element of terrestrial ecosystem carbon budgeting [4]. Forest structural parameters also serve as crucial indicators for assessing forest growth, evaluating forest health, and determining site quality [5]. Traditional forest resource surveys rely primarily on sampling...
theory and ground surveys, but these methods are labor-intensive, time-consuming, and challenging to implement for continuous sampling over vast areas. Remote sensing and inversion techniques have been widely recognized as an effective means of acquiring continuous forest structural parameter information over extensive regions [6–8].

Light Detection and Ranging (LiDAR) is an emerging active remote sensing technology that efficiently and accurately monitors ecosystems at various spatial scales [9]. It offers strong penetration capabilities, providing rich information on the three-dimensional structure of forest canopies and the intensity of the return. LiDAR data holds great potential in accurately estimating forest biophysical properties and structural parameters [10]. However, LiDAR data is expensive to collect and limited in its coverage, making it challenging to monitor forests on a large scale [11]. With the continuous development of technology, LiDAR has been integrated into satellites to constitute spaceborne LiDAR systems, which provide the ability to systematically sample global forest ecosystems [12].

The National Aeronautics and Space Administration (NASA) launched the ICESat-2 in 2018, equipped with the Advanced Topographic Laser Altimeter System (ATLAS). ICESat-2 incorporates multi-beam, micro-pulse, and photon-counting LiDAR technology [13], measuring ice sheet and glacier elevation change and sea ice freeboard and enabling the determination of the heights of Earth’s forests [14]. ATLAS utilized a more sensitive single-photon detector and a higher pulse repetition rate, capturing smaller and denser photon points with a finer footprint spacing (0.7 m) [15].

Forestry applications of spaceborne LiDAR have seen significant development in recent years. ICESat-2 products were combined with other various data to estimate forest height, canopy cover, and above-ground biomass in many previous studies [16,17]. However, the quality of these ICESat-2 products is not perfect. The ICESat-2 forestry products only focus on the forest height parameters, but the other forest parameters are still blank. The ATL08 product records forest height parameters within footprint segments of 100 m along the track. Queinnec et al. [12] compared the accuracy of ICESat-2 ATL08 forest height measurements in different stand ages, demonstrating that the accuracy is significantly reduced in over-mature stands with complex structure and forest height variability. Existing studies showed that the estimation of forest structural parameters based on ICESat-2 data was strongly reliable, but there was still potential for an improvement in accuracy [16–21]. Also, the weak power and diurnal laser signal from ATLAS significantly impacted the accuracy of forest height estimation [22]. Because of the laser propagation characteristics, complex terrain relief can affect the accuracy of forest height estimation [23]. Additionally, its low resolution also restricts its application in the forestry. Therefore, it is necessary to cluster the primary and raw ATL03 data rather than relying solely on the derived products.

Denoising approaches for spaceborne LiDAR data were classified into three categories: local statistical analysis algorithms, raster image processing algorithms, and point cloud clustering analysis algorithms [24]. Raster image processing algorithms provide new insights for extracting terrain contours from photon-counting LiDAR point cloud data [25]. Prior to the official launch of ICESat-2, Magruder et al. [26], and Wang et al. [27] proposed an improved Canny edge detection denoising algorithm, demonstrating the potential of an image processing method for photon-counting LiDAR point cloud data.

Ye et al. [28] compared the performance of various filtering algorithms on ATL03 product denoising and found that the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Ordering Points to Identify the Clustering Structure (OPTICS) algorithms achieved high accuracy. Zhong et al. [29] proposed a robust algorithm for photon-denoising and bathymetric-estimating combined DBSCAN and two-dimensional window filter, achieving a Root Mean Squared Error (RMSE) value of 0.30 m. However, regardless of the denoising method employed, it is imperative for the original data to accurately represent the real ground objects. Particularly in mountainous and dense forest areas in tropical regions, the penetration ability of LiDAR is finite, making it challenging to completely acquire ground footprints and significantly restricting the forest detection
capability [30]. This underscores the criticality of discerning and excluding data that significantly diverge from reality.

Previous research has combined spaceborne LiDAR data with other types of data to more accurately measure different aspects of forest structure. Wang et al. [31] proposed a hybrid model for estimating forest canopy heights by utilizing fused multimodal spaceborne LiDAR data and optical imagery. Wang et al. [32] extracted spectral features and vegetation indices based on ICESat-2 data and Landsat-8 OLI images and then constructed a random forest model for the inversion of above-ground biomass (AGB). Similarly, ICESat-2 ATL03 footprint data values can also be features of the inversion model.

In the data-rich world, images constitute a substantial portion of all measurements conducted [33]. The advancements in Computer Vision (CV) technology have rendered it applicable across diverse domains. CV technology offers simplicity in execution, robust interpretability, and high performance. Digital image processing techniques originate from enhancing pictorial information for human interpretation [34]. They facilitate the conversion of abstract digital forms into concrete visual feature information. With the development and widespread availability of various image processing toolsets such as OpenCV and scikit-image, techniques continuously improve in efficiency, reduce operational barriers, and lower learning costs, significantly accelerating the industrialization process of these methods [33,35].

Kurdi projects point clouds onto horizontal planes to obtain geospatial information in forestry applications [36]. In this study, according to the strip-like scanning pattern of ICESat-2, Profile Raster Images of Footprints (PRIF) were defined as the two-dimensional spatial distribution maps of the footprints, captured from a viewpoint parallel to the scanning trajectory in close proximity to the Earth’s surface along the satellite’s orbital path. The distance along the track of the footprints was the horizontal coordinate of PRIF, and the elevation was the vertical coordinate. When processing spaceborne LiDAR data, previous studies have hardly considered the spatial distribution characteristics of the PRIF, resulting in a massive loss of visual information. Focusing on this gap, deep learning methods were proposed, based on the regression between PRIF and forest structural parameters. Deep learning is a branch of neural networks encompassing supervised, unsupervised, and reinforcement learning techniques [37]. By incorporating multiple neural network layers, deep learning models can learn hierarchical representations of data and uncover potential patterns and regularities. Deep learning has shown significant advancements in various domains, emphasizing its effectiveness in extracting abstract feature representations from raw input data for better generalization [38,39].

The accuracy of ICESat-2 forest height products and research need to be further improved, and ATL08 products still contain an knowledge gap in canopy cover. This study aimed to explore spatial distribution features based on ATL03 footprints, proposing a framework to detect terrain, recognize terrain vegetation signal footprints, and estimate forest structural parameters. The research assembled the advantages of OPTICS, CV, and deep learning and integrated them into a framework called the Terrain Signal Neural Network (TSNN). The objectives of this study include the following: (1) to develop a terrain detection approach combining the PRIF of ICESat-2 ATL03 data with adaptive thresholding binarization and median filtering denoising to extract the terrain trend line; (2) to construct a terrain vegetation signal footprint recognition approach combining OPTICS clustering and Gaussian denoising algorithms for accurate terrain and vegetation identification; and (3) to construct estimation models by training standardized PRIF and using three deep learning algorithms (convolutional neural network (CNN), ResNet50, and EfficientNetB3) to estimate forest structural parameters (forest height and canopy cover).

2. Materials and Methods

In this section, we begin by presenting an overview of the research area in this paper, followed by an introduction to the primary data utilized. Subsequently, the details of the technical processes and specific methods employed within the study are introduced.
2.1. Study Area

The Gaofeng Forest Farm is located in Xingning District, Nanning City, Guangxi Province (22°51′–23°02′N, 108°06′–108°31′E), situated south of the Tropic of Cancer [40]. The average yearly temperature in the region is approximately 21.6 °C with an annual precipitation of around 1300 mm and an average relative humidity of 79%. The terrain comprises hilly landscapes with elevations ranging from 100 to 460 m and slopes varying from 6° to 35°. The region possesses deep red soil, which is conducive to the growth of subtropical and tropical tree species. Forest coverage in the forestry farm is approximately 87%, with a volume of 2.65 million m³, encompassing an operational land area of 593.3 km². Prominent tree species include *Cunninghamia lanceolata*, *Eucalyptus grandis* × *E. urophylla*, and *Pinus massoniana* [41]. The study area within the state-owned Gaofeng Forest Farm includes the Jiepai and Dongsheng sub-forest farms, with an area of approximately 51.8 km² (Figure 1).

![Figure 1. Summary of the coverage and distribution of ICESat-2 ATL03 data in the study area. (a) Guangxi is located in southwest China and shares borders with Yunnan, Guizhou, Hunan, and Guangdong provinces; (b) Nanning is located in the center of Guangxi, and the Gaofeng Forest Farm is positioned in the middle of Nanning; (c) The topographical layout and data coverage of ICESat-2 ATL03 data (from October 2018 to October 2020) within the study area.](image-url)

2.2. Data Sources and Acquisition

In this section, we primarily introduce the datasets about ICESat-2/ATLAS data and Airborne Laser Scanner (ALS) LiDAR data.

2.2.1. ICESat-2/ATLAS

ICESat-2 is equipped with the laser sensor ATLAS (Table 1, source: NSIDC). It offers a range of products classified into five levels and named from ATL01 to ATL21. Among these products, ATL03 serves as a primary product that combines photon travel time, laser position, and attitude angle data to accurately determine the geodetic measurement...
location of photons received by ATLAS. For this study, 21 pieces of ICESat-2 ATL03 data were selected, covering the period from October 2018 to October 2020 (source: NASA Earthdata Search).

Table 1. ICESat-2 ATLAS main parameters of the platform and sensors.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orbital coverage range</td>
<td>88°S–88°N</td>
</tr>
<tr>
<td>Orbit altitude (km)</td>
<td>500</td>
</tr>
<tr>
<td>Design lifetime (years)</td>
<td>3</td>
</tr>
<tr>
<td>Laser wavelength (nm)</td>
<td>532</td>
</tr>
<tr>
<td>Orbital period (days)</td>
<td>91</td>
</tr>
<tr>
<td>Pulse emission frequency (Hz)</td>
<td>10,000</td>
</tr>
<tr>
<td>Footprint diameter (m)</td>
<td>~17</td>
</tr>
<tr>
<td>Along-track footprint spacing (m)</td>
<td>~0.7</td>
</tr>
<tr>
<td>Number of beams</td>
<td>6 beams, 3 pairs</td>
</tr>
<tr>
<td>Strong-weak beam separation (m)</td>
<td>90</td>
</tr>
<tr>
<td>Beam pair separation (km)</td>
<td>~3.3</td>
</tr>
<tr>
<td>Single pulse energy (J)</td>
<td>175 ± 17, 45 ± 5</td>
</tr>
<tr>
<td>Accuracy of altimetry (m)</td>
<td>0.1</td>
</tr>
</tbody>
</table>

2.2.2. Airborne Laser Scanner LiDAR Data

The ALS point cloud data of Gaofeng Forest Farm was acquired in February 2018 using a Riegl LMS-Q680i laser scanner (Table 2). High-resolution products such as a digital elevation model (DEM), digital surface model (DSM), and canopy height model (CHM) were generated from the ALS data. Additionally, the point cloud data served as the verification data for this study.

Table 2. Airborne LiDAR sensor and flight parameters.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flight altitude (m)</td>
<td>750</td>
</tr>
<tr>
<td>Pulse repetition rate (kHz)</td>
<td>300</td>
</tr>
<tr>
<td>Aircraft speed (km/h)</td>
<td>180</td>
</tr>
<tr>
<td>swath width (m)</td>
<td>402</td>
</tr>
<tr>
<td>Scan frequency (Hz)</td>
<td>80</td>
</tr>
<tr>
<td>Scan angle (°)</td>
<td>±30</td>
</tr>
<tr>
<td>Pulse wavelength (nm)</td>
<td>1550</td>
</tr>
<tr>
<td>Overlap</td>
<td>65%</td>
</tr>
<tr>
<td>Beam divergence (mrad)</td>
<td>0.5</td>
</tr>
<tr>
<td>Average point density (pts/m²)</td>
<td>9.58</td>
</tr>
<tr>
<td>Accuracy of altimetry (mm)</td>
<td>20</td>
</tr>
</tbody>
</table>

To generate a suite of products, the point cloud data acquired from the ALS was preprocessed for denoising [42], classification [43], and normalization. Subsequently, 2 m resolution products such as DEM, DSM, and CHM were generated using a spike-free algorithm [44]. Canopy cover represents the percentage of the forest canopy’s vertical projection in the forested area [45], playing a crucial role in accurate forest management and being an essential factor for estimating forest stock. In this study, the canopy cover was obtained by calculating the ratio of the vegetation count of the first echo (greater than 2 m) to the count of all first echoes [46]; the range of the values is from 0 to 1.

2.3. Summary of the TSNN

The research framework (Figure 2) TSNN for this study involves the following procedures: (1) Terrain detection: Utilizing image binarization with adaptive thresholding and median filtering denoising on ICESat-2 ATL03 footprint data to extract the terrain trend line. (2) Signal footprints recognition: Employing OPTICS and Gaussian denoising techniques
to extract the terrain vegetation signal footprints. (3) Neural network construction: Comparing three neural network models, i.e., CNN, ResNet50, and EfficientNetB3, building a predictive model using standardized PRIF in order to establish a relationship with forest height and canopy cover. The standardization of PRIF aimed to realize the filtering and form unification of PRIF.

Figure 2. The framework of the advanced terrain vegetation signal detection approach for forest structural parameters estimation using ICESat-2 data (PRIF—Profile Raster Images of Footprints; OPTICS—Ordering Points to Identify the Clustering Structure).

2.4. Terrain Detection

ATL03 signal footprints exhibit a more regular and denser distribution compared to noise footprints. However, there is a possibility of encountering areas where the density of noise footprint is even more dense. Previous studies calculated the peak height of elevation histograms, then the rough denoising buffers were generated based on the border of the upper and lower extension (depends on the height of the target feature) [47,48]. This probably results in deviations from the actual distribution range of signal footprints. To solve this problem, a suite of image processing methods to extract the terrain trend line were applied, including grayscale processing, image binarization through adaptive thresholding, and median filtering denoising techniques. Based on the rough denoising result, the footprints were clustered by OPTICS for further denoising. The parameters in
the following steps were optimized for precision through pre-experiments and achieved the highest accuracy in this study.

1. Using the distance along the track (Distance) as the abscissa and the altitude (Alt) as the ordinate, the footprints of each piece of ATL03 data were visualized as raster images. The formula for calculating the Scatter Matrix (SM) is presented in Equations (1) and (2).

\[
SM = \begin{bmatrix}
\sum_{i=1}^{n} f_{(row=0,col=0)}, & \ldots, & \sum_{i=1}^{n} f_{(row=0,col=1)}, & \ldots, & \sum_{i=1}^{n} f_{(row=0,col=x)} \\
\sum_{i=1}^{n} f_{(row=1,col=0)}, & \ldots, & \sum_{i=1}^{n} f_{(row=1,col=1)}, & \ldots, & \sum_{i=1}^{n} f_{(row=1,col=x)} \\
\sum_{i=1}^{n} f_{(row=2,col=0)}, & \ldots, & \sum_{i=1}^{n} f_{(row=2,col=1)}, & \ldots, & \sum_{i=1}^{n} f_{(row=2,col=x)} \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
\sum_{i=1}^{n} f_{(row=ymax,col=0)}, & \ldots, & \sum_{i=1}^{n} f_{(row=ymax,col=1)}, & \ldots, & \sum_{i=1}^{n} f_{(row=ymax,col=x)}
\end{bmatrix}
\]

(1)

\[
f_{(row=y,col=x)} = \begin{cases} 
1 & \text{Footprint is located in } (x,y) \\
0 & \text{Footprint is not located in } (x,y) 
\end{cases}
\]

(2)

where i represents the index of ATL03 footprints, while n represents the count of ATL03 footprints. The variables x and y denote the horizontal and vertical coordinates of the scatter plot within the PRIF, respectively. The symbols xmax and ymax indicate the maximum values of the horizontal and vertical coordinate axes.

2. The histogram of grayscale (GS) values of PRIF was calculated, and an adaptive threshold (AD) was set based on the inflection point of the histogram [49]. The raster images were binarized by the AD, assigning GS pixel values of 255 or 0. A grayscale value of 255 denoted presumed signal pixels (displayed in black color), while a value of 0 indicated presumed non-signal pixels (displayed in white color), as illustrated in Equation (3).

\[
GS_{(x,y)} = \begin{cases} 
255 & \text{GS}_{(x,y)} \geq AD \\
0 & \text{GS}_{(x,y)} < AD 
\end{cases}
\]

(3)

3. To effectively remove salt-and-pepper noise from the binarized images, a suite of median filtering operations was applied iteratively [50]. This approach enhanced the overall quality and clarity of the images by replacing each pixel with the median value of its neighboring pixels.

\[
GS_{(x,y)} = \text{Median} \left( \begin{bmatrix} 
GS_{(x-v,y-v)} & GS_{(x-v+1,y-v)} & \cdots & GS_{(x+v,y-v)} \\
GS_{(x-v,y-v+1)} & GS_{(x-v+1,y-v+2)} & \cdots & GS_{(x+v,y+v)} \\
\vdots & \ddots & \ddots & \ddots \\
GS_{(x-v,y+v)} & GS_{(x-v+1,y+v)} & \cdots & GS_{(x+v,y+v)}
\end{bmatrix} \right), \quad \nu = \frac{m-1}{2}, m = 2k + 1 (k \in \mathbb{Z})
\]

(4)

where m represents the size of the operator in the median filtering algorithm.

4. The columns (col = x) along the track direction were iterated. If the number of presumed signal pixels in each column exceeds a predefined threshold denoted as T in Equation (5), the footprints on the entire column are treated as non-signal pixels and their GS values are changed to 0, as shown in Equation (6).

\[
T = \text{Max} \left( 150, 30 \times \frac{n}{\text{Max(Distance)} - \text{Min(Distance)}} \right)
\]

(5)

\[
GS_{(x,c)} = \begin{cases} 
\sum_{i=1}^{\nu_{\text{max}}} (GS_{(x,i)} = 255) < T \\
0 & \sum_{i=1}^{\nu_{\text{max}}} (GS_{(x,i)} = 255) \geq T
\end{cases}
\]

(6)
5. The images were resampled with a resampling factor denoted as \( w (w = 15) \). The number of presumed signal pixels in each resampled window was then calculated to generate a Resample Matrix (RM) using Equation (7). By extracting the maximum value from the resampled results for each column along the track direction, the distribution of terrain trend points could be determined.

\[
RM = \frac{1}{w^2} \begin{bmatrix}
\sum_{row=1}^{w} \sum_{col=1}^{w} GS_{(row,col)} & \sum_{row=1}^{2w} \sum_{col=1+w}^{w} GS_{(row,col)} & \cdots & \sum_{row=xmax-w}^{2w} \sum_{col=1+w}^{w} GS_{(row,col)} \\
\sum_{row=1+w}^{2w} \sum_{col=1}^{w} GS_{(row,col)} & \sum_{row=1+w}^{2w} \sum_{col=1+w}^{w} GS_{(row,col)} & \cdots & \sum_{row=xmax-w}^{2w} \sum_{col=1+w}^{w} GS_{(row,col)} \\
\cdots & \cdots & \cdots & \cdots \\
\sum_{row=ymax-wcol=1} \sum_{col=1}^{w} GS_{(row,col)} & \sum_{row=ymax-wcol=1+w} \sum_{col=1+w}^{w} GS_{(row,col)} & \cdots & \sum_{row=xmax-wcol=1+w} GS_{(row,col)}
\end{bmatrix}
\] (7)

6. To eliminate the outliers of terrain trend points, adjacent terrain trend points with a vertical coordinate difference greater than 12 were removed. Due to the blank of original ATL03 data or the previous removal of entire columns in step 4, the distribution of terrain trend points along the track direction becomes discontinuous. Therefore, a first-degree linear function was utilized to interpolate the discontinuous part.

7. The pixel boundaries of the terrain trend points were extended by 50 m in both the upward and downward directions to create a buffer zone for the signal footprint distribution.

2.5. Signal Footprints Recognition

Signal footprints recognition was carried out by detecting the terrain trend line, which provided a rough denoising result of the ATL03 footprints. To further enhance the denoising process within the buffer zone of the signal footprint distribution, an improved OPTICS algorithm was applied [48].

The OPTICS algorithm clusters data based on spatial density distribution, similar to the DBSCAN algorithm using the searching area to calculate the relative distances among the scatter points [51]. However, unlike DBSCAN, OPTICS does not explicitly generate clusters but instead arranges objects in the dataset to produce an ordered list from which clusters of any density can be obtained. Figure 3 shows an illustration of the improved OPTICS algorithm.

After that, a normal distribution \( N~(\mu, \sigma^2) \) was fitted to the elevation distribution within the buffer zone. The outliers were removed as noise in Equation (8).

\[
\begin{align*}
\mu &= \frac{1}{n} \sum_{i=1}^{n} Alt_i \\
\sigma &= \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Alt_i - \mu)^2} \\
\text{Outlier} &\in (\infty, \mu - \sigma) \cup (\mu + 3\sigma, +\infty)
\end{align*}
\] (8)

To assess the accuracy of denoising method, three statistical evaluation metrics \( R, P \), and F-score were utilized, quantitatively evaluating the algorithm’s performance. \( R \) measures the percentage of accurately identified valid signal footprints among the total original signal footprints, as shown in Equation (9). On the other hand, \( P \) calculates the ratio of accurately identified valid signal footprints to the count of extracted valid signal footprints, as shown in Equation (10). The F-score combines the meaning of \( R \) and \( P \), as shown in Equation (11). In this study, actual signal and noise footprints were defined based on the ALS DEM and ALS DSM products. Signal footprints were identified as being within the envelopes of the DEM and DSM, while noise footprints were identified as being outside of these envelopes.

\[
R = \frac{TP}{TP + FN}
\] (9)
where TP represents the count of photons correctly recognized as signals, FN represents the count of photons not correctly recognized as signals, and FP represents the count of photons incorrectly recognized as signals.

\[
P = \frac{TP}{TP + FP} \quad (10)
\]

\[
F - score = \frac{2PR}{P + R} \quad (11)
\]

Figure 3. The illustration of the improved OPTICS algorithm. OPTICS was utilized to recognize the terrain vegetation signal footprints in this study. In this figure, the yellow scatter points represent noise footprints outside the buffer of DEM trend line, while the black scatters represent rough denoising footprints. The gray dots are the footprints covered by the searching area. OPTICS was employed to establish a search area (red ellipse areas with the red-colored searching center footprints) and cluster the footprints based on their spatial distribution correlation.

2.6. The Standardization of PRIF

Before constructing the deep learning model, PRIFs based on 30 m segments were created using the signal footprints. Then, the segments with significantly different signal footprint distributions from reality were excluded to prevent interference with the subsequent training. A suite of standardization methods of PRIF were designed to achieve this.

The standardization of PRIF included two primary parts, spatial rearrangement and outlier elimination. In each segment of the PRIF, all footprints were taken at a central location to extract DEM, CHM, DSM, and canopy cover values from ALS data. The ALS extraction value represented the actual forest structural parameters values. Then, the simulated forest structural parameters values were calculated within the segment using the spatial information of ICESat-2 footprints. By comparing the actual and simulated values, the parts of PRIF with significant differences were eliminated. Due to the complexity of environmental factors, there was a large error between some data and the real situation. Validation in conjunction with ALS data products ensured the reliability of PRIF samples for the model’s training.

First, PRIFs were normalized by calculating the RH0 (the best-fitting result for the ALS DEM product with a coefficient of determination \( R^2 = 1.00 \), Mean Absolute Error
(MAE) = 1.50 m) in each 5 m sub-segment. The normalized elevation of PRIF was then defined as to the distance from the RH0 for each footprint. Within a PRIF segment, footprints were rearranged based on the normalized elevation. To simulate the dispersion of footprints, the distance between each footprint and the previous one was calculated. Summing up the distance along the horizontal direction, a new horizontal coordinate for each footprint in the segment was generated. Then, the PRIFs were resampled to a specific size. The result of the spatially rearranged PRIF exhibited a common monotonically increasing curved function (depicted as standardized PRIF in Figure 4).

![Figure 4](image-url)

**Figure 4.** The standardization of Profile Raster Images of Footprints (PRIF) at different forest structural parameters values. (a) The standardization of PRIF for forest height; (b) The standardization of PRIF for forest height. Denoised PRIFs and standardized PRIFs are in pairs derived from the same ATL03 segment in the 2 rows of each sub-figure. Each column signifies the progressive increase in forest structural parameters values from left to right.

To ensure the reliability of the training samples, the outliers were eliminated by comparing the simulated and actual forest height, as well as the simulated and actual canopy cover. The accuracy of extracting canopy top footprints and ground point footprints based on multiple percentile heights was evaluated. Then, the simulated forest height was defined as the difference between the RH100 and RH0 (the best-fitting result for ALS DSM product with $R^2 = 0.99$, MAE = 3.98 m) of ATL03 footprints as the most accurate CHM fitting results. Finally, standardized PRIFs with differences greater than 3 m were eliminated when compared with the simulated and true values (Figure 4a).

Similarly, the PRIFs for canopy cover training were standardized. Considering the characteristics of tree species in the study area, the canopy cover of each segment was estimated by using the ratio of footprints with a relative height greater than 2 m to the total footprints. Any standardized PRIFs with a difference of more than 0.10 between the simulated and actual values were also eliminated (Figure 4b).

### 2.7. Deep Learning Model Construction

With the continuous development and progress of deep learning technology, the structure of deep learning networks has become very rich. However, there are still some challenges of overfitting, gradient vanishing, gradient exploding, and degradation phenomena in deep networks. Complex deep learning models may not necessarily achieve good training performance. In this study, three neural network models with optimized weights were constructed to transform the standardized PRIF, including CNN, ResNet50, and EfficientNetB3.
2.7.1. Convolutional Neural Network

The CNN is one of the most traditional deep learning networks, including an input layer, convolutional layers, sampling layers, fully connected layers, and an output layer. Typically, multiple convolutional and sampling layers are used, connected in an alternating pattern. Each convolutional layer is connected to a sampling layer, followed by another convolutional layer, and so on. In the convolutional layer, each neuron in the output feature map is locally connected to its input, obtaining its input value through a weighted sum of the corresponding connection weights, local inputs, and an offset value. This operation is akin to convolution, hence the name, convolutional neural network [52].

The CNN is integral to the architecture of all deep learning networks. In light of this fact, a simple CNN model was developed that consists of four convolutional layers and two pooling layers. This architecture enables the transformation of input images with specific dimensions and 3 channels into a tensor representation, which is in a computer-readable format containing 64 channels (Figure 5).

![Figure 5](image)

**Figure 5.** The construction of CNN model in this study. The output of each layer is represented by the gray rectangle, while the window size for each convolution step is indicated by the orange rectangle. Additionally, the pink square pyramid illustrates the connection relationship between the convolution window and the corresponding pixels in the output layer.

2.7.2. ResNet50

The development of CNN has led to an increase in the number of convolutional layers, allowing researchers to obtain more profound features. Nonetheless, simply increasing the number of layers may not enhance the learning capacity. When the network reaches a certain depth, the issue of random gradient disappearance arises, resulting in a decline in accuracy. Traditional methods to address this problem involve data initialization and regularization methods, which resolve the gradient disappearance but do not improve network accuracy. The introduction of the residual network addressed the gradient problem. It also improved the expressiveness of the features and enhanced detection or regression performance. Furthermore, the utilization of 1 × 1 convolutions reduces the number of parameters and the computational workload to some extent [53].

The ResNet50 network comprises 49 convolutional layers and one fully connected layer. The first part focuses on input convolution, regularization, activation function, and maximum pooling, without incorporating residual blocks. The second, third, fourth, and fifth parts all integrate residual blocks. In the ResNet50 network configuration, the residual blocks consist of three layers, resulting in a total of 49 convolutional layers. Additionally,
the network possesses a final fully connected layer, bringing the total number of layers to 50, which gives rise to the name ResNet50.

2.7.3. EfficientNetB3

EfficientNet has the aim of developing a more efficient approach, as implied by its name, to enhance existing technological advancements. Typically, models are frequently designed to be excessively wide, deep, or at a high resolution. These characteristics initially contribute to model development but soon lead to saturation, resulting in a model with excessive parameters and diminished efficiency. In contrast, EfficientNet incrementally increases these parameters in a more efficient manner. The EfficientNet suites of models progressively increase in level and complexity, while maintaining a consistent stem structure. The stem structure mainly consists of an input layer, rescaling, normalization, zero padding, Conv2D, batch normalization, and activation (Figure 6).

![EfficientNet Stem Structure](image)

**Figure 6.** The stem structure of EfficientNet.

2.7.4. Training of the Standardized PRIF

Each PRIF was labeled with forest height and canopy cover values. Then, they were converted into tensors. In this study, the standardized PRIFs were divided into training and test datasets using a 7:3 ratio. The performances of the estimation models were validated by comparing them with the actual values derived from ALS products. Finally, the $R^2$, RMSE, and rRMSE were employed to evaluate the precision of the extracted forest structural parameters.

3. Results

This section primarily presents the results obtained from the TSNN framework discussed in this paper. It encompasses the accuracy of ATL03 denoising combined with CV and OPTICS, the estimation accuracy of forest height and canopy cover using PRIF deep learning, and the optimization process of neural network models.

3.1. Terrain Detection

During the terrain detection process, the terrain trend line was extracted from ICESat-2 ATL03 footprint data (Figure 7). The trend line was then compared to the DEM product. The fit of the trend line to the ground was evaluated using the $R^2$ value of 0.97, the RMSE of 10.7 m, and the MAE of 9.32 m (Figure 8).

![Terrain Detection](image)

**Figure 7.** The illustration of terrain detection. (a) Visualization of original ATL03 footprint data: The horizontal axis represents the distance along the orbital direction, and the vertical axis represents the altitude of the footprints; (b) Distribution of the terrain trend line: The horizontal coordinate of an image pixel is represented by X, while the vertical coordinate is represented by Y.
3.2. Signal Footprints Recognition

The rough denoising process involved generating a buffer around the rough signal footprints based on the terrain trend line. When combined with DSM and DEM products, the denoising buffer contained 95.4% of the signal footprints within the boundaries of the DSM and DEM envelopes. The R and P were 0.95 and 0.52, respectively, resulting in an F-score of 0.67. Following the application of OPTICS clustering and Gaussian denoising, the performance of ATL03 footprint accurate denoising significantly improved. The R and P values were optimized to 0.81 and 0.64, respectively, resulting in an increased F-score of 0.72 (Figure 9).

3.3. Deep Learning Model Construction

The forest height and canopy cover were predicted by CNN, ResNet50, and EfficientNetB3 models training by the standardized PRIF in the test dataset. The prediction results of the three models for forest structural parameters are all satisfactory. Despite their differing levels of structural complexity, unexpectedly, the simplest one, the CNN model, exhibited the highest accuracy in canopy cover ($R^2 = 0.80$, $rRMSE = 18.5\%$, Figure 10d) but the lowest accuracy in forest height ($R^2 = 0.81$, $rRMSE = 17.6\%$, Figure 10a). Meanwhile, the EfficientNetB3 with complex structures and richer layers achieved the best performance in

![Figure 8](image1.png)

**Figure 8.** The accuracy assessment of terrain trend line extraction.

![Figure 9](image2.png)

**Figure 9.** The illustration of the signal footprints recognition. The horizontal axis represents the distance along the orbital direction, and the vertical axis represents the altitude of the footprints. The blue data points represent the airborne laser scanner (ALS) digital elevation model (DEM) results, while the green data points represent the ALS digital surface model (DSM) results. (a) The rough denoising result of ATL03; (b) The accurate denoising result of ATL03.

![Figure 10](image3.png)
forest height with an $R^2$ of 0.88, an RMSE of 2.17 m, and an rRMSE of 13.5% (Figure 10c). CNN, as a neural network model with a relatively simple structure, has demonstrated superior prediction capability. This is mainly attributed to the limited availability of feature parameters in standardized PRIF and the simplicity of pixel spatial distribution. Additionally, the scarcity of training images restricted the training of neural networks with extensive layers and complex structures.

![Graphs showing accuracy assessment of extracted forest structural parameters through independent sample testing.](image)

Figure 10. The accuracy assessment of extracted forest structural parameters through independent sample testing. (a–c) Accuracy of forest height using CNN, ResNet50, and EfficientNetB3; (d–f) Accuracy of canopy cover using CNN, ResNet50, and EfficientNetB3.

These experiments effectively demonstrate the potential of deep learning models in predicting forest structural parameters. The standardized PRIFs are capable of providing accurate estimations of forest structural parameters.

3.4. Optimization of Neural Network Model Training

The accuracy of the model is influenced by different parameters to varying degrees. To explore the relationship between parameters and accuracy, a suite of iterative experiments using various parameter gradients was conducted. It suggested that when generating the TSNN, setting the optimal value of each parameter as the initial value for the control variable was recommended. As an illustrative example, the optimization of four parameters for forest height estimation was presented, including learning rate, iteration, batch size, and image size (Figure 11).

![Graphs showing influence of parameters on accuracy.](image)

Figure 11. Cont.
3.4.1. Learning Rate

Setting an appropriate learning rate is vital when training a neural network, as it directly impacts whether a neural network can effectively converge to the global minimum. A higher learning rate can lead to undesirable consequences, causing the loss function to oscillate and preventing the network from reaching the global minimum. Instead, it remains in the vicinity of the global minimum without truly converging. On the other hand, selecting a much smaller learning rate aids in achieving convergence to the global minimum. However, this can be time-consuming, as it involves making minute adjustments to the network’s weights. Thus, training the neural network may require more time.

In this study, multiple iterations to determine the optimal learning rate were performed. The study showed a significant enhancement in the accuracy when the learning rate was controlled within the range of $10^{-5}$ to $10^{-2}$. Notably, setting the learning rate to $5 \times 10^{-4}$ achieved the highest accuracy ($R^2 = 0.81$, RMSE = 2.81 m, and rRMSE = 17.6%) (Figure 11a).

3.4.2. Iteration

The number of iterations, or epochs, indicates how many times the entire training dataset is propagated forward and backward through the neural network. After each iteration, the weights are updated using the back-propagation algorithm. Simply, iteration represents the complete passing time of the dataset during the training process of the neural network.

To determine the appropriate number of iterations during training, changes in training errors and validation errors were observed. When both the training and validation errors exhibit stability or when the validation error starts to increase, it indicates that the model has converged, and training should be stopped at this point.

Examining the results, a significant decrease in the accuracy of the test set was observed when the number of iterations exceeded 10. However, the accuracy reached its peak when the number of iterations was approximately between 150 and 175 (Figure 11b).

3.4.3. Batch Size

The batch size plays a crucial role in neural network training as it determines the duration of each epoch and the smoothness of gradient updates between iterations. Larger batch sizes result in faster training but slower convergence and higher memory usage. Some theories suggest that GPUs perform better with batch sizes that are powers of 2, such as 16, 32, 64, 128, etc. [54]. However, this study indicates that the impact of batch size on the CNN model is insignificant (GPU: NVIDIA GeForce RTX 2060, 6G). Regardless of the batch size, the $R^2$ accuracy consistently exceeded 0.70. However, as the batch size increased, the peak was observed at a batch size of 24 (Figure 11c), and then there was a discernible downward trend in $R^2$ scores.
3.4.4. Size of the PRIF

Before each training process, resizing the PRIFs to the same size is an important part of the standardization method. The impact of PRIF size varies depending on the target. In the case of targets like image learning, where the objective is to detect or identify characteristics such as animals or landscapes, the image size has minimal influence on the model. Thus, resizing the images to a smaller dimension is appropriate, as the focus lies in extracting specific features such as mountains or trees, which can still be accomplished with smaller images. Consequently, the size of the image does not significantly impact feature extraction in these tasks.

However, using excessively large images can lead to larger models. CNNs extract features from images, and the size of the feature maps generated by convolutional layers depends on factors such as the filter size, padding, stride, and input size. In regression tasks, the CNN typically flattens the output before reaching the regression layer. When the input image is too large and the convolutional layers are shallow, the dimension of the flattened layer becomes high and the number of parameters in the final output layer increases.

Unlike most complex images, PRIFs only contain a simple spatial distribution of footprints, which requires low image size to achieve a relatively high interpretability. Therefore, it is not necessary to set the image size to 224 × 224 pixels or higher, such as 640 × 640 pixels. The research result also confirms that the size of the PRIF had minimal impact on the accuracy of the CNN model. Accuracy improved as the size increased from 1 until it reached a certain threshold of 24. Beyond this threshold, further increases in size did not significantly affect accuracy. The accuracy curve exhibits oscillations and eventually declines as the image size increases (Figure 11d).

4. Discussion

Three aspects are discussed based on the advantages, disadvantages, and prospects of TSNN.

4.1. The Flexible TSNN

The TSNN consists of three modules: terrain detection, terrain vegetation signal footprints recognition, and deep learning model construction. In this study, various image processing methods were employed for terrain detection, including grayscale, binarization, resampling, and median filtering. Previous studies commonly utilized elevation histogram statistics to obtain the maximum density window and extract the near-surface buffer. For further denoising, the OPTICS clustering and Gaussian denoising algorithms were combined. Previous research also explored algorithms like the quadtree isolation and DBSCAN algorithms in this context [55,56]. In the section regarding building deep learning models, this paper lists three kinds of neural network models. The applications of CV using various algorithms are becoming more and more advanced; for example, MobileNet, AlexNet, and Transformer can also be used in similar applications.

There have already been many studies on estimating forest structural parameters (forest height and canopy cover) based on ICESat-2 data. We combined several studies with the highest citations in recent years as a comparison [16–20]. This study surpasses previous research benchmarks with an impressive RMSE of 3.34 m for forest height product inversion and interpolation based on ATL08 data, as well as an exceptional $R^2$ value of 0.77 derived from ATL08 photon classification counting. This indicates that the PRIF deep learning method in this study shows a great improvement in accuracy.

The combination of TSNN algorithms shows significant potential in forest structural parameters estimation. The development of forest structure models is a never-ending process toward finding the optimal process for building the predictive model, as any part of TSNN can be replaced with an equivalent module to achieve a similar or even optimized effect. The process has inherent flexibility that allows for a multitude of algorithm combinations, hence its name, which signifies the primary combination of target components. In the future, prediction models for forest structural parameters (e.g., AGB, tree density,
LAI, etc.) can be further derived based on spaceborne LiDAR data, leading to continuous improvement in prediction accuracy.

On the other hand, the single data requirement of TSNN enhances the flexibility and practicality of the framework. Once TSNN is successfully constructed, it is simple to input spaceborne LiDAR footprint data containing location and elevation information to predict various forest structural parameters. This approach not only ensures accuracy but also reduces data restriction, particularly as spaceborne LiDAR satellites are constantly being updated. TSNN can be further developed using spaceborne LiDAR datasets to obtain more kinds of forest parameter information.

4.2. The Predicament of Estimating More Forest Structural Parameters

During the training process, the challenges related to standardization and neural network construction need to be addressed. The deep learning estimation model of forest structural parameters proposed in this study is theoretically suitable for Gaofeng Forest Farm, but its universal applicability to all regions is limited. Different geographical environments exhibit diverse climatic conditions, variations in forest composition, and distinct growth characteristics. To ensure the adaptability of the model to forests in multiple regions, it is necessary to propose an appropriate standardization method, particularly regarding the segment length of training inputs and the resampling size of PRIF. Subsequently, expanding the available data resources can further enhance the robustness and effectiveness of TSNN. This underscores the importance of establishing comprehensive LiDAR data sharing platforms (e.g., G-LIHT, NEON, etc.) globally to facilitate model construction.

Several uncertainties existed in the application of TSNN in this study. First, the validation data were derived from ALS data, instead of ground surveys that directly provide forest structural parameters. Consequently, the generation of digital models like DEM, CHM, and DSM, as well as the extraction of canopy cover and individual trees, might be influenced by the data source. Specifically, the limited penetration capability of large airborne LiDAR in the canopy may result in the loss of certain three-dimensional structural information of the forest. Tree species identification based solely on PRIF appears to be unattainable. While acquiring high-resolution ground or near-ground LiDAR data is beneficial for tree species identification and forest 3D structure analysis, it poses greater challenges for spaceborne LiDAR due to its much larger footprints.

We initially considered utilizing PRIFs to train tree density and AGB. However, we ultimately abandoned this approach due to the uncertainty arising from the aforementioned reasons. The ALS system operates at a flight altitude of 750 m, resulting in relatively low point density and limited laser pulse penetration. Consequently, in the tree density experiment, the point cloud data collected is inadequate for the precise segmentation of individual trees, leading to significant under-segmentation and an underestimation of tree density. Moreover, beyond the uncertainty associated with ALS data acquisition, the virtual plots utilized in this study for tree density training were determined based on the distribution of ICESat-2 footprints. Consequently, certain plots did not exhibit typical plantation forest characteristics, resulting in relatively low tree density. Due to the necessity of tree species identification, this study did not complete forest biomass estimation. For a pure stand, it is simple to utilize the same biomass model to calculate the actual values of some forest biomass samples based on individual tree identification results. However, in complex forest conditions, it is essential to analyze and develop specific biomass prediction models for each distinct tree species.

4.3. The Prospect of Spaceborne LiDAR Data in Forest Structural Parameters Estimation

In recent years, there has been a surge in the popularity of machine learning algorithms, including boosting, support vector machines, and artificial neural networks. As a simulation of biological vision using computers and related devices, the advancement of CV technology has greatly enhanced machines’ ability to recognize images and has a profound impact on various aspects of our lives [57]. As a result, predicting forest height
from visualizations of spaceborne LiDAR footprints has become feasible. Through a suite of pooling and compression methods, computers can extract information from images, enabling the quantification of the spatial distribution characteristics of point clouds that may appear elusive.

The constant advancement of artificial intelligence technology has led to the emergence of new types of AI systems, such as machine learning and deep learning, which continually challenge people’s cognition. Presently, the world has entered the era of large models, where the availability of enormous data and increasing computational capabilities provides the groundwork for their development. In forestry, deep learning applications can leverage multi-source remote sensing data as a benchmark for big data to achieve intelligent monitoring and accurate forecasting. In an era of rapid changes in visual processing frameworks such as Yolo, Transformer, and Mamba, introducing deep learning into forestry applications has great potential.

Based on the high-precision forest structural parameter information obtained from TSNN, various methods can be employed to achieve wall-to-wall continuity estimation (e.g., geospatial interpolation, inversion model prediction). This enables the full utilization of the global observation advantage of spaceborne LiDAR, providing crucial data support for evaluating and supervising forest growth across large areas.

5. Conclusions

Spaceborne LiDAR ICESat-2 data holds significant potential for acquiring forest heights across extensive regions. To improve the accuracy of and application based on these data, a novel framework named TSNN was introduced that integrates CV, OPTICS, and deep learning methods. The objective was terrain detection, terrain vegetation signal footprints recognition, and deep learning model construction for estimating forest structural parameters. The accuracy of terrain detection was found to have an $R^2$ of 0.97. The denoising accuracy also reached a high level ($R = 0.81, P = 0.64$, and $F$-score = 0.72). By employing standardized PRIF for establishing the estimation models, three algorithms with different structural complexity all showed good performance. The CNN model performed best in measuring canopy cover, achieving an $R^2$ of 0.80 and an rRMSE of 18.5%. The EfficientNetB3 model with more complex structures and richer layers achieved the highest accuracy in forest height with an $R^2$ of 0.88, an RMSE of 2.17 m, and an rRMSE of 13.5%.

TSNN provides a new method to extract forest height and canopy cover based on ICESat-2 data, which effectively improves the estimation accuracy and proves the potential of CV and deep learning in forestry applications.

Author Contributions: Conceptualization, L.C.; methodology, Y.L. and X.S.; validation, Y.L. and X.S.; formal analysis, Y.L.; resources, L.C.; data curation, Y.L. and X.S.; writing—original draft preparation, Y.L.; writing—review and editing, Y.L., X.S. and L.C.; visualization, Y.L.; supervision, L.C.; funding acquisition, L.C. 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 (2017YFD0600904), the Natural Science Foundation of Jiangsu Province (BK20220415), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Data Availability Statement: Data is contained within the article.

Acknowledgments: The authors gratefully acknowledge the foresters in Gaofeng Forest for their assistance with data collection and sharing their experiences of the local forests. We also would like to acknowledge the graduate students from the Department of Forest Management at Nanjing Forestry University for providing suggestions to improve this study.

Conflicts of Interest: The authors declare no conflicts of interest.


47. Pang, S.; Li, G.; Jiang, Y.; Chen, Y.; Lu, Y.; Lu, D. Retrieval of forest canopy height in a mountainous region with ICESat-2 ATLAS. *For. Ecosyst.* **2022**, *9*, 100046. [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.