Article

A Mixed Broadleaf Forest Segmentation Algorithm Based on Memory and Convolution Attention Mechanisms

Xing Tang, Zheng Li, Wenfei Zhao, Kai Xiong, Xiyu Pan and Jianjun Li *

Faculty of Electronic Information and Physics, Central South University of Forestry and Technology, Changsha 410004, China; 20221100379@csuft.edu.cn (X.T.); 20231200594@csuft.edu.cn (Z.L.); znlyzwf@csuft.edu.cn (W.Z.); 20231100410@csuft.edu.cn (K.X.); 20231100392@csuft.edu.cn (X.P.)

* Correspondence: lijianjun_21@163.com

Abstract: Counting the number of trees and obtaining information on tree crowns have always played important roles in the efficient and high-precision monitoring of forest resources. However, determining how to obtain the above information at a low cost and with high accuracy has always been a topic of great concern. Using deep learning methods to segment individual tree crowns in mixed broadleaf forests is a cost-effective approach to forest resource assessment. Existing crown segmentation algorithms primarily focus on discrete trees, with limited research on mixed broadleaf forests. The lack of datasets has resulted in poor segmentation performance, and occlusions in broadleaf forest images hinder accurate segmentation. To address these challenges, this study proposes a supervised segmentation method, SegcaNet, which can efficiently extract tree crowns from UAV images under natural light conditions. A dataset for dense mixed broadleaf forest crown segmentation is produced, containing 18,000 single-tree crown images and 1200 mixed broadleaf forest images. SegcaNet achieves superior segmentation results by incorporating a convolutional attention mechanism and a memory module. The experimental results indicate that SegcaNet's $mIoU$ values surpass those of traditional algorithms. Compared with FCN, Deeplabv3, and MemoryNetV2, SegcaNet's $mIoU$ is increased by 4.8%, 4.33%, and 2.13%, respectively. Additionally, it reduces instances of incorrect segmentation and over-segmentation.

Keywords: crown segmentation; algorithm for mixed forest; convolutional attention mechanism; MemoryNetV2 algorithm

check for updates

Citation: Tang, X.; Li, Z.; Zhao, W.; Xiong, K.; Pan, X.; Li, J. A Mixed Broadleaf Forest Segmentation Algorithm Based on Memory and Convolution Attention Mechanisms. *Forests* **2024**, *15*, 1310. https://doi.org/10.3390/f15081310

Academic Editor: Eric Casella

Received: 2 July 2024
Revised: 21 July 2024
Accepted: 23 July 2024
Published: 26 July 2024

1. Introduction

1.1. Research Significance and Background

Surveying forest resources and information simply and efficiently has always been crucial for smart forestry. Forest management aims to maximize ecological benefits while maintaining forest ecology [1,2].

This study focuses on how to inventory forest resources in a cost-effective manner. Compared with point cloud and ground data acquisition, UAV data acquisition is less expensive. The number of trees can be counted by using deep learning algorithms to segment individual tree crowns in UAV images. Combined with flight parameters such as geographic location and flight altitude, this method can provide strong support for forest resource investigations.

Many forests are geographically remote with a complex topography, making it difficult and expensive to collect relevant information through manual surveys [3]. Therefore, the image data captured by UAVs are of great significance to forest resource assessment. In recent years, UAV remote sensing has achieved remarkable results in mapping tree information [4,5]. Specifically, UAVs can quickly and inexpensively acquire data such as the tree location and crown width [6], which can be used to operate and manage forests more effectively [7].

For forest resource assessment, the core objective is to survey forest resources at the lowest cost while achieving the highest economic and ecological benefits [8,9]. Parameters such as the number of trees, species, and vegetation closure are crucial for forest resource assessment [10,11]. Orthophoto maps based on UAV photography are an essential method for obtaining these parameters. Compared with point cloud data obtained using LiDAR and high-altitude remote sensing images, using UAVs to capture images is more convenient, and their processing is less costly. The first step in extracting crown information from UAV images involves identifying distinct features. However, challenges such as tree shading and differences between the background and target pose significant obstacles to accurate segmentation.

1.2. Research Status of Crown Segmentation Using Machine Learning and Deep Learning

Optical image data include satellite images and UAV images, which are used for different purposes based on their scales. Satellite images cover larger areas and are commonly used to distinguish forest land from non-forest land. In contrast, UAV images have a smaller scale but contain more detailed texture information, making them ideal for the recognition and segmentation of individual trees.

The traditional individual tree segmentation algorithm [12] is divided into two steps: the center point of the tree contour is identified, and then the tree contour is delineated based on this center point. Numerous segmentation algorithms have been developed based on this approach.

Kestur [13] applied the watershed algorithm and the principle of distance transformation to extract tree crown information based on the traditional method. Wang [14] utilized the maximum suppression and distance transformation methods to detect the center point of a tree outline and then used the watershed algorithm to extract the crown boundary. Jing [15] processed an image using Gaussian filtering and then extracted the crown information using the watershed algorithm after obtaining a binary image.

Traditional segmentation algorithms are often applied to scenarios with low crown cover, but they are not effective in cases with high crown cover. Wu [16] used Fast R-CNN to detect apple tree crowns and then segmented them using the U-Net network, achieving a precision of 91.1% and a recall of 94.1%. Zhang [17] improved the feature fusion method in Mask R-CNN and introduced boundary-weighted loss in the loss function to segment different tree species. Xue [18] applied an improved DeepLabv3 [19] for citrus tree crown segmentation, achieving a faster inference speed and smaller parameter counts with the lightweight MobileNetV3 network. Yan [20] annotated tree crowns in WorldView3 high-resolution images and utilized multiple CNN models for recognition, achieving an accuracy of 82.7%.

Facing the challenges of obtaining datasets for tree crown segmentation, Weinstein [21] proposed a semi-supervised deep learning approach. This method helps mitigate the lack of research data in the field of tree crown detection. Braga [22] developed a method for creating datasets by using manually extracted tree crowns as samples. These samples are randomly placed in the background of high-resolution satellite images, allowing for the large-scale batch production of datasets.

At present, most datasets used for crown segmentation focus on single-tree segmentation, and there is a lack of data for mixed broadleaf forest segmentation. In dense mixed broadleaf forests, segmentation inaccuracies caused by crown overlap have been persistent challenges. The research on segmentation algorithms for dense broadleaf forests remains insufficient [23,24].

1.3. Primary Research Focus

To perform effective segmentation, appropriate methods are needed to distinguish the boundaries of different tree crowns. Existing segmentation methods can be categorized into threshold [25], color, and learning-based segmentation methods [26].

Threshold- and color-based segmentation methods, such as the watershed algorithm, perform segmentation by setting a threshold, measuring the distance between the starting point and surrounding pixels. Previous studies have shown that color is a more efficient feature for differentiating between plants and backgrounds in images [27].

Vegetation can be distinguished from non-vegetation by examining the pixel differences between color channels. Better results are often achieved by using multiple color spaces and selecting the optimal pairings among them [28–30]. This method is suitable for images where the background is differentiated from the target. However, it performs poorly on more complex images.

Learning-based supervised and unsupervised segmentation methods can address the shortcomings of threshold- and color-based methods, handling a variety of complex conditions. Unsupervised learning segmentation methods can segment trees without labeled data but often face challenges with image tilt, variable lighting, and other complex scenarios. Supervised learning methods require a large amount of labeled data for training. However, there is a lack of public datasets for crown segmentation in mixed forests.

To address the shortcomings of supervised learning methods, we developed a crown image segmentation algorithm with high accuracy and strong generalization. This algorithm was designed to extract crown information from UAV-captured images of mixed broadleaf forests. Additionally, we created a dataset specifically for crown segmentation in mixed broadleaf forests. The specific contributions of this study are as follows:

(1) A dataset is created for crown segmentation, containing approximately 18,000 single-tree crown images and 1200 mixed forest images.
(2) A semantic segmentation network, SegcaNet, is proposed for segmenting crowns based on convolutional attention and memory mechanisms.

Pre-training has been proven to accelerate model convergence and improve performance. When building our model, we incorporated the concept of pre-training. In the first stage, single-tree images were used for pre-training to capture local features and eliminate interference from the background and other factors. In the second stage, global images were used for training to acquire global features and enhance the model's actual performance. During data processing, we cropped the images taken by drones into a specified size, selected high-quality images for annotation, and obtained 18,000 single-tree canopy images and 1200 densely mixed broadleaf forest images.

2. Materials and Methods

2.1. Research Area and Equipment

The experimental location is a mixed broadleaf forest area in Xishan, Yunnan (latitude 25°07′ N and Longitude 102°62′ E). Xishan in Yunnan Province has a subtropical plateau monsoon climate and a good ecological environment, which is suitable for broadleaf forest growth. The data collection area is flat and has a simple topography; its location is shown in Figure 1.

A DJI M300RTK(DJI Company, Guangdong, China), equipped with a DJI H20T four-sensor spectral camera, was used to collect data. The DJI H20T includes four types of cameras, including wide-angle and zoom cameras operating in the 400–700 nm wavelength range.

The flight started at 2 p.m., maintaining an altitude of 130 m and a speed of 5 m/s. The forward overlap was 80%, and the lateral overlap was 75%. To maximize the quality and clarity of the captured images, the data were collected in clear and windless weather, covering many areas of the mixed broadleaf forest. The equipment is shown in Figure 2.

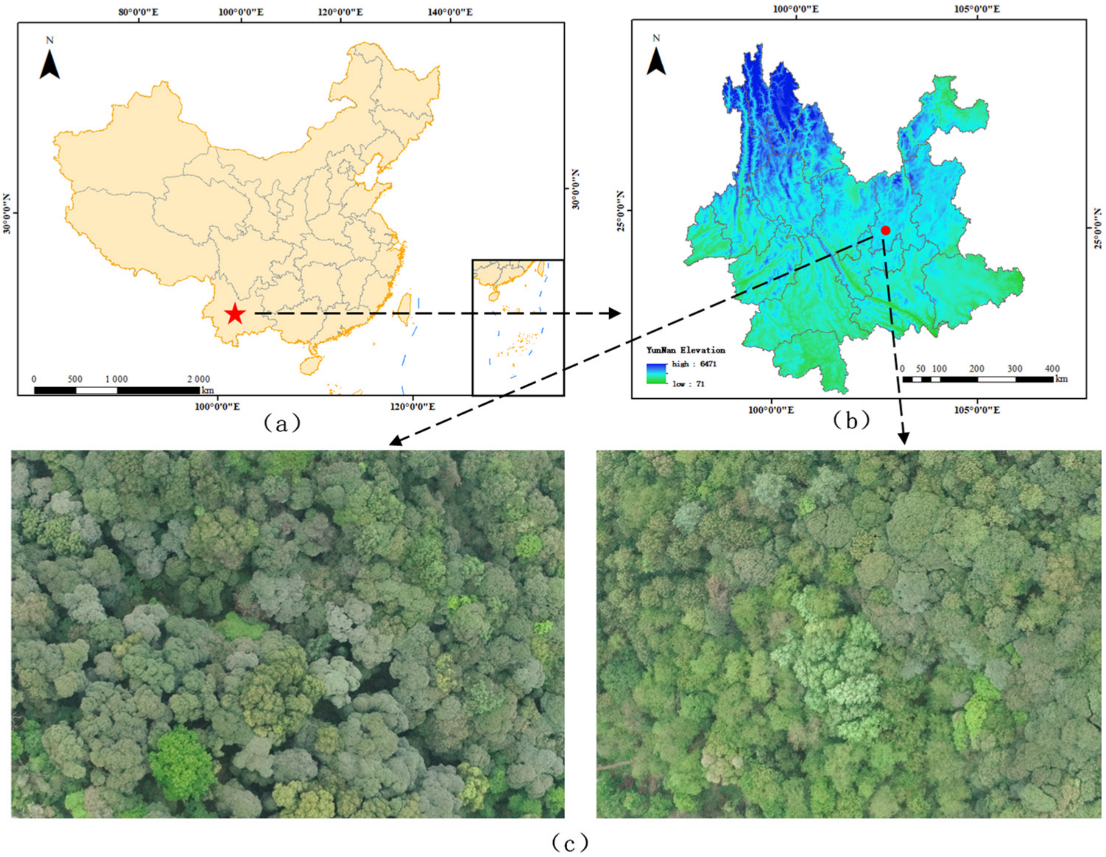


Figure 1. (**a**) Study area in China; (**b**) elevation and topography of Yunnan region; (**c**) images of mixed forests captured using UAVs.

Figure 2. (**a**) UAV DJI-M300RTK; (**b**) DJI-Zenmuse H20T spectral camera.

2.2. Datasets and Image Preprocessing Methods

Images of mixed forests obtained using UAVs under natural conditions often exhibit highly covered crowns. Due to the flight angle, natural conditions, and tree growth state, the crowns can appear tilted. Therefore, it is necessary to preprocess the images. We selected high-quality images from a large collection as benchmark images and labeled them.

Figure 3 shows some of the datasets and labels used for training. Due to the mixed forest region, complex background content, high pixel similarity, and blurred crown edges, segmentation can easily lead to confusion. After labeling the images, two sets of data were used for training. The contours of individual tree crowns (about 18,000) were extracted from the labeled data. These individual tree crowns were fused into images with dimensions of 960 × 960 and used for pre-training.

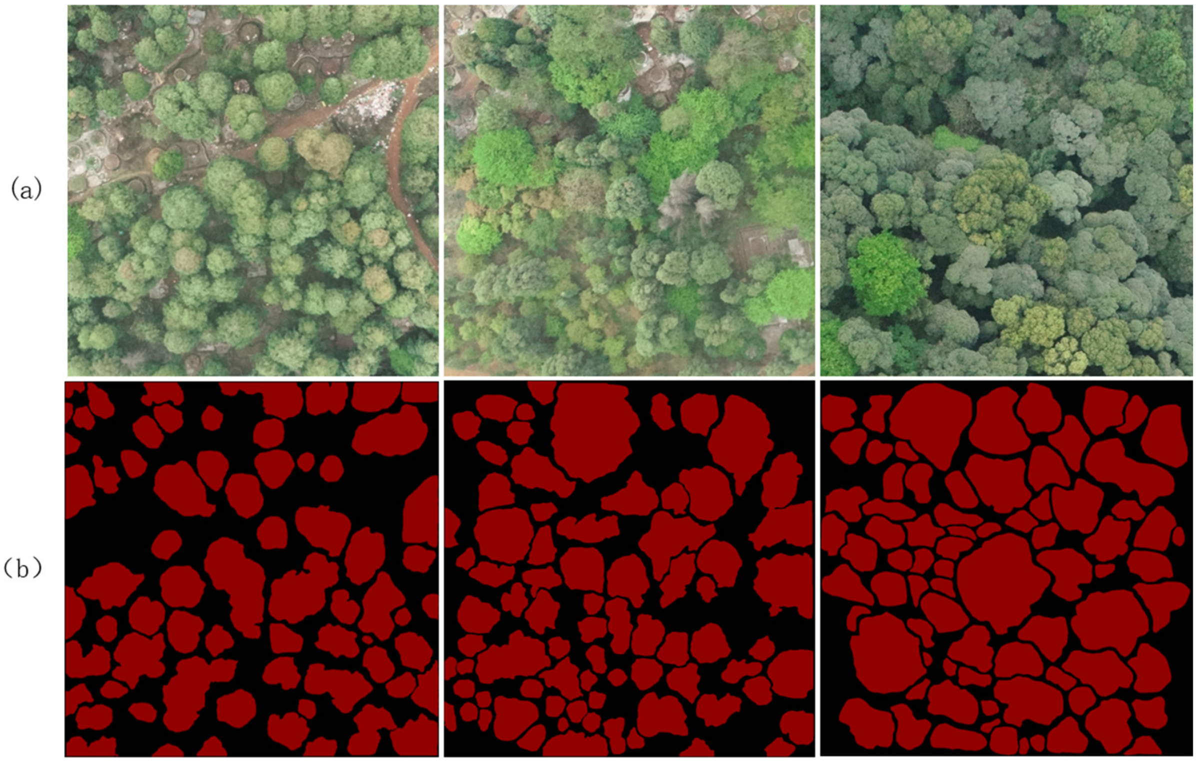


Figure 3. Selected broadleaf forest images taken by UAV: (**a**) original image; (**b**) labeled image.

Based on the pre-training, the global images were used for model training. The background information of the global mixed forest images is complex, and the background pixels have a high similarity to the crown pixels, which may have caused interference in training. However, these global images also contain global semantic information.

The small regions that include only single-tree crowns contain less interference but lack global semantic information. Based on these considerations, images containing only single-tree crowns were used in the pre-training phase to learn the features of the tree crowns. In the second phase, the global images of the mixed forest were used for training.

In mixed forests, although the canopies of different tree species tend to be occluded, resulting in a complex pattern of crown shapes, the crowns still tend to favor specific shapes. In addition, the color and crown boundaries of different tree species are also important features for distinguishing trees in a mixed forest region.

3. Research Methods

3.1. Overall Workflow

The tree crown segmentation algorithm includes five parts: data preprocessing, dataset allocation, network training, crown extraction, and accuracy evaluation. As shown in Figure 4, the orthorectified image is first cropped and processed, from which high-quality images are selected for labeling. A single-tree crown is extracted from the labeled images and fused into a pre-training dataset. Using the pre-trained model, the global mixed forest image is input into the SegcaNet network for training. The iteration with the optimal parameters is selected to examine the training accuracy and results. The overall workflow is shown in Figure 4.

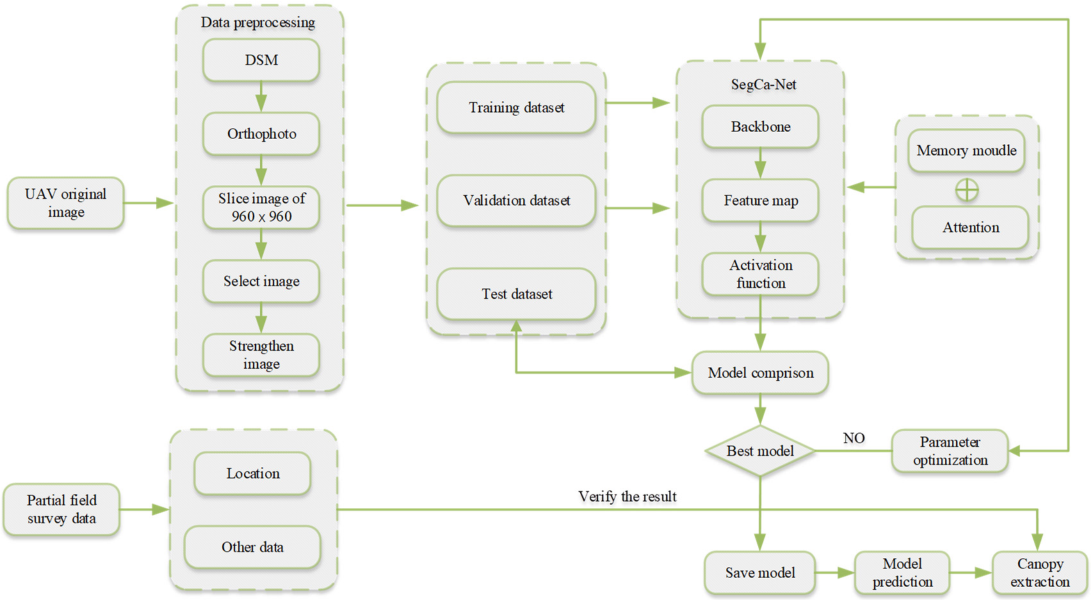


Figure 4. Overall workflow diagram.

3.2. Introduction to SegcaNet Network Architecture

MemoryNetV2 [31,32] is a highly effective semantic segmentation network that enhances model segmentation by reducing the distance between the same classes through a memory module and a cross-image information mining module. The memory module stores the historical information of each category, and after aggregating this information into the probability distribution, it can calculate the distance between instances of the same class. SegcaNet is an improvement based on MemoryNetV2, which enhances the model by using a convolutional attention mechanism and a partially looped feature pyramid.

As shown in Figure 5, after an image is fed into the backbone network, the generated feature vectors are processed in different branches, some of which are fed into Cloformer [33] after ASPP [34], while others are fed into Cloformer through the memory module and the cross-image information mining module. Finally, these two parts, together with the feature vectors directly output from the backbone, are fused and fed into the partially looped feature pyramid.

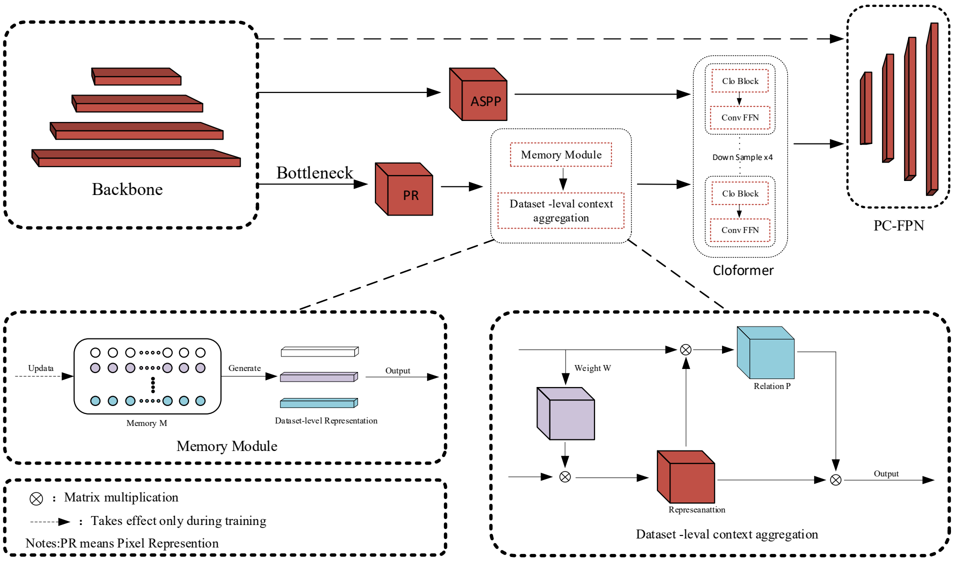

Backbone
Bottleneck
ASPP
PR
Memory Module
Dataset -leval context aggregation
Clo Block
Conv FFN
Down Sample x4
Clo Block
Conv FFN
Cloformer
PC-FPN
Updata
Memory M
Generate
Output
Dataset-level Representation
Memory Module
⊗ : Matrix multiplication
: Takes effect only during training
Notes:PR means Pixel Represention
Weight W
Relation P
Represeanattion
Output
Dataset -leval context aggregation

Figure 5. Diagram of SegcaNet network structure.

3.3. Network Improvements

Improving the receptive field of the network and obtaining multi-scale feature vectors are effective means of enhancing the network performance. Since the produced dataset contains forest images at different scales, there is a potential dataset dispersion problem. Relying solely on the MemoryNetV2 memory module may not allow for enough dataset-level semantic distribution information to be collected, leading to poor segmentation in the forest images at different scales. Therefore, the following modifications are made to the network:

(1) In the original MemoryNetV2 network, the feature vectors output by the backbone network are directly concatenated after passing through the dilated convolution and memory modules. However, in complex broadleaf forest areas, there are many trees with different scales, shapes, and features. The samples to be learned contain a large number of complex samples, and the original features are not at sufficient scales. The feature fusion is not sufficient for learning the more complex tree crowns. By using Cloformer with the convolutional attention mechanism after the dilated convolution and cross-image information mining modules, the feature extraction ability is enhanced. The method of feature fusion is also changed, replacing the original concatenation process with a partially looped feature pyramid. Compared with the traditional top-down and bottom-up fusion methods, this bi-directional propagation feature pyramid can better fuse global and local information.

(2) Cross-entropy is a loss function commonly used in deep learning [35]. According to the different classes in the network, the cross-entropy loss function converts the output into a probability between 0 and 1.

The cross-entropy loss function is defined as

$$L_{cr} = -p_t \log(\overline{p}_t) - (1 - p_t)\log(1 - \overline{p}_t) \tag{1}$$

However, cross-entropy loss in the crown segmentation task reduces the boundary accuracy of the crown. Additionally, due to the irregular and complex samples in the dataset, the learning difficulty is increased. To address these issues, the Focal loss function [36] is introduced into the loss function. This Focal loss function sets a dynamic scaling factor. It increases the focus on complex samples, addressing the imbalance of easy and difficult samples during single-stage training.

The specific definition of the loss function is as follows:

$$L_{FL} = -\alpha_t (1 - p_t)^{\gamma} \log p_t \tag{2}$$

In the above loss function, the cross-entropy loss function is used to address the imbalance of positive and negative samples, while the Focal loss function distinguishes between simple and complex samples.

The final form of the focal loss function is as follows:

$$\begin{cases} L_{FL} = -\alpha_t (1 - p_t)^{\gamma} \log p_t & y = 1 \\ L_{FL} = -\alpha_t (1 - p_t)^{\gamma} \log(1 - p_t) & y = 0 \end{cases} \tag{3}$$

Based on the above considerations, we combined the Focal loss function and the cross-entropy loss function by setting a weight factor. This approach addresses the sample imbalance issue in the dataset.

The final loss function is defined as

$$L_{loss} = \varepsilon L_{cr} + (1 - \varepsilon) L_{FL} \tag{4}$$

In Formulas (1)–(4), p_t represents the probability of the class, α_t represents the dynamic scaling factor, L_{cr} represents the cross-entropy loss function, L_{FL} represents the Focal loss function, and ε represents the weight factor. The range of ε and α_t is between 0 and 1.

3.4. Partial Looped Feature Pyramid

FPN [37] improves the segmentation accuracy of the network by constructing a multi-scale pyramid structure to obtain multi-scale semantic information. In the network proposed in this paper, the idea of the feature pyramid is utilized. The feature map extracted from the backbone is input into the lateral layers, and a feature pyramid with partial loops is used to pay more attention to the detailed parts of the semantic information. Then, top-down fusion is performed on the lateral layers.

As shown in Figure 6, the construction of the feature pyramid mainly utilizes two sets of information. One set contains the feature vector directly from the input of the backbone. The other set contains the feature vector processed by the memory module and Cloformer. The loops of the partially looped feature pyramid are located in the second and third layers of the feature pyramid. In addition to the direct input from the backbone, these two layers contain as much detailed information as possible, for which loop feature fusion can better process the detailed information contained in the image.

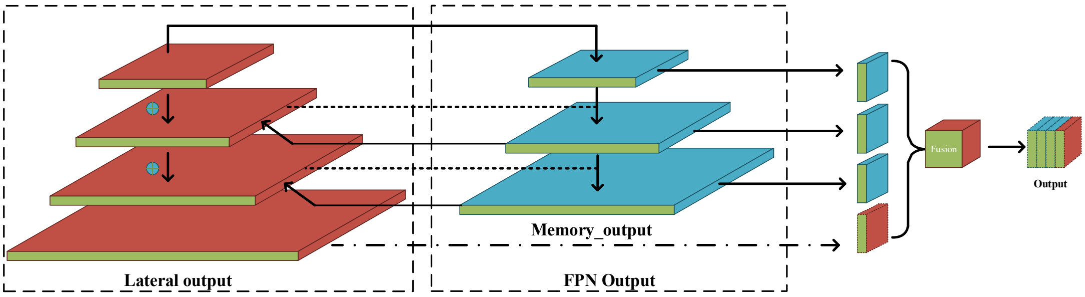


Figure 6. Partially looped feature pyramid.

3.5. Cloformer

Usually, the following two aspects can be considered to improve the performance of a segmentation network: one is the enhancement of the feature extraction capacity by changing the backbone of the network, and the other is the improvement of the model's performance by enhancing its ability to mine contextual information. The attention mechanism selectively focuses on more important semantic information and allocates limited computational resources to more valuable semantic regions. This can enhance the model while keeping the number of parameters and network layers relatively small.

Cloformer introduces a convolutional attention module that combines the attention mechanism with convolutional operations to better capture local information within the image. By using shared weights and context-aware weights, it can better handle the relationship between features at different locations in the image. The principle of the convolutional attention module is shown in Figure 7.

The classic Transformer generates Q, K, V vectors after linear transformations [38], and AttnConv adopts this same concept. As shown in Figure 7, the generated Q, K, V vectors share the weights of the depthwise separable convolution after the linear transformations. Q, K, V all utilize these shared weights to aggregate local information. The difference is that Q and K, after aggregating local information, calculate the Hadamard product. The Hadamard product is then processed through a fully connected layer and an activation function to generate context-aware weights.

$$\begin{gathered} Q, K, V = FC(Input) \\ Q_1, K_1, V_1 = DWconv(Q, K, V) \\ Attn = Tanh\left(\frac{FC(Swish(FC(Q_1 \odot K_1)))}{\sqrt{d}}\right) \end{gathered} \tag{5}$$

In Formula (5), FC and $DWconv$ represent the fully connected layer and depthwise separable convolution, respectively. Q, K, V represents the vectors after the linear transformations, Q_1, K_1, V_1 represents the vectors after the depthwise separable convolution,

$Q_1 \odot K_1$ represents the Hadamard product of Q_1, K_1, *Swish* and *Tanh* represent the activation function, and d represents the number of channels of the token.

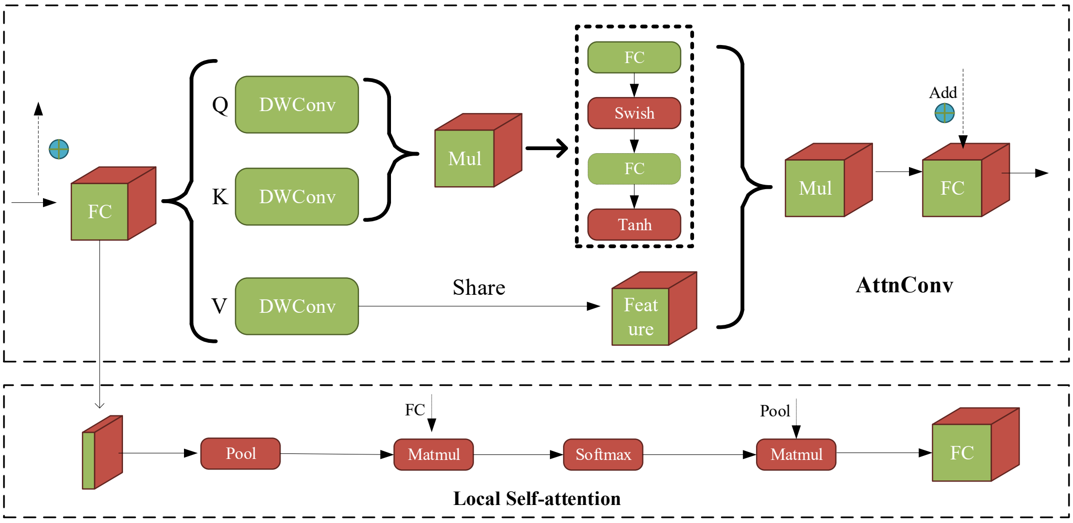


Figure 7. AttenConv+Local self-attention. Mul represents the Hadamard product, pool represents average pooling, Matmul represents matrix multiplication, FC represents the fully connected layer, and Feature refers to the feature map. Softmax is the activation function. The remaining parts are explained in Equation (5).

Due to the use of depthwise separable convolution in the convolutional attention module, the amount of computation is largely reduced. The difference between ConvFFN and FFN is that the deep convolution after the activation function can efficiently aggregate local information, allowing for direct downsampling without the need for additional operations. The principle of ConvFFN is shown in Figure 8.

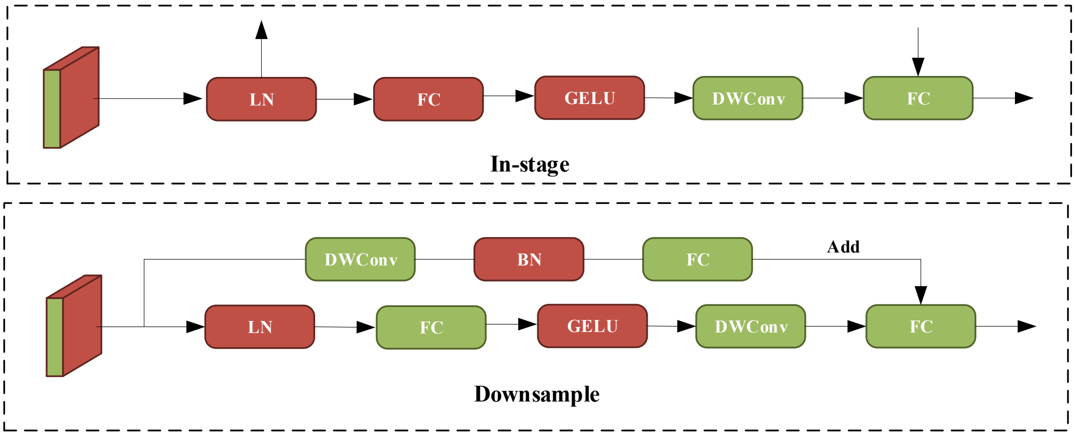


Figure 8. ConvFFN schematic. LN and BN denote layer normalization and batch normalization. GELU is an activation function.

3.6. Introduction to Backbone Structure

Many neural networks use ResNet, UNet [39,40], and other backbone networks as their feature extraction networks and have achieved good results. However, with the development of deep learning, there are numerous outstanding feature extraction networks that have been validated on several public datasets.

To achieve the goals of forest management, it is essential to accurately separate targets in mixed broadleaf forest areas. The segmentation results are used as the basis for forest operations and management. Unsatisfactory segmentation results can easily lead to incorrect decisions in forest management. In the network proposed in this paper, Swinbase-Transformer [41], which has a strong feature extraction ability, is used as the backbone network. This backbone network has a superior performance to other backbone networks.

3.7. Evaluation Metrics

The overall performance of the segmentation network can be evaluated using several metrics, including recall (Re), precision (Pr), mean Intersection over Union ($mIoU$), and balanced F-score (F_1). These metrics are used to test the effectiveness of the model. The calculation methods for each metric are as follows:

$$\begin{aligned} \text{Re} &= \frac{\sum TP}{\sum TP + \sum FN} \times 100\% \\ Pr &= \frac{\sum TP}{\sum TP + \sum FP} \times 100\% \\ F_1 &= \frac{2 \times Pr \times \text{Re}}{Pr + \text{Re}} \times 100\% \\ mIoU &= \frac{1}{k+1} \times \sum \frac{\sum TP}{\sum TP + \sum FN + \sum FP} \times 100\% \end{aligned} \tag{6}$$

In Formula (6), TP represents the number of pixels correctly segmented into the crown region, TN represents the number of pixels correctly segmented into other regions, FP represents the number of pixels incorrectly segmented into the crown region, and FN represents the number of pixels incorrectly segmented into other regions. $k + 1$ represents the number of categories; in this paper, $k = 1$.

4. Result

4.1. Comparative Experiment

To assess the model's ability to segment tree crowns of varying densities, several comparative experiments were conducted. Table 1 compares FCN, Deeplabv3, MemoryNetV2, and SegcaNet. The results show that SegcaNet outperforms the other models.

Table 1. Experimental results of different methods.

Segmentation Methods	Backbone	$Re(\%)$	$Pr(\%)$	$mIoU(\%)$	$F_1(\%)$
Threshold Segmentation	/	/	/	/	/
FCN	Swin	89.26%	78.37%	77.56%	83.46%
Deeplabv3	Swin	92.31%	77.23%	78.03%	84.10%
MemoryNetv2	Swin	91.51%	84.44%	80.23%	87.83%
SegcaNet	Swin	91.68%	85.27%	82.36%	88.35%

From the results in Table 1, it can be seen that the traditional segmentation algorithms, such as the watershed algorithm, do not perform well. For the more complex situation of a mixed broadleaf forest, it is difficult to differentiate the boundaries between the crowns. When using the traditional watershed algorithm, the ground, grass, and other non-crown areas are incorrectly labeled as crowns, and the boundaries between the segmented crowns are not clear.

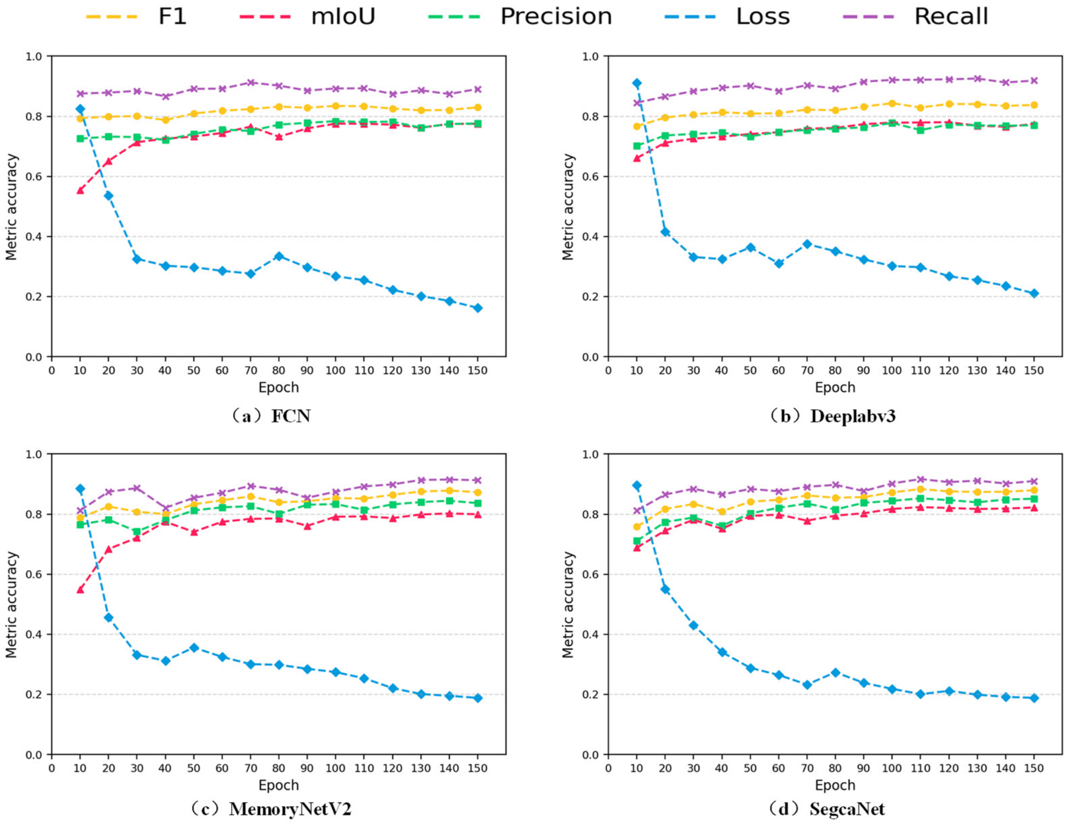
The deep learning-based image segmentation algorithm performs well. The FCN algorithm replaces the fully connected layer with a fully convolutional layer. It upsamples using inverse convolution for pixel-by-pixel classification. This approach relaxes the input image size requirement. Compared with the watershed algorithm, FCN significantly reduces erroneous segmentation. However, FCN mainly utilizes local information for prediction, leading to local discontinuities in the segmentation results. The Pr and $mIoU$ achieved by FCN are 78.37% and 77.56%, respectively.

The DeepLabv3 model meets the basic requirements for crown segmentation, effectively distinguishing most of the crowns. However, it is limited by its network architecture, which lacks sufficient shallow features and detailed semantic information. This limitation leads to an inadequate recognition of crown edges and details. As a result, its Pr and $mIoU$ are 78.03% and 77.23%, respectively.

MemoryNetV2 achieves better results than FCN and Deeplabv3. The memory module store historical information during training, while the cross-image information mining module obtains more information, enabling the dynamic use of data. However, due to the large number of images at different scales and under various natural conditions in the dataset, MemoryNetV2 still occasionally misidentifies the background as a tree crown. Its Pr and $mIoU$ are 84.44% and 80.23%, respectively.

The SegcaNet segmentation network proposed in this paper introduces the convolutional attention mechanism in Cloformer. This mechanism, which benefits from depthwise separable convolution, effectively reduces the computational effort of the network while ensuring excellent results. SegcaNet uses Swinbase as the backbone of the network and utilizes approximately 18,000 single-tree crown images for pre-training. The resulting weight files are then used to train the model. According to the experimental test data, the model achieves excellent results, with Re, Pr, $mIoU$ recall, and F_1 scores of 91.68%, 85.27%, 82.36%, and 88.35%, respectively. Compared with the other models, SegcaNet can effectively differentiate crown edges and details, and it can reduce the incorrect classification of non-crown areas as crown areas.

As shown in Figure 9, the curve fluctuations during the iteration process of FCN and MemoryNetV2 are relatively strong, while those of SegcaNet and Deeplabv3 are relatively gentle. As the training proceeds, the total loss gradually decreases, indicating that the network gradually learns the required features. Due to network differences, FCN and MemoryNetV2 exhibit more fluctuations, but, after more than 120 epochs, all networks tend to converge, and the loss approaches the minimum. Among all the compared networks, SegcaNet demonstrates the best performance.



(a) FCN

(b) Deeplabv3

(c) MemoryNetV2

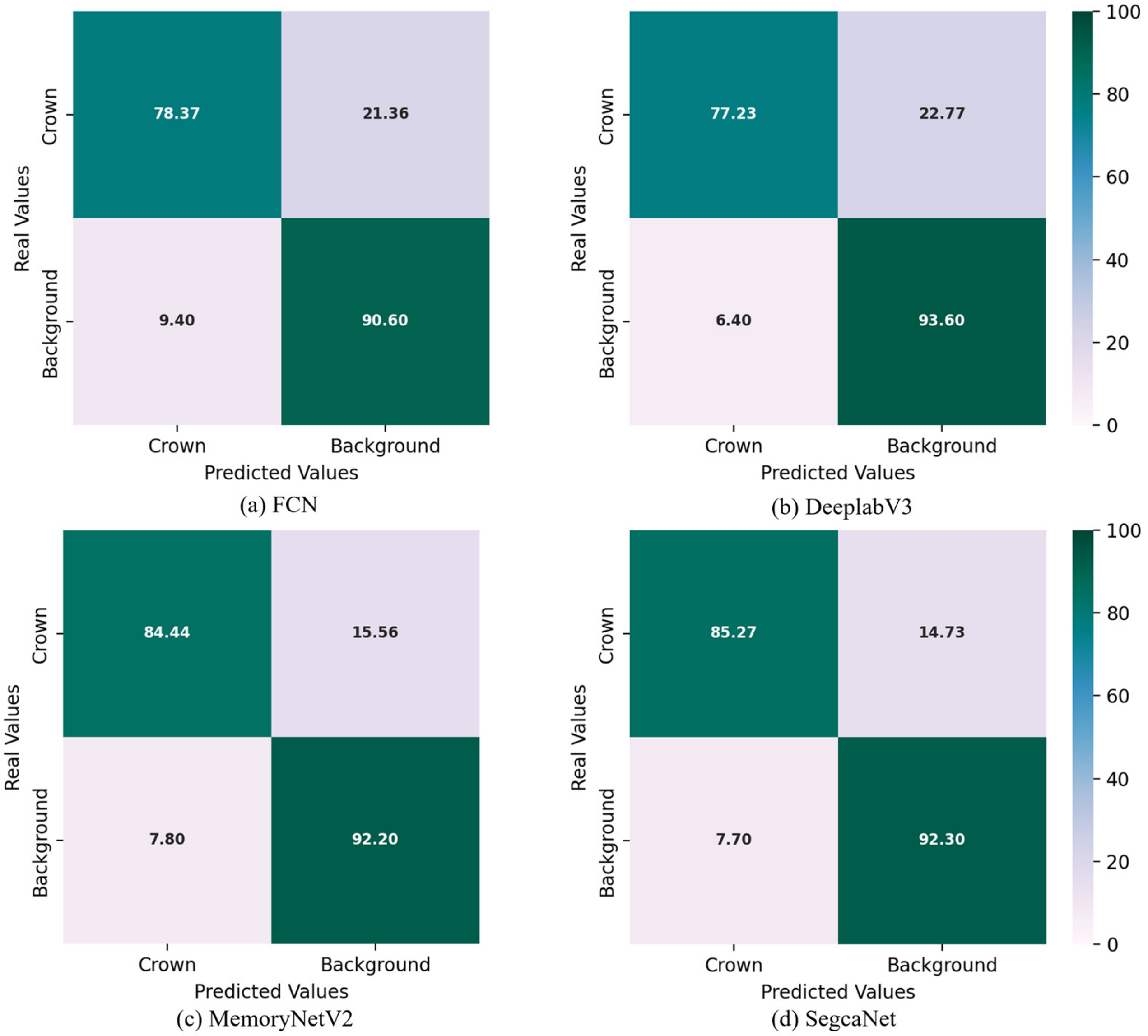
(d) SegcaNet

Figure 9. Change trends of various indicators of the network being evaluated. The horizontal axis is the number of training epochs, and the vertical axis is the metric accuracy.

A confusion matrix is an important tool for evaluating the performance of a classification model. It compares the model's predictions with the true labels and visualizes the results as a matrix.

Figure 10 shows the confusion matrices for all the networks used in this study. These values correspond to those in Table 1.

In the comparative experiment, we applied the network model to the test images to evaluate its performance. The results are shown in Figure 11.



(a) FCN (b) DeeplabV3 (c) MemoryNetV2 (d) SegcaNet

Figure 10. Confusion matrices for all compared networks.

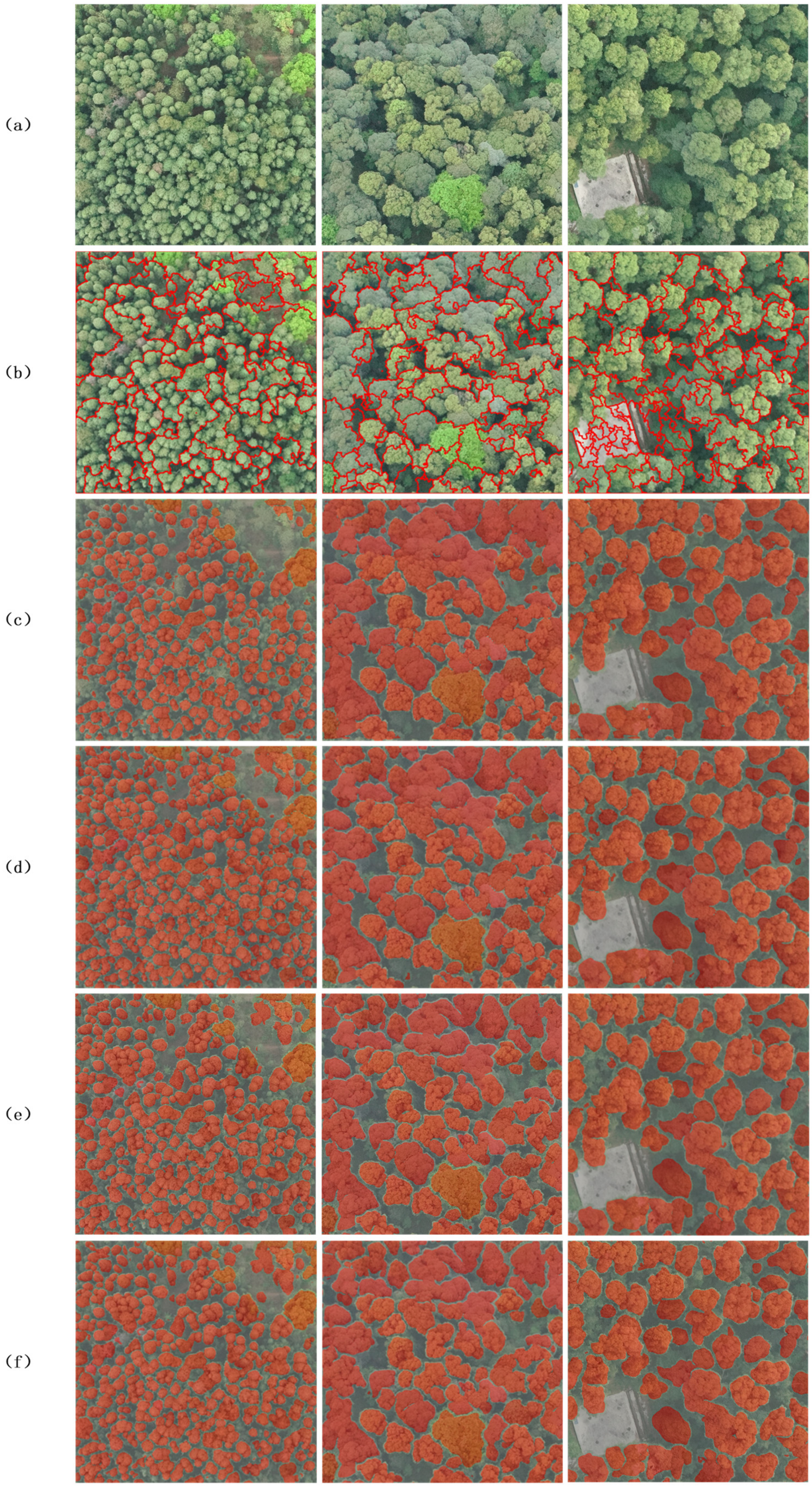


Figure 11. Experimental results for different networks: (**a**) original image; (**b**) watershed algorithm (**c**) FCN network; (**d**) Deeplabv3 network; (**e**) MemoryNetV2 network; (**f**) SegcaNet network.

4.2. Ablation Experiment

In this section, we describe the ablation experiments conducted on MemoryNetV2. We compare the effects of using different modules while keeping all other parameters equal. The results confirm the effectiveness of our proposed improvements.

From the results of the ablation experiments (Table 2), it can be seen that, with the addition of Cloformer alone, the $mIoU$ of the network improves by 1.13%. With the addition of both Cloformer and dilated convolution, the $mIoU$ improves by 1.65%. With the addition of Cloformer, ASPP, and PC-FPN, the $mIoU$ improves by 2.13%.

Table 2. Experimental effects using different modules.

Methods	Re(%)	Pr(%)	$mIoU$(%)	F_1(%)
MemoryNetv2	91.51%	84.44%	80.23%	87.83%
MemoryNetv2 + Cl	90.26%	83.56%	81.36%	86.78%
MemoryNetv2 + Cl + ASPP	90.93%	84.98%	81.88%	87.85%
MemoryNetv2 + Cl + ASPP + PC-FPN	91.68%	85.27%	82.36%	88.35%

Cl, Cloformer; PC-FPN, partially looped feature pyramid; ASPP, dilated convolution.

5. Discussion

In this study, in order to segment individual tree crowns in mixed broadleaf forests, a dataset for mixed broadleaf forests was created. Additionally, a network called SegcaNet was proposed to quickly extract crowns from mixed broadleaf forests.

Although there are many studies on single-tree segmentation, many use traditional algorithms and point cloud data for processing [42–44]. Some researchers have used Mask R-CNN and other neural networks to segment individual tree crowns in high-resolution images [45,46]. These methods have shown promising results in urban environments, but their effectiveness decreases in dense forests [47].

Traditional segmentation algorithms cannot effectively handle broadleaf forest images with complex backgrounds. Supervised learning methods are superior to traditional algorithms, but they are limited by the network's feature extraction capabilities, which can result in discontinuous and incorrect segmentation.

The original MemoryNetV2 faces divergence issues due to dataset variability. Its memory module cannot collect enough semantic information, resulting in poor segmentation performance.

SegcaNet improves the performance by introducing partially looped feature pyramids and Cloformer based on the convolutional attention mechanism. The partially looped feature pyramids improve the precision of segmenting crown edges by using more detailed information for fusion. Cloformer uses depthwise separable convolution to construct a convolutional attention mechanism, thereby focusing more computing resources on more important areas of the image, reducing the amount of calculation and improving the performance of the network. Comparative experiments and ablation studies confirmed the effectiveness of these improvements.

The tree height, diameter at breast height (DBH), and species are essential parameters in ground surveys. With these parameters, it is possible to calculate the biomass, carbon storage, and timber volume, achieving the objective of forest resource assessment. However, plot surveys require the measurement of every tree, which is labor-intensive and resource-consuming. Additionally, due to their location and terrain, some forest areas are inaccessible to ground survey personnel, making it difficult to evaluate these regions accurately.

The crown segmentation method for mixed broadleaf forests proposed in this paper focuses on the inventory of forest resources. The algorithm can accurately identify crowns in mixed broadleaf forest areas. By combining these data with specific parameters obtained using UAVs, it is possible to estimate the crown width of a tree. This method offers valuable support for conducting forest resource surveys.

Although the SegcaNet network proposed in this paper achieved excellent segmentation results, it still has some shortcomings. For example, in mixed broadleaf forests with complex backgrounds and a high canopy density, weeds and shrub areas may still be misidentified as canopy areas. This issue could be mitigated through more high-quality data preprocessing steps. Additionally, apart from the efficient and accurate segmentation of the crown area in mixed broadleaf forests, this paper does not discuss its further applications. In future studies, we will consider how the segmented results can be used more precisely for forest resource inventory and forest management planning.

6. Conclusions

This study proposes a mixed broadleaf forest canopy segmentation algorithm, SegcaNet, and creates a dataset for the segmentation of mixed broadleaf forests. This dataset includes 1200 panoramic images and 18,000 single-tree crown images.

Compared to other networks, SegcaNet has the following advantages: (1) By introducing a partial looped feature pyramid, the algorithm pays more attention to the details of tree crowns, reducing segmentation errors and over-segmentation in mixed broadleaf forests. (2) The convolutional attention mechanism and memory mechanism address the dataset divergence caused by images of different scales, improving the model's performance.

Based on the experimental results, the following conclusions can be drawn: (1) Compared to commonly used supervised learning methods such as FCN, Deeplabv3, and MemoryNetV2, the proposed method achieves the best overall performance. The $mIoU$ is 4.8% higher than FCN, 4.33% higher than Deeplabv3, and 2.13% higher than MemoryNetV2. Additionally, the computational load of the network is reduced to some extent by using depthwise separable convolution in the convolutional attention mechanism. (2) This study also demonstrates the significant application potential of artificial intelligence in forest resource surveys. The proposed algorithm can provide a reference for the application of deep learning in forestry.

Author Contributions: X.T.: Writing—original draft, Data curation, Validation, and Visualization. Z.L., K.X., W.Z., J.L. and X.P.: Writing—review and editing and Formal analysis. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National program on key research during the Fourteenth Five-year Plan Period (2022YFD2200505); the Hunan Provincial Natural Science Foundation, Dongting Lake Forest Structure and Ecosystem Service Scale Correlation and Optimization Model (2022JJ31000).

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments: We are very grateful to all those who assisted in the data collection and experimentation. We also thank the anonymous reviewers for their helpful comments and suggestions.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Kuuluvainen, T.; Angelstam, P.; Frelich, L.; Jõgiste, K.; Koivula, M.; Kubota, Y.; Lafleur, B.; Macdonald, E. Natural Disturbance-Based Forest Management: Moving Beyond Retention and Continuous-Cover Forestry. *Front. For. Glob. Chang.* **2021**, *4*, 629020. [CrossRef]
2. Aggestam, F.; Konczal, A.; Sotirov, M.; Wallin, I.; Paillet, Y.; Spinelli, R.; Lindner, M.; Derks, J.; Hanewinkel, M.; Winkel, G. Can Nature Conservation and Wood Production Be Reconciled in Managed Forests? A Review of Driving Factors for Integrated Forest Management in Europe. *J. Environ. Manag.* **2020**, *268*, 110670. [CrossRef]
3. Liu, Y.; Gong, W.; Hu, X.; Gong, J. Forest Type Identification with Random Forest Using Sentinel-1A, Sentinel-2A, Multi-Temporal Landsat-8 and DEM Data. *Remote Sens.* **2018**, *10*, 946. [CrossRef]
4. Miraki, M.; Sohrabi, H.; Fatehi, P.; Kneubuehler, M. Individual Tree Crown Delineation from High-Resolution UAV Images in Broadleaf Forest. *Ecol. Inform.* **2021**, *61*, 101207. [CrossRef]
5. Santos, A.A.D.; Marcato Junior, J.; Araújo, M.S.; Di Martini, D.R.; Tetila, E.C.; Siqueira, H.L.; Aoki, C.; Eltner, A.; Matsubara, E.T.; Pistori, H.; et al. Assessment of CNN-Based Methods for Individual Tree Detection on Images Captured by RGB Cameras Attached to UAVs. *Sensors* **2019**, *19*, 3595. [CrossRef] [PubMed]
6. Xie, Y.; Wang, Y.; Sun, Z.; Liang, R.; Ding, Z.; Wang, B.; Huang, S.; Sun, Y. Instance Segmentation and Stand-Scale Forest Mapping Based on UAV Images Derived RGB and CHM. *Comput. Electron. Agric.* **2024**, *220*, 108878. [CrossRef]

7. Guimarães, N.; Pádua, L.; Marques, P.; Silva, N.; Peres, E.; Sousa, J.J. Forestry Remote Sensing from Unmanned Aerial Vehicles: A Review Focusing on the Data, Processing and Potentialities. *Remote Sens.* **2020**, *12*, 1046. [CrossRef]
8. Taye, F.A.; Folkersen, M.V.; Fleming, C.M.; Buckwell, A.; Mackey, B.; Diwakar, K.C.; Le, D.; Hasan, S.; Ange, C.S. The Economic Values of Global Forest Ecosystem Services: A Meta-Analysis. *Ecol. Econ.* **2021**, *189*, 107145. [CrossRef]
9. Dubois, H.; Verkasalo, E.; Claessens, H. Potential of Birch (*Betula pendula* Roth and *B. pubescens* Ehrh.) for Forestry and Forest-Based Industry Sector within the Changing Climatic and Socio-Economic Context of Western Europe. *Forests* **2020**, *11*, 336. [CrossRef]
10. Zhen, Z.; Quackenbush, L.; Zhang, L. Trends in Automatic Individual Tree Crown Detection and Delineation—Evolution of LiDAR Data. *Remote Sens.* **2016**, *8*, 333. [CrossRef]
11. Fassnacht, F.E.; Hartig, F.; Latifi, H.; Berger, C.; Hernández, J.; Corvalán, P.; Koch, B. Importance of Sample Size, Data Type and Prediction Method for Remote Sensing-Based Estimations of Aboveground Forest Biomass. *Remote Sens. Environ.* **2014**, *154*, 102–114. [CrossRef]
12. Huang, H.; Li, X.; Chen, C. Individual Tree Crown Detection and Delineation From Very-High-Resolution UAV Images Based on Bias Field and Marker-Controlled Watershed Segmentation Algorithms. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 2253–2262. [CrossRef]
13. Kestur, R.; Angural, A.; Bashir, B.; Omkar, S.N.; Anand, G.; Meenavathi, M.B. Tree Crown Detection, Delineation and Counting in UAV Remote Sensed Images: A Neural Network Based Spectral–Spatial Method. *J. Indian Soc. Remote Sens.* **2018**, *46*, 991–1004. [CrossRef]
14. Wang, L.; Gong, P.; Biging, G.S. Individual Tree-Crown Delineation and Treetop Detection in High-Spatial-Resolution Aerial Imagery. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 351–357. [CrossRef]
15. Jing, L.; Hu, B.; Noland, T.; Li, J. An Individual Tree Crown Delineation Method Based on Multi-Scale Segmentation of Imagery. *ISPRS J. Photogramm. Remote Sens.* **2012**, *70*, 88–98. [CrossRef]
16. Wu, J.; Yang, G.; Yang, H.; Zhu, Y.; Li, Z.; Lei, L.; Zhao, C. Extracting Apple Tree Crown Information from Remote Imagery Using Deep Learning. *Comput. Electron. Agric.* **2020**, *174*, 105504. [CrossRef]
17. Zhang, C.; Zhou, J.; Wang, H.; Tan, T.; Cui, M.; Huang, Z.; Wang, P.; Zhang, L. Multi-Species Individual Tree Segmentation and Identification Based on Improved Mask R-CNN and UAV Imagery in Mixed Forests. *Remote Sens.* **2022**, *14*, 874. [CrossRef]
18. Xue, X.; Luo, Q.; Bu, M.; Li, Z.; Lyu, S.; Song, S. Citrus Tree Canopy Segmentation of Orchard Spraying Robot Based on RGB-D Image and the Improved DeepLabv3+. *Agronomy* **2023**, *13*, 2059. [CrossRef]
19. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv* **2017**, arXiv:1706.05587.
20. Yan, S.; Jing, L.; Wang, H. A New Individual Tree Species Recognition Method Based on a Convolutional Neural Network and High-Spatial Resolution Remote Sensing Imagery. *Remote Sens.* **2021**, *13*, 479. [CrossRef]
21. Weinstein, B.G.; Marconi, S.; Bohlman, S.; Zare, A.; White, E. Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks. *Remote Sens.* **2019**, *11*, 1309. [CrossRef]
22. Braga, J.R.G.; Peripato, V.; Dalagnol, R.; Ferreira, M.P.; Tarabalka, Y.; Aragão, L.E.O.C.; de Campos Velho, H.F.; Shiguemori, E.H.; Wagner, F.H. Tree Crown Delineation Algorithm Based on a Convolutional Neural Network. *Remote Sens.* **2020**, *12*, 1288. [CrossRef]
23. Krůček, M.; Král, K.; Cushman, K.; Missarov, A.; Kellner, J.R. Supervised Segmentation of Ultra-High-Density Drone Lidar for Large-Area Mapping of Individual Trees. *Remote Sens.* **2020**, *12*, 3260. [CrossRef]
24. Sani-Mohammed, A.; Yao, W.; Heurich, M. Instance Segmentation of Standing Dead Trees in Dense Forest from Aerial Imagery Using Deep Learning. *ISPRS Open J. Photogramm. Remote Sens.* **2022**, *6*, 100024. [CrossRef]
25. Yu, B.; Liu, Y.; Zhao, T. Counting of Pine Wood Nematode Disease Trees Based on Threshold Segmentation. *J. Phys. Conf. Ser.* **2021**, *1961*, 012033. [CrossRef]
26. Wang, A.; Zhang, W.; Wei, X. A Review on Weed Detection Using Ground-Based Machine Vision and Image Processing Techniques. *Comput. Electron. Agric.* **2019**, *158*, 226–240. [CrossRef]
27. Lu, Y.; Young, S.; Wang, H.; Wijewardane, N. Robust Plant Segmentation of Color Images Based on Image Contrast Optimization. *Comput. Electron. Agric.* **2022**, *193*, 106711. [CrossRef]
28. Sabzi, S.; Abbaspour-Gilandeh, Y.; García-Mateos, G. A Fast and Accurate Expert System for Weed Identification in Potato Crops Using Metaheuristic Algorithms. *Comput. Ind.* **2018**, *98*, 80–89. [CrossRef]
29. Jothiaruna, N.; Joseph Abraham Sundar, K.; Karthikeyan, B. A Segmentation Method for Disease Spot Images Incorporating Chrominance in Comprehensive Color Feature and Region Growing. *Comput. Electron. Agric.* **2019**, *165*, 104934. [CrossRef]
30. Abdalla, A.; Cen, H.; El-manawy, A.; He, Y. Infield Oilseed Rape Images Segmentation via Improved Unsupervised Learning Models Combined with Supreme Color Features. *Comput. Electron. Agric.* **2019**, *162*, 1057–1068. [CrossRef]
31. Jin, Z.; Yu, D.; Yuan, Z.; Yu, L. MCIBI++: Soft Mining Contextual Information Beyond Image for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 5988–6005. [CrossRef] [PubMed]
32. Jin, Z.; Gong, T.; Yu, D.; Chu, Q.; Wang, J.; Wang, C.; Shao, J. Mining Contextual Information Beyond Image for Semantic Segmentation. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 7211–7221.
33. Fan, Q.; Huang, H.; Guan, J.; He, R. Rethinking Local Perception in Lightweight Vision Transformer. *arXiv* **2023**, arXiv:2303.17803.

34. Yu, F.; Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. *arXiv*, 2016; arXiv:1511.07122.
35. Li, L.; Doroslovacki, M.; Loew, M.H. Approximating the Gradient of Cross-Entropy Loss Function. *IEEE Access* **2020**, *8*, 111626–111635. [CrossRef]
36. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
37. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
38. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
40. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer: Cham, Switzerland, 2015.
41. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. *arXiv* **2021**, arXiv:2103.14030.
42. Yu, J.; Lei, L.; Li, Z. Individual Tree Segmentation Based on Seed Points Detected by an Adaptive Crown Shaped Algorithm Using UAV-LiDAR Data. *Remote Sens.* **2024**, *16*, 825. [CrossRef]
43. Liu, Y.; Chen, D.; Fu, S.; Mathiopoulos, P.T.; Sui, M.; Na, J.; Peethambaran, J. Segmentation of Individual Tree Points by Combining Marker-Controlled Watershed Segmentation and Spectral Clustering Optimization. *Remote Sens.* **2024**, *16*, 610. [CrossRef]
44. Xu, J.; Su, M.; Sun, Y.; Pan, W.; Cui, H.; Jin, S.; Zhang, L.; Wang, P. Tree Crown Segmentation and Diameter at Breast Height Prediction Based on BlendMask in Unmanned Aerial Vehicle Imagery. *Remote Sens.* **2024**, *16*, 368. [CrossRef]
45. Yao, Z.; Chai, G.; Lei, L.; Jia, X.; Zhang, X. Individual Tree Species Identification and Crown Parameters Extraction Based on Mask R-CNN: Assessing the Applicability of Unmanned Aerial Vehicle Optical Images. *Remote Sens.* **2023**, *15*, 5164. [CrossRef]
46. Fu, H.; Zhao, H.; Jiang, J.; Zhang, Y.; Liu, G.; Xiao, W.; Du, S.; Guo, W.; Liu, X. Automatic Detection Tree Crown and Height Using Mask R-CNN Based on Unmanned Aerial Vehicles Images for Biomass Mapping. *For. Ecol. Manag.* **2024**, *555*, 121712. [CrossRef]
47. Lv, L.; Li, X.; Mao, F.; Zhou, L.; Xuan, J.; Zhao, Y.; Yu, J.; Song, M.; Huang, L.; Du, H. A Deep Learning Network for Individual Tree Segmentation in UAV Images with a Coupled CSPNet and Attention Mechanism. *Remote Sens.* **2023**, *15*, 4420. [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.