Real-Time Counting and Height Measurement of Nursery Seedlings Based on Ghostnet–YoloV4 Network and Binocular Vision Technology

Xuguang Yuan 1, Dan Li 2,*, Peng Sun 1, Gen Wang 1 and Yalou Ma 1

1 Forestry Information Engineering Laboratory, Northeast Forestry University, Harbin 150040, China
2 College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China
* Correspondence: ld725725@126.com

Abstract: Traditional nursery seedling detection often uses manual sampling counting and height measurement with rulers. This is not only inefficient and inaccurate, but it requires many human resources for nurseries that need to monitor the growth of saplings, making it difficult to meet the fast and efficient management requirements of modern forestry. To solve this problem, this paper proposes a real-time seedling detection framework based on an improved YoloV4 network and binocular camera, which can provide real-time measurements of the height and number of saplings in a nursery quickly and efficiently. The methodology is as follows: (i) creating a training dataset using a binocular camera field photography and data augmentation; (ii) replacing the backbone network of YoloV4 with Ghostnet and replacing the normal convolutional blocks of PANet in YoloV4 with depth-separable convolutional blocks, which will allow the Ghostnet–YoloV4 improved network to maintain efficient feature extraction while massively reducing the number of operations for real-time counting; (iii) integrating binocular vision technology into neural network detection to perform the real-time height measurement of saplings; and (iv) making corresponding parameter and equipment adjustments based on the specific morphology of the various saplings, and adding comparative experiments to enhance generalisability. The results of the field testing of nursery saplings show that the method is effective in overcoming noise in a large field environment, meeting the load-carrying capacity of embedded mobile devices with low-configuration management systems in real time and achieving over 92% accuracy in both counts and measurements. The results of these studies can provide technical support for the precise cultivation of nursery saplings.

Keywords: deep learning; YoloV4; Ghostnet; binocular vision; sapling detection

1. Introduction

The relatively new worldwide trend of ‘precision forestry’ refers to the use of high-tech sensors and analytical tools to support site-specific forest management for the conservation and use of forest resources [1,2]. According to McKinsey & Company Research, precision forestry plays an important role in nursery and forest management, forestry fees, timber delivery and value chains [3]. The global precision forestry market is projected to be worth USD 6.1 billion by 2024 [4] and has become an important industry in China.

Afforestation and reforestation operations constitute an important part of forest management [5], and the quality of seedlings produced by nurseries is related to the survival rate of planted trees, so it is crucial to advance the level of research to nursery techniques. The number of seedlings per unit area is a common indicator used to characterise seedling production. The rapid and accurate identification of saplings and the detection of the number of saplings per unit area play an important role not only in estimating production, but also in breeding and plant phenotyping. The height of a sapling is not only a reflection of its current growth status and the conditions required for its cultivation, but also determines its production trend and yield size as a cash crop. The traditional methods of
testing, counting and measuring the height of saplings using manual sampling estimates are not only unstable and less timely [6], but they are also labour intensive. When forest nurseries need to produce seedlings in a sustainable manner, they must be produced at minimal cost and with a minimal input of resources [7]. Therefore, it is imperative to speed up the mechanisation and automation of seedling production management to significantly reduce labour time and eliminate heavy and inefficient manual work while maintaining a high accuracy.

With the development of technologies such as deep learning [8,9] and computer vision and the substantial increase in computers’ computing power in recent years, the real-time extraction of the number of saplings in a nursery can be accomplished using the fast detection characteristics of neural networks; real-time extraction of the height of saplings in a nursery can be accomplished using binocular cameras to build depth images. Many studies have applied deep learning techniques to forestry timber species recognition, e.g., Jozef Martinka [10] used Matlab to build a deep neural network for detecting timber species, and identified colour temperature pictures of light with 97.9% accuracy. Shustrov [11] used four neural network structures (AlexNet, VGG-16, GoogLeNet and ResNet-50) for fir, pine and spruce wood, respectively, with accuracies of over 90%, 90%, 80% and 70% for their four networks, respectively. Deep learning has also frequently been used in wood knots, surface defects in wood veneers and predicting wood properties. For example, Wei et al. [12], Mohan et al. [13] and Urbonas et al. (2019) [14] designed neural networks for identifying timber knots and timber veneer surface defects, and they achieved detection accuracies of 70%–95%. The detection of tree vegetation in cities using techniques such as UAVs and deep learning has also been studied. For example, Xi et al. [15] used two instance segmentation networks (BlendMask and Mask R-CNN) to segment ginkgo tree canopies in urban environments after a dimensionality reduction in the UAV multispectral images, while Zheng et al. [16] used YoloV4-Lite to segment high-resolution remote sensing images of woods on campus, which were detected and localised. In a tree height measurement study, Prada, E [17] measured the height of a single tree by means of a UAV-based LIDAR scanning sensor. In the detection of forests or native woods, Castilla G et al. [18] used a point cloud of images from a UAV to measure the height of individual conifer seedlings, noting in the paper that the accuracy was high for seedlings above 30 cm, but not applicable to height measurements of seedlings below 30 cm; in 2019, Puliti et al. [19] used a random forest model to estimate the height of 580 circular Norwegian plots, and Imamgholiloo et al. [20] used 2.5 cm GSD DIPCs (defoliated and defoliation) to estimate the average height of small trees within 15 plots in a conifer-dominated regeneration stand in Finland, but both studies involved trees above 1 metre in height.

After comparing 40 studies on deep learning in agroforestry applications, Kamilaris et al. [21] found that deep learning has higher accuracy in image recognition and is better than commonly used image processing techniques. Using deep learning for detection can help to obtain deeper features and produce more accurate classification results [22–25]. It is divided into three broad categories of classical target detection algorithms: (1) the region-based convolutional neural network (R-CNN) family based on the region [26], which has the highest accuracy but whose algorithm is complex and time consuming; (2) the regression-based YOLO (you only look once) series and SSD (single shot multibox detector), which are fast, small and efficient [27]; (3) density estimation-based methods, i.e., estimating the number of targets through learning target features, and combining the corresponding linear mapping and spatial features to construct a density map [28]. Compared with other neural network detection models, the YOLO model significantly improves detection speed while ensuring detection accuracy [29,30]. This, coupled with the advantage of the smaller overall size of the YOLO model, makes it ideal for mobile embedded device applications, e.g., the improved Fast YOLO model by Redmon [31] was able to process an impressive 155 frames per second. J Wang et al. [32] used YOLOV4-Tiny combined with a ZED 2 stereo camera for 3D reconstruction to obtain 3D coordinates of pixels in the current scene, and calculated the distance between the centre of the potted plant and the optical centre of the binocular
left camera while completing the identification of the flower species. They completed real-time positioning and ranging of flowers with a real-time detection frame rate of 16 FPS and an average absolute error of 18.1 mm for flower centre positioning, with a maximum positioning error of 25.8 mm for flower centres under different light radiation conditions.

The above studies provide an insight into the detection of forest trees. However, the application of these techniques to nurseries with complex backgrounds and smaller individual saplings for target counting and height measurement still suffers from inaccuracies and is time consuming and labour intensive [33]. This study proposes the use of the Ghostnet–YoloV4 network and binocular vision technology to solve this type of problem in order to obtain the number and height of target saplings in real time, as well as to investigate whether the improved network can have a better detection capability to satisfy the rudimentary management equipment of most small nurseries with a lower computing power and the use of inexpensive binocular cameras. Field tests were carried out in nurseries to check that the method achieves the practical requirements of sapling counting and height measurement and enables intelligent mechanical operation at minimal cost. The short-term aim is to use the research results to help nursery staff reduce the burden of manually counting saplings and measuring height, while the long-term goal is to improve the effectiveness of forestry machinery automation and lay the foundation for intelligent forestry management.

2. Materials and Methods

2.1. Process of Nursery Sapling Detection Based on Ghostnet–YoloV4 Network and Binocular Cameras

As shown in Figure 1, the original left images of various saplings were first obtained using binocular cameras, labelled and then data augmented to produce a dataset, which was then fed into an improved neural network for training to obtain detection weights. At the same time, the binocular camera was stereo-calibrated and calibrated to obtain its internal parameters; when the neural network can detect the saplings in the still image and the BM algorithm integrated into the network can acquire the depth image, the nursery can be accessed for the real-time counting and height measurement of the saplings. The number and height of saplings acquired in real time are displayed directly in the image, and more detailed information on the location and height of the saplings can be seen in the PyCharm output window.

Figure 1. Detection process of tree saplings collected in the field environment.
2.2. Datasets

2.2.1. Data Collection

The original experimental data were images of nursery saplings taken using the left binocular camera, each from a different angle. Considering the effects of different complex backgrounds and camera parameters in the natural environment, some lower-quality images were excluded from the data pre-processing in this study. The original images of the saplings are shown in Figure 2. Images of (a) large, (b) medium and (c) small spruce (the Latin name is *Picea asperata* Mast; large, medium and small indicate its body size at different growth periods), (d) Mongolian scotch pine (the Latin name is *Pinus sylvestris* var. mongholica Litv) and (e) Manchurian ash (the Latin name is *Fraxinus mandshurica* Rupr) saplings were collected from Xiaoling Town Forestry, Acheng District, Harbin, Heilongjiang Province, China, on 6 May 2022 from 13:30 to 16:30. Using a binocular camera, 500 images (640 × 480 pixels, 100 KB-130 KB in size) were taken of each sample, totalling 2500 original images. There were 1500 images left after screening. The dataset format of PASCAL VOC was used for this study, and the label file (xml format) was created manually using ‘Labellmg’ as the labelling tool.

![Figure 2. Images of five saplings: (a) large, (b) medium and (c) small spruce; (d) Manchurian ash; (e) Mongolian scotch pine.](image)

2.2.2. Data Augmentation

YOLOV4 [34] comes with data augmentation functions, such as Mosaic, which can enrich various objects and backgrounds, greatly improving YOLOV4 performance and effectively solving the problem of the poor detection of smaller volume objects in model training. However, for the YOLO series, obtaining a larger dataset for the training and validation of the network model results in training weights that better fit this type of data and the better detection of the network model [35,36]. To prevent overfitting, this study used 15 methods to expand the original data, including flipping, adding noise, cutting, rotating, stretching and adjusting brightness, as shown in Table 1. The images were expanded simultaneously with the label files, and the expanded images and label files were divided into a training set and a validation set in the ratio of 9:1. This resulted in a training set of 22,500 images and a validation set of 2500 images. The testing of the training results was carried out directly at the nursery for the real-time detection of saplings, and the correctness of the method was evaluated by comparing the manual counting results with the system counting results to verify the timeliness and robustness of the method.

2.3. Experimental Architecture

The experiments were conducted on Windows 10 using the pytorch 1.7.1 deep learning framework. The experimental platform was PyCharm 2020.2.3 with built-in python version 3.6.5. The binocular camera calibration software was Matlab 2021a. The hardware equipment configuration was as follows: binocular camera model: WN-L2110.K350L; triangular mount and checkerboard calibration board; Intel Core i7-10710U@1.1 GHz hexa-core computer processor; 16 GB memory; NVIDIA GeForce MX350 graphics card (2 GB video memory and 8 GB virtual memory); 512 GB hard drive. As this study required a close fit of the detection frame to the sapling, the upper and lower edges of the labelled rectangular frame were overlapped with the vertices and bottom points of the sapling during labelling;
the confidence threshold for network training and detection was set to high to improve the fit of the detection frame to the sapling during detection.

Table 1. Data augmentation dataset examples.

| X-Axis Flip | Y-Axis Flip | XY-Axis Flip | Salt Noise | Gaussian Noise |
|-------------|-------------|--------------|------------|----------------|
| ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) | ![Image](image4.png) | ![Image](image5.png) |
| ![Image](image6.png) | ![Image](image7.png) | ![Image](image8.png) | ![Image](image9.png) | ![Image](image10.png) |

2.4. Ghostnet–YoloV4 Network Architecture

YoloV4 is essentially a large CNN network that is more speed conscious, converting the detection problem into a regression problem, and various optimized versions are under development [34]. The main function of the CSPDarknet53 of the YoloV4 network is to perform the initial feature extraction of the input image; the SPP is used to build a bottom-up feature pyramid to improve the perceptual field; the enhanced feature extraction network PANet is then used to perform more expressive feature fusion; and finally, YoloHead uses the extracted features for detection. Although real-time, high-quality and convincing target detection results can be applied on a single GPU, a network structure is more complex, and the number of parameters and operations is high, which presents a considerable burden for embedded devices [16]. This study, therefore, introduced Ghostnet into the YOLOV4 network to enhance feature extraction while reducing the network’s transport burden.

2.4.1. Ghostnet

Ghostnet is a new lightweight neural network architecture proposed by a Huawei researcher in a paper published in June 2020 [37]. The abstract of the paper states that the network is designed to help solve the problem of the highly difficult deployment of convolutional neural networks in embedded devices; Ghostnet networks outperform Google MobileNetV3 and Facebook’s FBNet across the board. Instead of normal convolution, the authors propose a novel Ghost module, as shown in Figure 3, that uses fewer parameters to generate more feature maps. The Ghost module applies a series of linear transformations to generate these redundant feature maps, reducing the amount of computation caused by some convolution operations. This operation generates several Ghost feature maps that extract the required information from the original features at a small cost.
The Ghost module is plug-and-play and can be stacked to produce the Ghost bottleneck, which is a lightweight neural network known as Ghostnet. The authors conducted comparative experiments on the ImageNet classification dataset and showed that the network can perform fast inferences on mobile devices. This study, therefore, introduced Ghostnet to YoloV4 for deployment on laptops for real-time detection.

2.4.2. Ghostnet–YoloV4 Improvement Method

The main function of YoloV4’s CSPDaeknet53 is to perform a series of convolution operations on the input image to complete the initial feature extraction and obtain a feature map. The main function of Ghostnet is to bypass some of the convolution operations and use cheaper linear operations to obtain feature maps. Therefore, our improved approach is to replace the CSPDaeknet53 with Ghostnet for the initial feature extraction of the input image, which allows us to maintain similar recognition performance while reducing the computational cost of the generic convolutional layer.

CSPDaeknet53 will eventually output three effective feature layers. Their feature map height, width and number of channels are: $52 \times 52 \times 256$, $26 \times 26 \times 512$ and $13 \times 13 \times 1024$, respectively, a subsequent enhancement of the feature extraction network construction. If we directly replace Ghostnet with CSPDaeknet53 to process the input image and output the feature map, it will result in a mismatch between the input feature map height and width and the number of channels of SPP and PANet. Therefore, we segmented the Ghostnet sequence model constructed using the Ghost bottleneck by deriving a list of Ghostnet’s cfgs parameters, then derived the positions of the height and width of the feature maps satisfying SPP and PANet, and finally, removed them as the input feature maps for SPP and PANet.

The cfgs parameter table is shown in Table 2, where $k$ represents the convolutional kernel size, indicating the feature extraction capability across feature points; $t$ represents the channel count size of the first Ghost module; $c$ represents the final output channel count of the bottleneck structure; $SE$ indicates whether the attention mechanism is used, when $SE > 0$ indicates use; $s$ represents the step size, if 2 it will compress the height and width of the incoming feature layer. The output is the derived height, width and number of channels of the output feature map after each stage. The image will then change from $416 \times 416 \times 3$ to $208 \times 208 \times 16$, and then enter the feature extraction into the table.

As we can see from the table, the height and width of the output feature maps for stage 3, stage 4 and stage 5 matched the height and width of the final output feature map of CSPDaeknet53. After Ghostnet was used to extract these three feature layers and passed into YoloV4 as the initial feature extraction output, our study also needed to modify the number of input channels for the convolution operation used in the SPP and PANet. This was carried out by making the input channel equal to the output channel, so that the number of output channels for Ghostnet as the backbone feature extraction network was the same as the number of input channels of the subsequent enhanced feature extraction network. At this point, the modification that introduced Ghostnet into YoloV4 was completed.
Table 2. Cfgs parameters.

| stage | K | t | c | SE | s | Output |
|-------|---|---|---|----|---|--------|
| 1     | 3 | 16| 16| 0  | 1 | 208 × 208 × 16 |
| 2     | 3 | 48| 24| 0  | 2 | 104 × 104 × 24 |
|       | 3 | 72| 24| 0  | 1 | 52 × 52 × 40  |
| 3     | 5 | 72| 40| 0.25 | 1 | 26 × 26 × 112 |
|       | 5 | 120| 40| 0.25 | 1 | 13 × 13 × 160 |

2.4.3. PANet Improvements

The main role of YoloV4’s enhanced feature extraction network PANet is to perform feature fusion on the three initial effective feature layers. In this way, better features are extracted, and three more effective feature layers are obtained, resulting in a higher detection accuracy of the detection network YoloHead. The number of parameters in PANet is mostly concentrated on the ordinary convolution of 3 × 3. Therefore, to further reduce the number of parameters, we used the depth-separable convolution \[38,39\] to replace all the 3 × 3 ordinary convolutions used in PANet. We used the summary function to traverse the entire network structure and used its output network structure as a basis to derive the Ghostnet–YoloV4 network structure diagram with reference to the YoloV4 network, as shown in Figure 4.

![Figure 4. Ghostnet–YoloV4 network.](image-url)
The same approach as above was used to introduce the Mobilenetv3 modification to YoloV4 to compare the number of parameters. The summary function was used to calculate the total number of network parameters and to obtain the total number of parameters for the two different networks before and after the PANet modification. The total number of parameters is shown in Table 3, where (1) represents the original YoloV4 network; (2) represents Mobilenetv3–YoloV4, where Mobilenetv3 is introduced to replace the YoloV4 backbone; (3) represents Ghostnet–YoloV4, where Ghostnet is introduced to replace the YoloV4 backbone; (4) represents the introduction of Mobilenetv3 to replace the YoloV4 backbone and the modification of PANet for Mobilenetv3–YoloV4; and (5) represents the introduction of Ghostnet to replace the YoloV4 backbone and the modification of PANet for Ghostnet–YoloV4. The number of network parameters was significantly reduced. Compared with the Mobilenetv3–YoloV4 network, the Ghostnet–YoloV4 network has the smallest number of parameters, thus reducing many unnecessary calculations in the repetitive training and detection process.

Table 3. Total net parameter.

| Method         | (1)     | (2)     | (3)     | (4)     | (5)     |
|----------------|---------|---------|---------|---------|---------|
| Total          | 64,363,101 | 39,989,933 | 39,062,013 | 11,729,069 | 11,428,545 |

2.5. Integration of Binocular Vision Technology with YoloHead

2.5.1. Principle of Binocular Stereo Vision for Height Measurement

The height of a sapling is the length from the root neck of the sapling to the slight top of the main trunk, and its length at production is defined as the height of the sapling. As shown in Figure 5, point P is the apex of the sapling, and point Q is the point of contact between the rootstock and the ground, so the length L of PQ is the height of the sapling.

![Figure 5. Schematic of seedling height.](image)

The 3D coordinates of objects in the actual scene can be determined using binocular stereo vision techniques [32,40]. For 3D reconstruction using binocular cameras, we first obtained the depth information of the current binocular view and then performed a conversion from the camera coordinate system to the world coordinate system to obtain the 3D coordinates of the object.

As shown in Figure 6, the left and right eye cameras were fixed at points B and E; point P is the actual point to be measured; C and C’ are the imaging points of point P in the binocular image, and the distance between the two points is d; the distance BE between the left and right cameras is the baseline distance m; GE is the focal length f of the binocular camera; and the depth of point P is D. Then, the depth information D can be determined according to the triangular relationship

\[ D = \frac{f \times m}{d} \]  

(1)
working steps of the BM algorithm are described below.

(2) Pre-filtering: the window of the image filter is shifted in parallel on the acquired images. Figure 6. Parallax distance schematic.

The distance d is the parallax of the binocular view and was calculated using camera calibration parameters and a stereo matching algorithm. In this paper, the classical BM stereo matching algorithm was used to obtain the parallax depth map of the binocular image, and the left camera optical centre was used as the origin for the 3D reconstruction, which in turn extracted the 3D coordinates of the vertex P (X₁, Y₁, and Z₁) and the base point Q (X₂, Y₂, and Z₂) of the sapling, so that the tree height L was

\[ L = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2} \]  (2)

2.5.2. Binocular Camera Calibration

In this study, the Camera Calibrator toolbox of Matlab was used for calibration. More than 40 images of the checkerboard grid were taken with the binocular camera and transferred to Matlab to obtain the calibration parameters for the binocular camera and complete the PyCharm parameter file. The calibration can be carried out by clicking on the histogram to remove the poorly angled checkerboard images and reduce the calibration error. The parameters to be filled in were the left and right camera internal parameters, the radial and tangential aberrations of the left and right cameras, the rotation matrix of camera 2 with respect to camera 1, the translation matrix of camera 2 with respect to camera 1 and the image size. The internal reference was used to correct for aberrations in the captured binocular images, to obtain an isochronous map (where the corresponding pixels of the binocular images are on a straight line) and to apply it to a depth map in stereo matching.

2.5.3. BM Stereo Matching Algorithm

BM stands for bidirectional matching. It works by dividing the frames of the two cameras into a number of small squares for matching: moving the small squares to match the small squares in the other image, finding the pixel positions of the different squares in the other image, and then combining the relationship data of the two cameras (translation and rotation matrices in the calibrated parameters) to calculate the actual depth of the object to generate the corresponding depth map. The BM algorithm is extremely fast and has a great speed advantage in stereo matching work for binoculars in real time. The main working steps of the BM algorithm are described below.

(1) Image acquisition: the images taken by the left and right eye cameras of the binocular camera at the same time, and the acquired images are orthorectified according to the camera parameters and distortion coefficients, and then enter the pre-filtering process.

(2) Pre-filtering: the window of the image filter is shifted in parallel on the acquired images to normalize the brightness of each part of the image while highlighting the relevant details in the image and enhancing the texture in the image, so that the features in the image can be extracted more easily. (3) This step is the most critical step in the BM algorithm, which is to match the highest matching pair of points in the left and right images, i.e., for each pixel point in the left (right) image, find the best matching pixel point in the right (left) image, and only find the best matching pixel point in the right (left) image with the maximum
correctness. (4) Post-filtering: the purpose of post-filtering is to filter the high quality matching points and to remove the incorrect matching points to the maximum extent.

2.5.4. Introduction of the YoloHead Method for Binocular Vision Technology

In order to achieve both the counting and height measurement of saplings, this study integrated the binocular vision code into YoloHead, the detection part of YoloV4. The basic principle is that the left view of each frame of a sapling was used as the image to be detected and fed into the Ghostnet–YoloV4 network for detection, where the current sapling category and number were quickly obtained and displayed from the top left corner of the output window; at the same time of detection, each binocular image of a sapling was stereo matched using the BM algorithm to obtain a depth map to complete the 3D reconstruction, thus obtaining the 3D position information of the current scene pixels and storing it in the points_3d function; then, the 2D coordinates of the detection frame of each sapling obtained by detecting the left image of the sapling were taken out, the vertices and base points of the detection frame of the height-fitted sapling were replaced by the vertices and base points of the sapling and converted to the corresponding 3D coordinates using the points_3d function, and finally the sapling height was obtained. In order to solve the shortcomings of a small number of sapling key points with a poor fit to the detection frame and saplings falling over or sloping, the solution of the manual selection of sapling key points was introduced: a mouse click event was added to the depth map of saplings, and the height of a small number of saplings was found in real time by clicking on the top and bottom points of the depth image of the saplings that fit poorly.

2.6. Output Window Design

For the real-time inspection, three windows were designed to display the results. The depth window shows the 3D depth image of the current scene; the left window shows the centre of the sapling in the current scene and the corresponding height; and the video window shows the current detection frame and number of saplings. The two parameters were set on the depth window adjustment bur: ‘num’ is the difference between the maximum and minimum parallax value, while ‘blocksize’ is the matching block size. Both parameters have a large impact on the depth map and adjusting them can significantly improve the display and accuracy of the depth map when the distance to the target varies. For real-time detection, adjust the height of the tripod and the angle of the binocular camera to optimally fit the sapling in the field of view, and adjust the parameters above the depth image of the binocular camera to optimally fit the sapling in the current scene when the block outline is complete and resembles a sapling.

2.7. Results Statistics Method

Manual testing was carried out with a research team of five people, and the average results of the five people were compared with the computerized testing results to verify the accuracy of the method. As the effective depth of the binocular camera for this experiment was 3–10 m, and the nursery field was large, we could not extract the number and height of saplings from the whole field at once. Therefore, for each type of spruce sapling, the experiment was set up to detect and count at three different points in the field.

3. Results

3.1. Training Parameters and Results

The Ghostnet–YoloV4 network parameters and training results are shown in Table 4. The original detection image size was 640 × 480, but the network was automatically scaled to 416 × 416 before being fed into the network for training and detection. mAP values and recall rates were not particularly high due to the use of data augmentation. We spent a total of nearly 120 h training 25,000 images, resulting in a training loss of 0.35. We achieved a frame rate of 15 FPS for real-time detection, which meets the requirements for real-time detection.
Table 4. Training parameters and results.

| Parameter | Epoch | Confidence | Input Shape | Batch |
|-----------|-------|------------|-------------|-------|
|           | 400   | 0.9        | 416 x 416   | 2     |

| Result   | MAP   | Recall    | Frame Rate | Loss  |
|----------|-------|-----------|------------|-------|
|          | 89.8% | 87.03%    | 15         | 0.35  |

3.2. Presentation of Sapling Detection Results

As shown in Figure 7, the spruce saplings varied in colour, size, texture and planting density during the three different growth periods. The colour and texture features helped the network to distinguish saplings from their surroundings, which directly affected the accuracy of the counts. The different sizes and planting densities of the saplings made the depth images of each sapling different. The larger the sapling, the more accurate the height measurement; the sparser the planting, the lower the level of occlusion, and the more accurate the count. From the diagram, we can quickly determine the number and height of saplings of this type.

![Spruce saplings detection images](image)

**Figure 7.** Spruce saplings’ detection images: (a) large, (b) medium and (c) small spruce.

The results of the real-time inspection of Mongolian scotch pine are displayed in Figure 8, which also shows the complete design of the output window during detection. For sparse Mongolian scotch pine, no adjustment of the camera shooting angle was required; counting within the effective depth of the binocular camera was virtually error-free and the height measured by the system was very close to the actual height.
As shown in Figure 9, the slender trunk of the sapling was relatively large compared to the crown of the Manchurian ash sapling; the extended crown of the sapling obscured the Manchurian ash saplings from one another, which resulted in the poor detection of the Manchurian ash saplings. Additionally, the roots of the Manchurian ash saplings in the back row of the camera were easily blocked by the saplings near the binoculars, making it difficult to detect the roots of some of the saplings. Moreover, the cadres of the saplings were relatively small, so the smaller trunks of the saplings were not visible in the depth image, making it difficult to match the bottom of the frame with the bottom of the saplings. In this study, upon adjusting the position of the binocular camera downwards, the roots of the willow were exposed as much as possible, while the manual intervention and manual selection of the top and bottom points of the sapling for the severely obscured willow could improve the detection accuracy of the Manchurian ash sapling.

Figure 9. Manchurian ash saplings’ detection images.

3.3. Analysis of Test Results

Table 5 demonstrates the accuracy of detection for three different forms of spruce saplings. The table shows the number of spruce saplings and the average height of the saplings for the three forms; for the 3D coordinates of the centre point obtained simultaneously, they could be used to locate the saplings for future operations, such as precise automatic watering and the application of pesticides. For each point, the following measures were calculated: TP indicates the number of true saplings correctly detected as saplings; FP indicates the number of false saplings incorrectly detected as saplings; FN indicates the number of true saplings incorrectly detected or missed; count indicates the average number of saplings counted manually; H indicates the average height of saplings measured by the system in cm; and TH indicates the average height of trees measured manually in centimetres.
### Table 5. Test results for three forms of spruce saplings.

| Category           | Point | TP | FP | FN | Count | H    | TH       |
|--------------------|-------|----|----|----|-------|------|----------|
| Large spruce       | 1     | 6  | 0  | 0  | 6     | 130.4 cm | 127.3 cm |
|                    | 2     | 11 | 0  | 0  | 11    | 126.8 cm | 130.1 cm |
|                    | 3     | 7  | 0  | 0  | 7     | 137.8 cm | 133.9 cm |
| Medium spruce      | 1     | 20 | 0  | 2  | 22    | 26.3 cm  | 28.5 cm  |
|                    | 2     | 33 | 1  | 3  | 36    | 25.8 cm  | 28.2 cm  |
|                    | 3     | 14 | 0  | 1  | 15    | 27.2 cm  | 30.1 cm  |
| Small spruce       | 1     | 24 | 1  | 1  | 25    | 17.4 cm  | 20.2 cm  |
|                    | 2     | 33 | 2  | 1  | 34    | 14.8 cm  | 16.6 cm  |
|                    | 3     | 19 | 0  | 0  | 19    | 18.3 cm  | 17.1 cm  |
| Manchurian ash     | 1     | 12 | 0  | 1  | 13    | 51.5 cm  | 53.2 cm  |
|                    | 2     | 17 | 0  | 2  | 19    | 54.6 cm  | 57.4 cm  |
|                    | 3     | 19 | 0  | 1  | 20    | 49.8 cm  | 52.3 cm  |
| Mongolian scotch pine | 1    | 6  | 0  | 0  | 6     | 66.8 cm  | 68.7 cm  |
|                    | 2     | 17 | 1  | 0  | 17    | 72.1 cm  | 74.5 cm  |
|                    | 3     | 9  | 0  | 0  | 9     | 70.2 cm  | 73.4 cm  |

For the large spruce saplings, the nurserymen chose to plant them at a higher density in order to ensure their growth rate, so that they were counted with 100% accuracy. As the crowns of the large spruce saplings were farther away from the roots on the ground, it was easy to distinguish between them, and because these saplings were taller, the binocular camera took photographs from the side, so that the top and bottom points were selected more accurately, so their counting accuracy was also higher. The medium and small spruces were photographed diagonally downwards. The medium spruce was denser, and the shading between saplings had a greater impact on detection, making it easier for two or even more adjacent saplings to be mistakenly detected as one. The small spruce tilted and fell easily and the plants were shorter, making it easier to find the wrong top and bottom points of the saplings when taking height measurements, and resulting in a slightly lower accuracy than the other two spruce forms. However, the small spruce had a lower spacing, so the number of missed detections was lower, but there were slightly more false detections due to the similarity of its form to the surrounding weeds.

The three saplings showed the best detection results in terms of counting results for the Mongolian scotch pine. This is because camphor pine was more sparsely planted, while the three spruce and Manchurian ash sapling species were more densely planted, so the number of errors and omissions at each detection point for camphor pine was relatively low in comparison. In terms of height measurement, the saplings of Mongolian scotch pine were the furthest apart from one another, so the root and crown features were the most pronounced and the height measurements were the most accurate for Mongolian scotch pine saplings. In contrast, although manual intervention improved the detection accuracy of Manchurian ash, there were still a small number of missed detections, especially in the case of the smaller saplings that were relatively close to one another and could easily be detected as one sapling because they were too close to a slightly larger sapling.

The number of saplings was minimal when comparing the system detection with the manual detection, which shows that the data source was a good fit for the network detection function and, therefore, worked well. The slightly larger difference in the sapling height measurement is due to the errors inherent in the binocular camera and the fact that the roots of some saplings were not detected, which made the difference between the bottom point of the rectangular frame and the bottom point of the sapling too large and ultimately pulled down the average height. During the inspection, the binocular camera was very sensitive to changes in light, which was highlighted by the depth map display of Manchurian ash and spruce. When comparing tests of medium-sized spruce in shade and direct sunlight, and comparing tests under cloudy Manchurian ash and sunny Mongolian scotch pine, we found that under good lighting conditions, the grey contours of the saplings
on the depth image were very close together, which also made the height measurements of spruce and Mongolian scotch pine saplings under sunlight more accurate. The reflection of sunlight brought out the colour and texture characteristics of the saplings and allowed more accurate results for spruce and Mongolian scotch pine saplings. In the shade or on cloudy days, the grey-scale contours of the Manchurian ash and medium-sized spruce on the depth images differed less from the background and neighbouring saplings, and the colour and texture characteristics were somewhat reduced, which caused a reduction in the accuracy of both the height measurements and count results. Adequate light conditions made the features, such as texture and colour, of the saplings more visible, facilitating the detection of the network. It is worth noting that when there was sufficient light, the rate of missed detection of multiple saplings into one was significantly reduced.

Table 6 shows the overall detection accuracy of the saplings, from which the Mongolian scotch pine benefitted from its larger spacing, with the highest count accuracy of 96.97% and a high measurement accuracy of 96.55%. Although the front and rear of the Manchurian ash were obscured, it could still count and measure with a high accuracy of over 92%, which could be further improved by combining it with human intervention.

Table 6. Total counting and measuring accuracy.

| Spruce | Manchurian Ash | Mongolian Scotch Pine | Total |
|--------|----------------|-----------------------|-------|
| TP     | 167            | 48                    | 32    | 247   |
| TP + FP + FN | 179          | 52                    | 33    | 264   |
| Count accuracy | 93.30%    | 92.31%                | 96.97%| 93.56%|
| Height accuracy  | 92.93%    | 95.70%                | 96.55%| 95.06%|

3.4. Network Performance Analysis

In order to verify the performance of the improved network, the following four networks were trained separately and tested for comparison to complete the ablation experiment, and the results are shown in Table 7, where (1) represents the original YoloV4 network; (2) represents Ghostnet–YoloV4, where Ghostnet is introduced to replace the YoloV4 backbone; (3) represents YoloV4 with PANet modification only; and (4) represents the introduction of Ghostnet to replace the YoloV4 backbone and the modification of PANet for Ghostnet–YoloV4. This dataset used the original dataset of saplings, a total of 1500 images, divided into a training and validation set in a ratio of 8:2, and training was carried out using four neural networks of 400 epochs. Of these, (1) had the longest training time and (4) had the shortest training time, and it had the best training results with a MAP value of 92.93%. Additionally, for all four networks, the real-time frame rate reached the maximum value of 15 FPS for this binocular camera. As the amount of training and detection data for the four networks was not very large, it was possible to have better performance but poorer detection results. In order to control for possible uncertainties of this type, we kept the influence of external factors as low as possible: all variables were the same, the detection locations were identical and the four neural networks were run separately for counting and altimetry. Accuracy calculations were still carried out using computer testing compared to manual testing. As can be seen in Table 7, Ghostnet–YoloV4, which introduces Ghostnet to replace the YoloV4 backbone and modifies the PANet, demonstrated the highest accuracy in counting and height measurement.

Table 7. Ablation experiment.

|          | (1)      | (2)      | (3)      | (4)      |
|----------|----------|----------|----------|----------|
| MAP      | 83.9%    | 85.0%    | 85.3%    | 87.7%    |
| LOSS     | 1.75     | 0.93     | 1.68     | 0.55     |
| Count accuracy | 91.27%   | 92.91%   | 90.07%   | 93.82%   |
| Height accuracy  | 93.26%   | 93.81%   | 92.59%   | 94.19%   |
4. Discussion

4.1. Reliability of the Ghostnet–YoloV4 Network

The experimental results show that the Ghostnet–YoloV4 network achieves good accuracy in the real-time counting of all three saplings. This result validates the prediction that the use of the Ghostnet network and deep separable convolution to improve YoloV4 not only reduces the network load massively but also has better detection results. The detection speed of the Ghostnet–YoloV4 network is very high, judging from the real-time frame rate of 15 FPS achieved. From the above, it is clear that there is no obstacle to deploying the Ghostnet–YoloV4 network on personal computers. It is also possible to apply the neural network to other mobile devices, such as mobile phones and tablets, in the future, which will greatly enhance the practical and generalisation capabilities of the network and can be applied to more fields for detection.

It is worth noting that Ghostnet–YoloV4 has a much lower number of parameters compared to YoloV4 when training the network, so it is much faster, thereby saving computer training time. This makes sense for practical applications, as for each different tree species, we need to carry out data collection, labelling and network training, and with a very large variety of saplings in the nursery, the training time is particularly important when conducting large-scale tree counts. If the training time is too long, it will cause a reduction in detection capability as the saplings grow and, more importantly, will delay the production process.

4.2. Binocular Camera 3D Reconstruction Capability

This experiment used images of three tree saplings taken using a binocular camera as the main study dataset, and the binocular view allowed the reconstruction of spatial location information. We chose a low-resolution binocular camera lens of only 640 × 480 in consideration of two factors: (1) Since the training data had 25,000 images, if the resolution of each image was increased so that the training set size was greater, it would massively increase the training time, making it unfavourable for both experimental research and field applications. (2) The network real-time processing and binocular camera real-time 3D reconstruction of higher-resolution images make the speed of the images inconsistent and causes delays, so we needed to choose lower-resolution images in consideration of the real-time effect. Although the lower-resolution images are sufficient to support the extraction of key points of the saplings and their height measurement, the accuracy and generalisability of the system would be improved if the inexpensive system could process the high-resolution images quickly. This will be possible as hardware computing power increases and information sources become more abundant. On the one hand, the increase in computing power will allow the system to obtain better information on colour, texture and depth for learning and detection, which will certainly improve the accuracy of counting and height measurement; on the other hand, other data sources, such as UAV point clouds [17–20] and hyperspectral imagery [41–45], can effectively reconstruct the structure of individual tree types and thus help their detection.

It is clear from the experimental results that the binocular camera can complete a 3D reconstruction of the current scene and generate a depth map. The depth map contains the 3D coordinates of the pixel points, from which we can obtain the vertex and base points of the saplings, as well as the centre point. The vertex and base points were used to calculate the height of the saplings, while the centroids could be used to position them. Compared to studies that estimate tree height using point cloud images from a drone, we have the advantage that height measurements can be carried out for small saplings shorter than 30 cm, and the binocular camera is cheaper and simpler to operate. The centroid location and height estimation of saplings could provide the basic capability to sense, distinguish, measure and locate target objects in future fully automated nursery management, which in turn would enable unmanned cultivation operations, such as automatic watering, fertilization and temperature control.
However, the cheapness of the binocular camera dictates the simplicity of its hardware system architecture. As a result, compared to sensors such as UAVs and LiDAR, binocular cameras can pose larger data errors and more tedious pre-calibration and other tasks. Binocular cameras are susceptible to terrain, light and weather [46], which negatively affects processes, such as subsequent altimetry and localisation and limits the ability of binocular cameras to generalise.

In addition, the binocular camera is unable to measure the height of saplings that are heavily obscured by one another, and the binocular camera will output a height value containing a large error due to the lack of access to the key points of the saplings. For this reason, we propose a method for manually selecting key points for height measurement. This method solves the problem of low fit and missed saplings through simple human-machine interaction; in addition, by changing the position of the binocular camera, the height of some heavily occluded saplings can be detected. However, occlusion is still a problem in vision technology, and the accuracy of counting and height measurement using the binocular camera in this study was greatly reduced in the case of dense hibiscus and other saplings. To address these shortcomings, we will introduce other sensors in future work, such as using LiDAR to acquire point cloud data [17], to segment saplings for height measurement and enhance the generalisability of the system.

4.3. Experimental Errors

Although experimental errors are avoided as much as possible, some errors are inevitable due to objective conditions [47]. The errors mainly originate from the following: (1) There is always some discrepancy between the parameter calibration results and the real parameters of the binocular camera in actual use, which is due to the errors inherent in the binocular camera and the errors in the tessellation calibration images taken. (2) Saplings that are too close together can be mistakenly detected as one sapling, which is due to the system identification errors caused by the mutual forking of sapling branches. (3) There will always be a partial incomplete fit between the sapling and the sapling detection frame, and for very dense saplings, the count and measurement accuracy will be drastically reduced, which is a drawback of the Yolo series using rectangular frames as the detection tool. (4) Poor lighting conditions will lead to a reduced differentiation between saplings and their surroundings, making sapling features weaker and causing detection errors; for saplings on sloping ground, unsuitable detection points will decrease the accuracy rate. These errors can cause some discrepancies between the manually measured height TH and the system measured height H. This can be countered by applying manual intervention and finding the best camera angle, as described above. Considering that there may be saplings that are in overgrown grass, causing the roots to be obscured, we need to add the estimated height of the grass to the average height of the saplings.

To reduce binocular camera errors, the expensive Zed2 integrated binocular camera can be applied, which poses a much lower risk of calibration error and allows for higher-resolution images, but this will require more hardware, such as a computer graphics card. In addition, the option of using techniques, such as density mapping [28], to estimate the number of saplings or LiDAR may alleviate the difficulty of counting and measuring the height of dense saplings.

5. Conclusions

This study constructed and enhanced a sapling image dataset using commercially available inexpensive binocular cameras to sample nursery data, and proposed a framework for the counting and height measurement of saplings using Ghostnet–YoloV4 networks combined with binocular vision techniques. The following conclusions are drawn:

1. The Ghostnet network, which is suitable for loading into mobile devices, was introduced into the YoloV4 network with an improved PANet, and compared with networks such as Mobilenetv3–YoloV4, and it was found that the Ghostnet–YoloV4 network had the lowest total number of network parameters. Through field testing,
it was found that the Ghostnet–YoloV4 network had more than 92% accuracy for all three types of tree saplings, and the overall accuracy was still above 90% even under different light and terrain conditions. These results validate the reliability of the Ghostnet–YoloV4 network.

2. Binocular vision technology was integrated into the Ghostnet–YoloV4 network detection section to complete the 3D reconstruction of the current binocular view to obtain sapling heights. The results show that the binocular camera extracted sapling heights with an overall accuracy of 92.2%, which is sufficient to support accurate forestry in nurseries and nursery management. The field detection accuracy can also be improved if the binocular camera parameters and shooting positions are adjusted according to different light sapling morphologies and if human intervention is added.

This study constructed a new network structure with high detection accuracy and applicability. This demonstrates the feasibility of using a low-profile binocular camera and a personal computer to achieve the real-time counting and height measurement of nursery saplings. It can currently be used to help nursery staff reduce the burden of manual inspection. In the future, it could also be used to help automate forestry machinery for the real-time detection, classification, localization and acquisition of the shape and size of objects of interest around the machine, providing guidance for subsequent automated operations.

Author Contributions: Methodology, X.Y. and D.L.; resources, X.Y. and Y.M.; software, X.Y. and P.S.; writing, X.Y.; format calibration, Y.M. and G.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to them also being necessary for future essay writing.

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

References
1. Dash, J.; Pont, D.; Brownlie, R.; Dunningham, A.; Watt, M.S. Remote sensing for precision forestry. N. Z. J. For. 2016, 60, 15–24.
2. Panagiotidis, D.; Abdollahnejad, A. Accuracy Assessment of Total Stem Volume Using Close-Range Sensing: Advances in Precision Forestry. Forests 2021, 12, 6. [CrossRef]
3. Zhao, H.; Wang, Y.; Sun, Z.; Xu, Q.; Liang, D. Failure Detection in Eucalyptus Plantation Based on UAV Images. Forests 2021, 12, 1250. [CrossRef]
4. PR Newswire. US Global Precision Market Projected to Reach $6.1 Billion by 2024, at a CAGR of 9% during 2019–2024; PR Newswire: New York, NY, USA, 2019.
5. Boja, N.; Boja, F.; Teusdea, A.; Vidrean, D.; Marcu, M.V.; Iordache, E.; Duță, C.I.; Borz, S.A. Resource Allocation, Pit Quality, and Early Survival of Seedlings Following Two Motor-Manual Pit-Drilling Options. Forests 2018, 9, 665. [CrossRef]
6. Fernandez-Gallego, J.A.; Kefauver, S.C.; Gutierrez, N.A.; Nieto-Taladriz, M.T.; Araus, J.L. Wheat ear counting in-field conditions: High throughput and low-cost approach using RGB images. Plant Methods 2018, 14, 22. [CrossRef]
7. Boja, N.; Borz, S. Seedling Growth Performance of Four Forest Species with Different Techniques of Soil Tillage Used in Romanian Nurseries. Forests 2021, 12, 782. [CrossRef]
8. Long, J.; Shelhammer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2015, 39, 640–651. [CrossRef]
9. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444. [CrossRef]
10. Martinka, J. Neural networks for wood species recognition independent of the colour temperature of light. Eur. J. Wood Wood Prod. 2021, 79, 1645–1657. [CrossRef]
11. Shustrov, D. Species Identification of Wooden Material Using Convolutional Neural Networks; Lappeenranta University of Technology: Lappeenranta, Finland, 2018.
12. Wei, Q.; Chui, Y.H.; Leblon, B.; Zhang, S.Y. Identification of selected internal wood characteristics in computed tomography images of black spruce: A comparison study. J. Wood Sci. 2009, 55, 175–180. [CrossRef]
13. Mohan, S.; Venkatachalapathy, K.; Sudhakar, P. Hybrid optimization for classification of the wood knots. J. Theor. Appl. Inf. Technol. 2014, 63, 774–780.
14. Urbonas, A.; Raudonis, V.; Maskeliūnas, R.; Damaševičius, R. Automated Identification of Wood Veneer Surface Defects Using Faster Region-Based Convolutional Neural Network with Data Augmentation and Transfer Learning. Appl. Sci. 2019, 9, 4898. [CrossRef]
15. Xi, X.; Xia, K.; Yang, Y.; Du, X.; Feng, H. Evaluation of dimensionality reduction methods for individual tree crown delineation using instance segmentation network and UAV multispectral imagery in urban forest. Comput. Electron. Agric. 2021, 191, 106506. [CrossRef]
16. Zheng, Y.; Wu, G. YOLOv4-Lite-Based Urban Plantation Tree Detection and Positioning With High-Resolution Remote Sensing Imagery. Front. Environ. Sci. 2022, 14, 641. [CrossRef]
17. Rodríguez-Puerta, F.; Gómez-García, E.; Martín-García, S.; Pérez-Rodríguez, F.; Prada, E. UAV-Based LiDAR Scanning for Individual Tree Height and Measurement in Young Forest Permanent Trials. Remote Sens. 2021, 14, 170. [CrossRef]
18. Castilla, G.; Filattrautl, M.; McDermid, G.J.; Gartrell, M. Estimating Individual Conifer Seedling Height Using Drone-Based Image Point Clouds. Forests 2020, 11, 924. [CrossRef]
19. Puliti, S.; Solberg, S.; Granhus, A. Use of UAV photogrammetric data for estimation of biophysical properties in forest stands under regeneration. Remote Sens. 2019, 11, 233. [CrossRef]
20. Imanbollahi, M.; Saarinen, N.; Markelin, L.; Rosnell, T.; Näsi, R.; Hakala, T.; Honkavaara, E.; Holopainen, M.; Hyypä, J.; Vastaranta, M. Characterizing seedling stands using leaf-off and leaf-on photogrammetric point clouds and hyperspectral imagery acquired from unmanned aerial vehicle. Forests 2019, 10, 415. [CrossRef]
21. Kamilaris, A.; Prenafta-Boldu, F.X. Deep learning in agriculture: A survey. Comput. Electron. Agric. 2018, 147, 70–90. [CrossRef]
22. Liu, B.; Yu, X.C.; Yu, A.Z.; Wan, G. Deep convolutional recurrent neural network with transfer learning for hyperspectral image classification. J. Appl. Remote Sens. 2018, 12, 026028. [CrossRef]
23. Chen, Y.S.; Jiang, H.L.; Li, C.Y.; Jia, X.P.; Ghamisi, P. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Trans. Geosci. Remote Sens. 2016, 54, 6232–6251. [CrossRef]
24. Makantasis, K.; Karantzalos, K.; Doulamis, A.; Doulamis, N. Deep supervised learning for hyperspectral data classification through convolutional neural networks. In Proceedings of the IGARSS 2015—2015 IEEE International Geoscience and Remote Sensing Symposium, Milan, Italy, 26–31 July 2015; pp. 4959–4962. [CrossRef]
25. Alipourfard, T.; Arefi, H.; Mahmoudi, S. A novel deep learning framework by combination of subspace-based feature extraction and convolutional neural networks for hyperspectral images classification. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 4780–4783. [CrossRef]
26. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
27. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525. [CrossRef]
28. Lu, H.; Cao, Z.; Xiao, Y.; Zhuang, B.; Shen, C. TassellNet: Counting maize tassels in the wild via local counts regression network. Plant Methods. 2017, 13, 1–17. [CrossRef] [PubMed]
29. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. arXiv 2018, arXiv:1804.02767.
30. Zhang, C.; Li, T.; Zhang, W. The Detection of Impurity Content in Machine-Picked Seed Cotton Based on Image Processing and Improved YOLO V4. Agronomy 2021, 12, 66. [CrossRef]
31. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [CrossRef]
32. Wang, J.; Gao, Z.; Zhang, Y.; Zhou, J.; Wu, J.; Li, P. Real-Time Detection and Location of Potted Flowers Based on a ZED Camera and a YOLO V4-Tiny Deep Learning Algorithm. Horticulturae 2022, 8, 21. [CrossRef]
33. Sozzi, M.; Cantalamessa, S.; Cogato, A.; Kayad, A.; Marinello, F. Automatic Bunch Detection in White Grape Varieties Using YOLOv3, YOLOv4, and YOLOv5 Deep Learning Algorithms. Agronomy 2022, 12, 319. [CrossRef]
34. Bochkovskiy, A.; Wang, C.Y.; Liao, H. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv 2020, arXiv:2004.10934.
35. Ma, J.; Du, K.; Zheng, F.; Zhang, L.; Gong, Z.; Sun, Z. A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network. Comput. Electron. Agric. 2018, 154, 18–24. [CrossRef]
36. Ding, W.; Taylor, G. Automatic Moth Detection from Trap Images for Pest Management. Comput. Electron. Agric. 2016, 123, 17–28. [CrossRef]
37. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. GhostNet: More features from cheap operations. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 1577–1586. [CrossRef]
38. Li, W.; Chen, H.; Liu, Q.; Liu, H.; Wang, Y.; Gu, G. Attention Mechanism and Depthwise Separable Convolution Aided 3DCNN for Hyperspectral Remote Sensing Image Classification. Remote Sens. 2022, 14, 9. [CrossRef]
39. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [CrossRef]
40. Mohamed, A.; Deng, Y.; Zhang, H.; Wong, S.H.F.; Uheida, K.; Zhang, Y.X.; Zhu, M.-C.; Lehmann, M.; Quan, Y. Photogrammetric evaluation of shear modulus of glulam timber using torsion test method and dual stereo vision system. Eur. J. Wood Wood Prod. 2021, 79, 1209–1223. [CrossRef]
41. Ghiyamat, A.; Shafri, H.Z.M. A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment. *Int. J. Remote Sens.* 2010, 31, 1837–1856. [CrossRef]

42. Ghiyamat, A.; Shafri, H.Z.M.; Mandiraji, G.A.; Shariff, A.R.M.; Mansor, S. Hyperspectral discrimination of tree species with different classifications using single- and multiple-endmember. *Int. J. Appl. Earth Obs. Geoinf.* 2013, 23, 177–191. [CrossRef]

43. Yao, W.; Krzystek, P.; Heurich, M. Tree species classification and estimation of stem volume and DBH based on single tree extraction by exploiting airborne full-waveform LiDAR data. *Remote Sens. Environ.* 2012, 123, 368–380. [CrossRef]

44. Sun, C.; Huang, C.; Zhang, H.; Chen, B.; An, F.; Wang, L.; Yun, T. Individual Tree Crown Segmentation and Crown Width Extraction From a Heightmap Derived From Aerial Laser Scanning Data Using a Deep Learning Framework. *Front. Plant Sci.* 2022, 13, 914974. [CrossRef]

45. Xue, X.; Jin, S.; An, F.; Zhang, H.; Fan, J.; Eichhorn, M.P.; Jin, C.; Chen, B.; Jiang, L.; Yun, T. Shortwave Radiation Calculation for Forest Plots Using Airborne LiDAR Data and Computer Graphics. *Plant Phenomics* 2022, 2022, 9856739. [CrossRef]

46. Li, S.; Lideskog, H. Implementation of a System for Real-Time Detection and Localization of Terrain Objects on Harvested Forest Land. *Forests* 2021, 12, 1142. [CrossRef]

47. Chen, C.; Jing, L.; Li, H.; Tang, Y. A New Individual Tree Species Classification Method Based on the ResU-Net Model. *Forests* 2021, 12, 1202. [CrossRef]