drones

MDPI

Article

Deep Learning-Based Pine Nematode Trees' Identification Using Multispectral and Visible UAV Imagery

Bingxi Qin [1], Fenggang Sun [1], Weixing Shen [2], Bin Dong [2], Shencheng Ma [2], Xinyu Huo [1] and Peng Lan [1,*]

[1] College of Information Science and Engineering, Shandong Agricultural University, Tai'an 271018, China
[2] Taishan Forest Pest Management and Quarantine Station, Tai'an 271018, China
* Correspondence: lanpeng@sdau.edu.cn; Tel.: +86-156-500-911-99

Abstract: Pine wilt disease (PWD) has become increasingly serious recently and causes great damage to the world's pine forest resources. The use of unmanned aerial vehicle (UAV)-based remote sensing helps to identify pine nematode trees in time and has become a feasible and effective approach to precisely monitor PWD infection. However, a rapid and high-accuracy detection approach has not been well established in a complex terrain environment. To this end, a deep learning-based pine nematode tree identification method is proposed by fusing visible and multispectral imagery. A UAV equipped with a multispectral camera and a visible camera was used to obtain imagery, where multispectral imagery includes six bands, i.e., red, green, blue, near-infrared, red edge and red edge 750 nm. Two vegetation indexes, NDVI (Normalized Difference Vegetation Index) and NDRE (Normalized Difference Red Edge Index) are extracted as a typical feature according to the reflectance of infected trees in different spectral bands. The YOLOv5 (You Only Look Once v5)-based detection algorithm is adopted and optimized from different aspects to realize the identification of infected pine trees with high detection speed and accuracy. e.g., GhostNet is adopted to reduce the number of model parameters and improve the detection speed; a module combining a CBAM (Convolutional Block Attention Module) and a CA (Coordinate Attention) mechanism is designed to improve the feature extraction for small-scale pine nematode trees; Transformer module and BiFPN (Bidirectional Feature Pyramid Network) structure are applied to improve the feature fusion capability. The experiments show that the mAP@0.5 of the improved YOLOv5 model is 98.7%, the precision is 98.1%, the recall is 97.3%, the average detection speed of single imagery is 0.067 s, and the model size is 46.69 MB. All these metrics outperform other comparison methods. Therefore, the proposed method can achieve a fast and accurate detection of pine nematode trees, providing effective technical support for the control of a pine nematode epidemic.

Keywords: pine wood nematode disease; deep learning; multispectral imagery; YOLOv5

check for updates

Citation: Qin, B.; Sun, F.; Shen, W.; Dong, B.; Ma, S.; Huo, X.; Lan, P. Deep Learning-Based Pine Nematode Trees' Identification Using Multispectral and Visible UAV Imagery. *Drones* **2023**, *7*, 183. https://doi.org/10.3390/drones7030183

Academic Editor: Eben Broadbent

Received: 28 January 2023
Revised: 23 February 2023
Accepted: 4 March 2023
Published: 7 March 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Pine wilt disease (PWD) is one of the major forest diseases caused by *Bursaphelenchus xylophilus*. It can be transmitted by various vectors [1]. It is highly infectious and poses a serious threat to ecological security. PWD spreads rapidly once infected among pine trees; it is difficult to prevent and control and has caused serious economic losses [2]. Therefore, identifying PWD-infected trees is urgent and helps prevent and control the epidemic.

Due to the strong pathogenicity, rapid spread and high lethality, pine trees are susceptible to PWD. With the gradual increase in global temperature, more and more areas of the earth are vulnerable to the threat of PWD [1]. The needles of healthy pine trees are green in color, while those of diseased pine trees die from lack of water due to leaf stem blockage, and the needles slowly change from green to yellow or red-brown [3]. The symptoms also include reduced resin exudation, pine needles turning yellow but not falling off, and, finally, withering and dying, which usually develops in late summer to late fall [4,5]. The color change provides distinguishable information for identification. Various PWD detection

methods have been proposed using drones equipped with visible light equipment according to differences in appearance between diseased and healthy pine trees. Ren et al. [6] used a new Global Multi-Scale Channel Adaptation Network to detect pine wood nematode trees; the average precision (AP) and the recall are 79.8% and 86.6%, respectively. Zhang et al. [7] proposed a U-Net-based discolored wood recognition method with imagery collected by fixed-wing UAVs and achieved an average accuracy of 95.17%. Liu et al. [8] used a full convolutional neural network (CNN) for pixel recognition of infected trees in visible imagery and achieved a detection accuracy of 97.86%. Tao et al. [9] proposed a CNN-based method for detecting PWD-infected trees with an accuracy of 97.38% in forest areas. Liu et al. [10] achieved an average detection accuracy of 85.45% using the Relief feature selection algorithm with a VGG (Visual Geometry Group) network combined with an attention mechanism. Xu et al. [11] used a two-stage target detection model to identify and achieve 82.42% accuracy. Huang et al. [12] used the improved YOLOv4 (You Only Look Once v4) [13] model for PWD identification and achieved an average detection accuracy of 87.92%. Sun et al. [14] used lightweight SSD300 to detect pine nematode trees and achieved an average detection accuracy of 97.22%. Zhou et al. [15] combined preprocessing means, such as imagery segmentation, imagery enhancement and coordinate conversion with improved residual networks to achieve 87% accuracy for diseased tree recognition. Hu et al. [16] designed a diseased tree recognition method combining deep CNN, deep convolutional generative adversarial network and an Ada-Boost classifier to achieve effective recognition in complex backgrounds. Oide et al. [17] used artificial neural networks combined with RGB and HSV (hue, saturation, value) color space datasets to achieve improved PWD-detection accuracy.

However, the visible imagery only contains the information in R, G, and B bands. PWD detection may suffer from several challenges, such as high altitude and a complex feature environment, which may degrade the detection performance in terms of accuracy and efficiency. One idea is a two-stage detection: first, coarse-grained identification by the UAV's onboard equipment, and then transmission of the remaining images to the ground station for further precise identification [18]. However, the accuracy of this method still needs to be improved. The multispectral equipment can acquire multispectral imagery, including more spectral bands such as green, red, near-infrared and infrared edge imagery information [19]. Multispectral imagery divides the feature spectral information into several spectral bands, which can obtain more information from many spectral channels and uncover the feature characteristics hidden in a narrow spectral range, thus meeting the accuracy requirements for PWD detection under complex ground environments [20]. Iordache et al. [21] studied the spectral reflectance characteristics and confirmed that the red wavelength (688 nm) region showed the most significant difference between healthy and infected trees. Kim et al. [22] investigated ten vegetation indices for PWD detection. They found that the indexes using red and infrared wavelengths showed greater changes before and after pine trees were infected by PWD.

As a result, the use of multispectral and hyperspectral devices to collect spectral information for PWD identification is gradually increasing. Iordache et al. [23] found that using the traditional RF (Random Forest) classification algorithm combined with multispectral images or hyperspectral images, the classification accuracy of both was higher than 91%. Qin et al. [24] proposed a SCANet-based neural network algorithm by utilizing UAV multispectral imagery with an overall accuracy of 79%. Ren et al. [25] designed a 3D CNN for early PWD detection and achieved an average recognition accuracy of 81.06%. Hae et al. [26] used multispectral data to generate vegetation indices, combined with visible light imagery to train a deep learning model, which was applied to detect suspected PWD trees, and achieved 86.63% mAP. Zhou et al. [27] proposed a multi-band image-fusion infected pine tree detector (MFTD), which can accurately detect the infected pine trees, especially those at the early stage, and the average precision values (AP@50) are 87.2%, 93.5%, and 84.8% for early, middle, and late stages, respectively.

In summary, most existing works are based on visible imagery for recognition, which is affected by terrain and features and are difficult to apply accurately in complex environments. For example, PWD usually occurs in a complex environment with various tree species, vegetation, bare ground and shadows, which may interfere with identifying infected trees and degrade the identification performance. Therefore, single visible imagery-based or single multispectral imagery-based recognition methods will inevitably be affected by the phenomenon of "same object with different spectrums" and "different objects with same spectrum", leading to various false detections and missed detections and affecting the recognition accuracy. In this paper, by combining visible imagery and multispectral imagery, a YOLOv5-based [28] detection algorithm is proposed to extract more spectral features of pine nematode trees. For the problem of a large number of UAV images and slow detection speed, we use GhostNet [29] to reduce the model complexity and improve the detection speed; for the problem that the model is difficult to detect small-scale pine nematode trees, we use a module combining a CBAM (Convolutional Block Attention Module) [30] and a CA (Coordinate Attention) [31] mechanism to improve the feature extraction ability; we use Transformer [32] module and BiFPN (Bidirectional Feature Pyramid Network) [33] structure to improve the feature fusion ability.

The main contributions of this paper are as follows:

- We compared the characteristic differences in multispectral bands between healthy and PWD-infected trees by conducting experiments with data collected in the field.
- We conducted experiments by fusing multispectral and visible images and conducted multiple comparisons and ablation experiments.
- A deep learning YOLOv5l-based PWD-detection approach is proposed by combining multispectral and visible UAV imagery.

2. Materials and Methods

2.1. Preparation

The study area is located in the eastern region of Qingdao City, Shandong Province of China (36°8′28.032″~36°8′32.748″ N and 120°37′55.74″~120°43′26.832″ E). The total area of the data collection area is about 6 km^2, as shown in Figure 1. The pine species in the area are mainly larch (*Larix gmelinii*), red pine (*Pinus koraiensis*) and black pine (*Pinus thunbergii*).

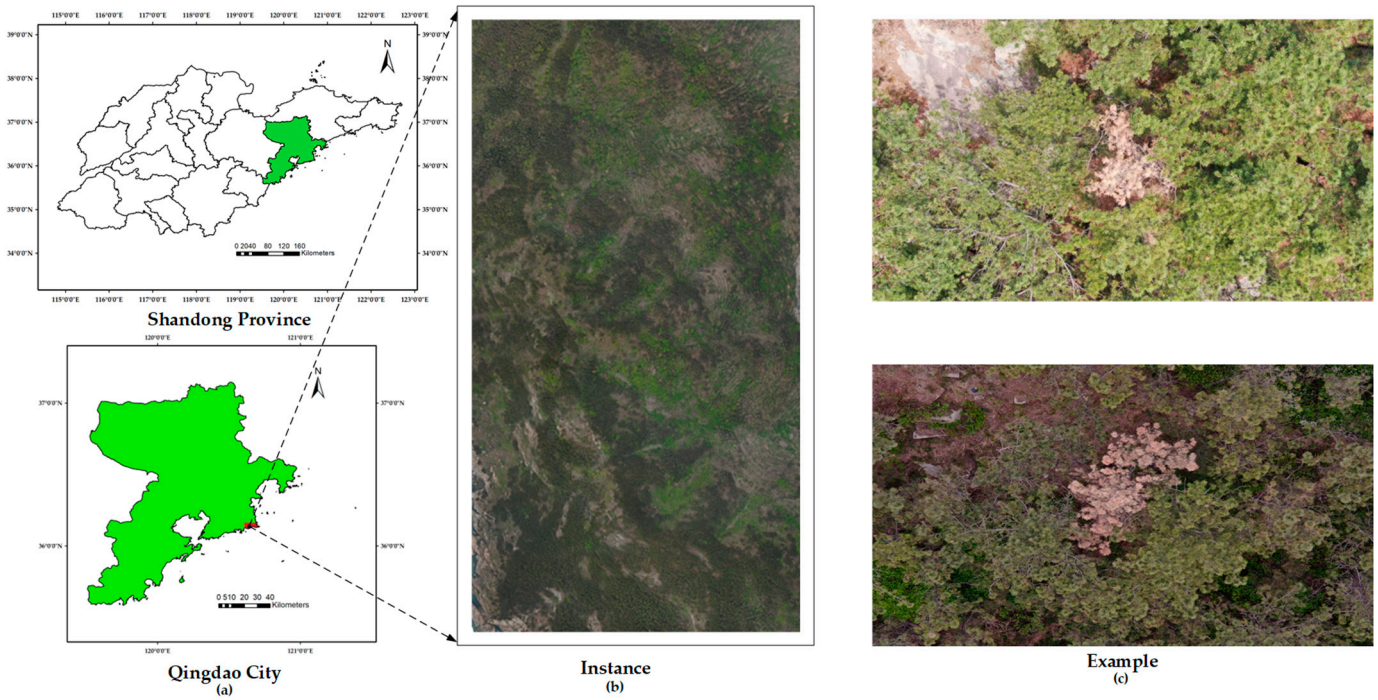


Figure 1. Location of study area (**a**) instance (**b**) and some examples (**c**).

The images were collected by a UAV (DJI M300 RTK) equipped with a multispectral camera (Changguang Yuchen MS600 Pro) and a visible light camera (Zenith H20). The multispectral camera captures images in six bands: red, green, blue, near-infrared, red edge and red edge 750 nm. The images were collected during 2–8 May 2022 and 1–10 July 2022, with UAV flight altitudes of 100 m and 350 m. The specific flight parameters are shown in Table 1. Finally, 3500 visible images and 1000 multispectral images were acquired.

Table 1. Specific flight parameters.

Flight Parameters	Visible Image	Multispectral Image
Flight altitude (m)	350	100
Flight speed (m/s)	15	10
Heading overlap rate (%)	80	85
Sideways overlap rate (%)	80	85
Shooting interval (s)	8	5

Multispectral images cannot be used directly, and multispectral and visible images may contain different resolutions. Therefore, the original images need several pre-processing procedures. (1) The multispectral images are pixel normalized by standard reflectance to adapt them as visible images for model training. To match with the channels of visible images, the DN (Digital Number) of multispectral images in six bands are normalized to the same range as the visible image, as shown in Figure 2. (2) Image cropping (e.g., image scaling and segmentation): as multispectral images and visible images are collected from different angles and heights, uniform scaling and cropping is required for calibration. The original multispectral images are 1280 × 960 pixels, uniformly scaled to 640 × 640 pixels. The original visible images are 5184 × 3888 pixels, and they are split into several smaller images with 1280 × 1280 pixels to avoid some targets being segmented and truncated between two adjacent small images; finally, an overlapped region is set between two adjacent small images, as shown in Figure 3.

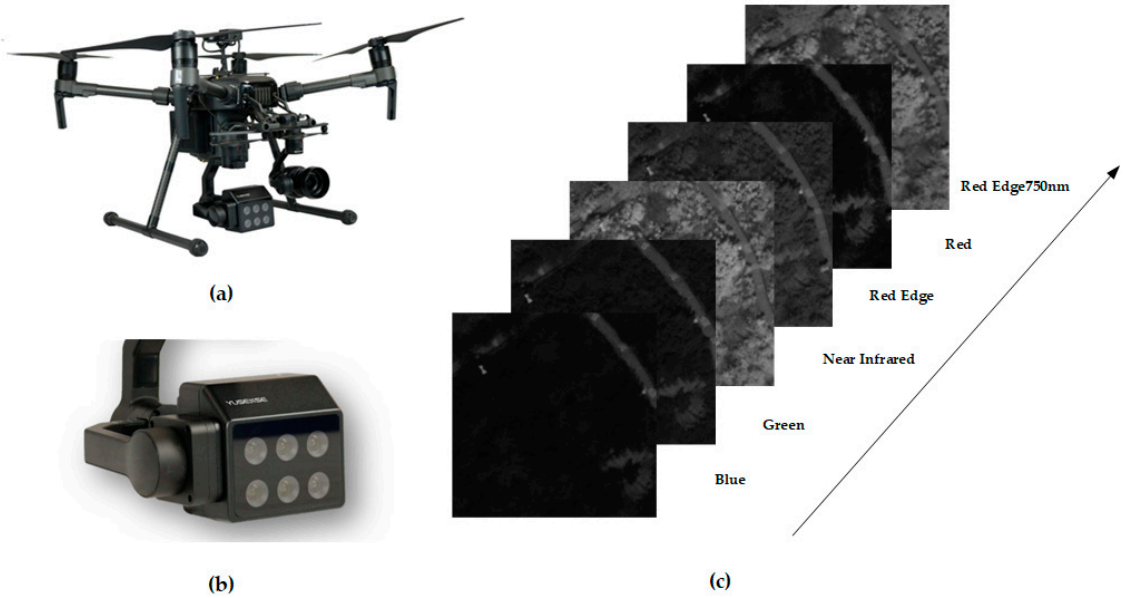


Figure 2. DJI M300RTK UAV (**a**) with MS600 Pro multispectral camera (**b**), it has six bands (**c**).

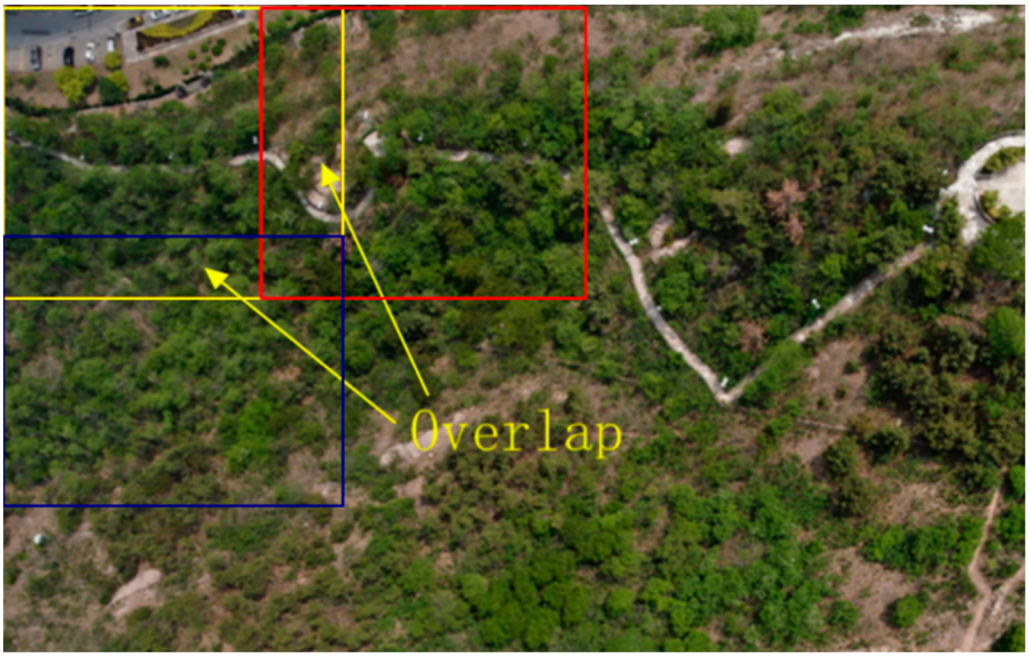


Figure 3. Overlap area setting. The yellow, red and blue boxes represent the small images after segmentation. The overlapping area is set to avoid the target being segmented and truncated.

The LabelImg tool was used to label the collected pine nematode images. The labeled box positions (xmin, xmax, ymin, ymax) and corresponding category information (sicktree) were saved as txt files to form a YOLO format dataset. In total, 4500 images (3500 visible images and 1000 multispectral images) are collected by UAV. As some images do not contain any PWD-infected trees, and after some per-processing, such as screening and image cropping, 1958 visible images and 965 multispectral images that contain PWD infected trees were kept for the following training and testing. The 1958 visible images are divided into training and validation sets with 1615 and 343 images according to the ratio 8:2, respectively. The multispectral images could not be used directly for recognition and are divided into training sets. The training set contains 2580 images, and to increase training volume to improve generalization ability, they are augmented to 15256 images by using DCGAN (Deep Convolution Generative Adversarial Networks) [34] and data enhancement (e.g., brightness changing, pair scaling, random rotation, vertical flip, diagonal flip, mirror flip, etc.). In addition, the validation set obtained 2058 images after image segmentation and filtering. The number of images in different sets is shown in Table 2.

Table 2. Number of images of infected trees.

Imagery Type	No. of Original Images	No. of Images in Training Set	No. of Images with Expansion in Training Set	No. of Images in Validation Set (Original)	No. of Images in Validation Set (Image Splitting)
Visible	1958	1615	9550	343	2058
Multispectral	965	965	5706	0	0
Total	2923	2580	15,256	343	2058

2.2. Feature Spectral Band Selection for Multispectral Images

For multispectral images, the feature spectral bands are selected by comparing reflectance differences of different bands between healthy and PWD-infected trees. The bands with high reflectance variability between healthy and PWD-infected trees are kept as the feature spectral bands, which are used to reduce training burden of the deep learning model. The reflectance is firstly extracted from six bands (red, green, blue, near-infrared, red edge and red edge 750 nm) of the multispectral image.

Then, the Normalized Difference Vegetation Index (NDVI) and Normalized Difference Red Edge Index (NDRE) can be used to measure vegetation health levels [35], and they are calculated as

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

$$NDRE = \frac{NIR - RED\ EDGE}{NIR + RED\ EDGE} \quad (2)$$

where NIR, RED and $RED\ EDGE$ denote the reflectance in the bands of near-infrared, red and red edge, respectively.

The comparison between healthy and infected trees is shown in Figure 4, and the comparison in NDVI and NDRE between healthy and infected trees is presented in Figure 5. For clarity, the comparison of reflectance in six bands, NDVI and NDRE, is also given in Table 3, including their range and median values.

As can be seen, the diseased and healthy trees exhibit a different reflectance in the six multispectral bands. For example, the reflectance range for the infected trees in red band is 5765~11,269, and the median is 8805, and for the healthy trees is 2461~5091 with a 3595.5 median. This is because infected trees die, causing the yellowing of pine needles with different water content than healthy trees, reflecting more red light. Similarly, in the near-infrared band, the reflectance varies from 12,677 to 21,987, and the median is 15,865 for infected trees and from 15,360 to 22,329 with a median of 19,371 for healthy trees. This is because infected trees have lower chlorophyll content than healthy trees due to dieback and absorb more light in the near-infrared band than healthy trees. In general, for the six multispectral bands, the reflectance differences in red, near-infrared, red edge and red edge 750 nm are the most obvious, and in green and blue bands, they are less obvious. Therefore, the bands of red, near-infrared, red edge and red edge 750 nm are selected as the feature spectral bands to distinguish PWD-infected trees from healthy trees. In terms of vegetation indexes, the differences in NDVI and NDRE between diseased and healthy trees are also significant. Therefore, NDVI and NDRE can also be used as effective indicators to identify PWD-infected trees and healthy trees.

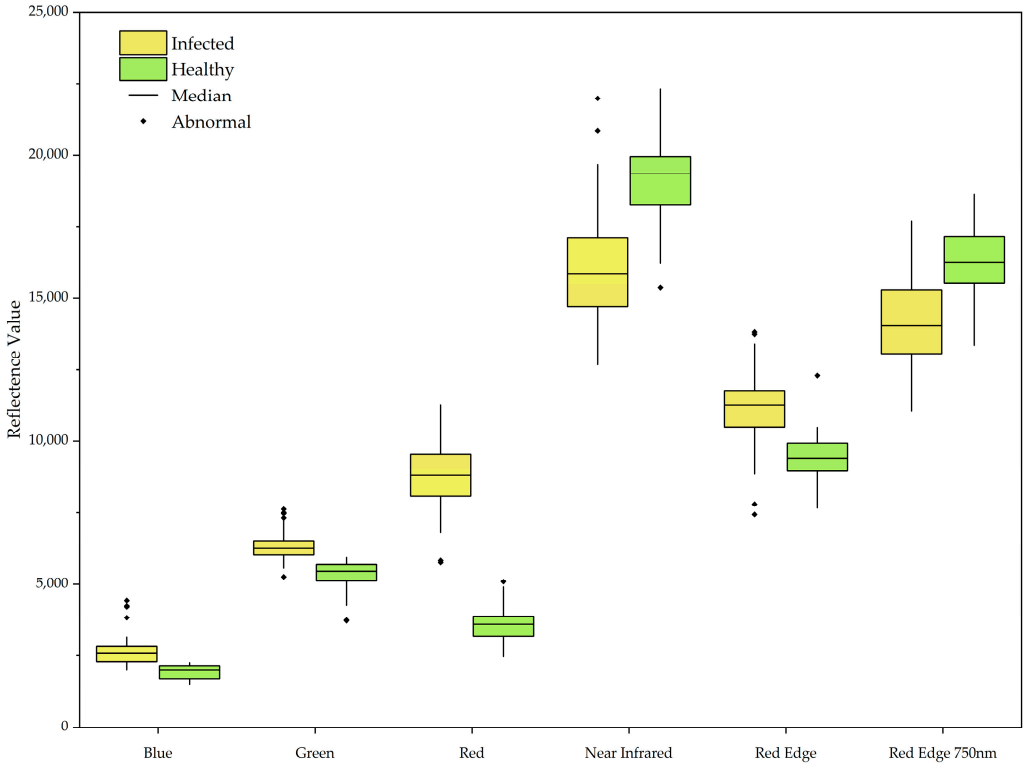


Figure 4. Comparison of reflectance values of infected trees and healthy trees in six bands.

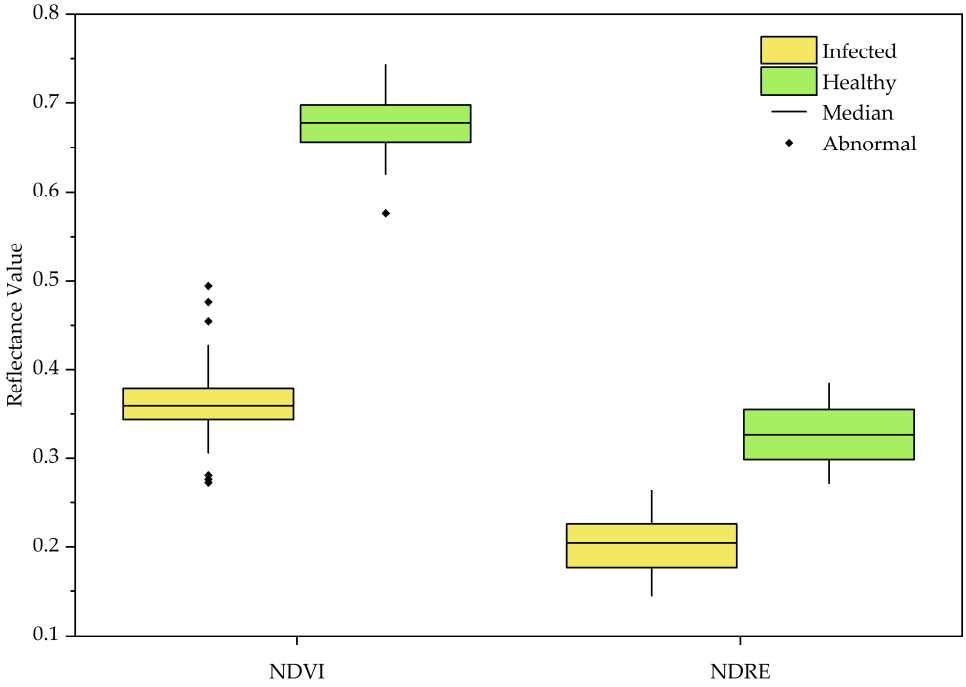


Figure 5. Comparison of the differences in reflectance values between infected and healthy trees in the two vegetation indices.

Table 3. Comparison of reflectance between diseased and healthy trees in six bands and two vegetation indices.

Band and Vegetation Index	Range of Changes		Median of Reflectance Range	
	Infected Trees	Healthy Trees	Infected Trees	Healthy Trees
Red	5765~11,269	2461~5091	8805	3595.5
Green	5261~7619	3725~5967	6251	5393.5
Blue	2003~4431	1500~2225	2449	1783.5
Near Infrared	12,677~21,987	15,360~22,329	15,865	19,371
Red Edge	7435~13,817	7660~12,297	11,258	9412.5
Red Edge 750 nm	11,053~17,694	13,355~18,638	14,038	16,254.5
NDVI	0.273~0.495	0.576~0.744	0.359	0.678
NDRE	0.145~0.264	0.272~0.385	0.204	0.327

Therefore, the bands in red, near-infrared, red edge and red edge 750 nm for multispectral images and the two vegetation indices NDVI and NDRE were significantly different in diseased and healthy trees and were selected as the feature bands. Combining these features and visible images to construct the detection model is expected to improve the detection performance compared to single visible imagery-based detection models.

2.3. Improved YOLOv5l-Based Detection Method

In this section, YOLO v5l [28] is used as the basic model to realize PWD detection, the input of which includes multispectral images, visible images and vegetation index images. YOLO v5l, a classical single-stage target detection algorithm, is small, fast in training and inference, and can be flexibly deployed and used. YOLO v5l consists of four main components: input layer, backbone network, bottleneck network, and prediction layer. The input layer is responsible for the image processing strategy and anchors the generation mechanism. The backbone network extracts features at different scales by multilayer convolution. The bottleneck network Neck uses a Feature Pyramid Network (FPN) structure to provide detection capabilities for targets at different scales. The prediction layer Prediction generates class, probability and location information of the targets to

be detected by applying three prediction branches for targets at different scales. The underlying structure of YOLO v5l is shown in Figure 6.

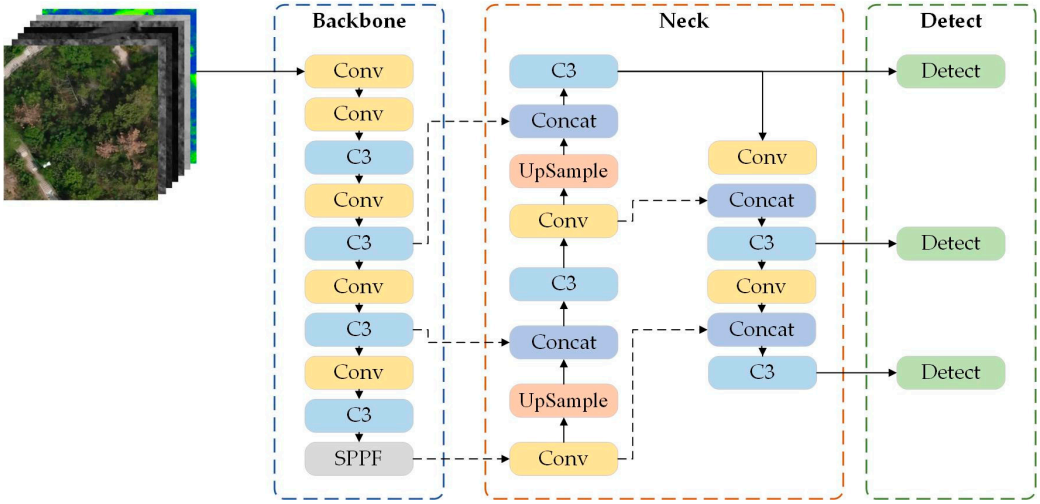


Figure 6. YOLOv5l Base Model Structure Diagram.

In complex terrain environments, the traditional YOLO v5l model has a weak feature extraction ability for infected trees due to shading and background, low detection accuracy, and is easy to occur false detection, and missed detection. Moreover, to cover a large-scale forest area, massive images are generated. To this end, we improve the YOLO v5l model from different aspects, including enhancing the model's ability to effectively extract and fuse multi-scale and multi-species features of multispectral images and visible images and reducing the complexity of the model, etc. The improved structure diagram is shown in Figure 7.

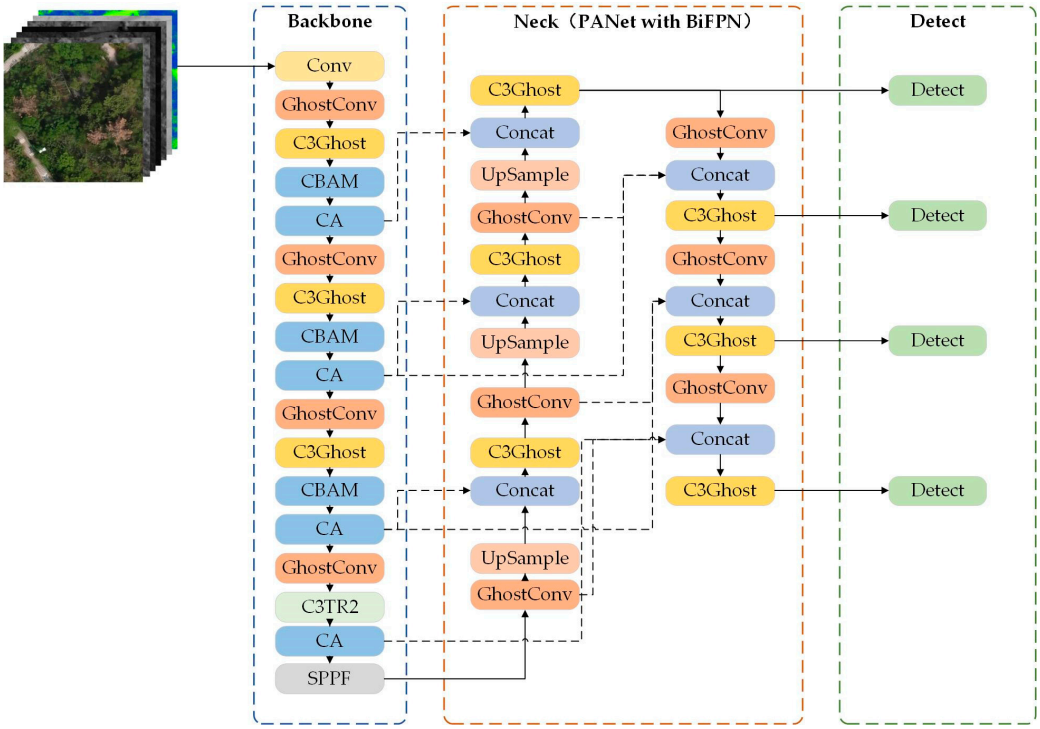


Figure 7. YOLOv5l improved model structure diagram.

(1) Model reduction with phantom network GhostNet.

Model detection speed and accuracy is the primary concern for PWD detection. In feature maps extracted by deep neural networks, rich or even redundant information usually ensures a comprehensive understanding of the input data but takes up a large number of computational resources and increases the complexity of the model. The idea of GhostNet [29] uses a series of linear variations to generate many "phantom" feature maps that can uncover the required information from the original features at a small cost to reduce the high computational complexity of extracting all the redundant features and improve the detection speed. The overall structure is shown in Figure 8.

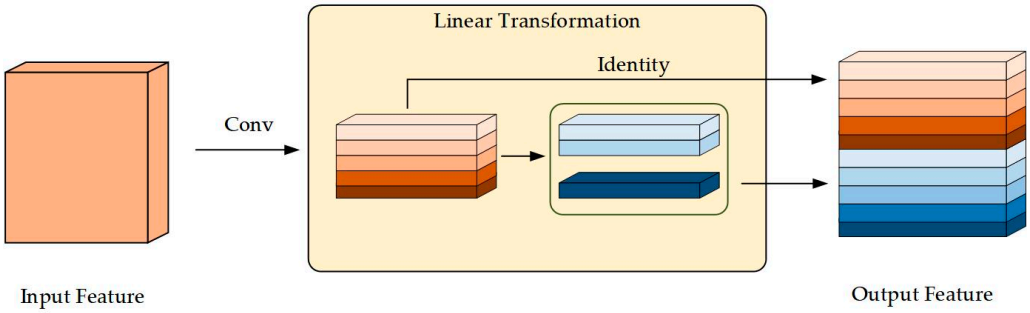


Figure 8. GhostNet overall structure diagram.

(2) Model feature extraction enhancement by combining CBAM and CA attention mechanisms.

A complex terrain environment interferes greatly with feature extraction of infected trees, and more attention to infected trees' features is needed to suppress the interfering features. The attention mechanism is used to increase the representational power of the network: focus on important features and suppress unnecessary features. CBAM [30] combines channel attention and spatial attention and consists of a channel attention module and a spatial attention module, whose structure is shown in Figure 9.

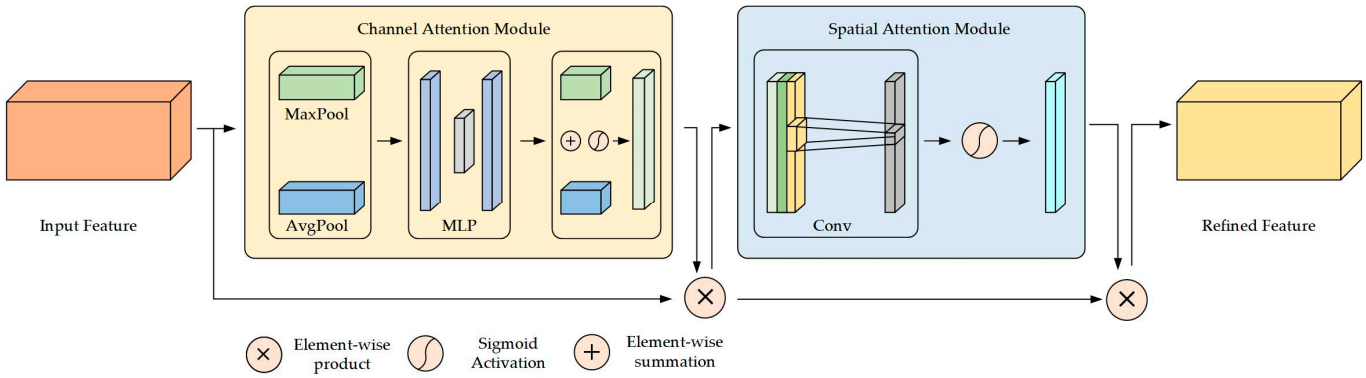


Figure 9. CBAM structure diagram.

In the channel attention module (CAM), the channel attention map is generated using the channel relationship between features. CAM selects "what" is meaningful to focus attention on the input image and computes the channel attention by compressing the dimensionality of the input feature mapping space. That is, the feature map F with size of $C \times H \times W$ (number of channels C, height H, width W) is passed through the average pooling layer $AvgPool$ and the maximum pooling layer $MaxPool$ to obtain two channel maps, respectively, which are subsequently mapped by a multilayer perceptron (MLP) shared by the two layers to obtain the one-dimensional channel attention map Mc. In the spatial attention module (SAM), the spatial attention map is generated by using the spatial relationship between features. SAM, as a supplement of CAM, pays more attention to the meaningful part "where", and the feature map F_1 generated by the CAM passes through

AvgPool and *MaxPool* to aggregate the channel information of one element map in turn to generate two two-dimensional (2D) feature maps. Then a standard convolution layer is used for connection and convolution operations to obtain the 2D spatial attention map *Ms*. *Ms* is multiplied with the feature map F_1 to obtain the final output feature map f.

Although CBAM brings a relatively significant performance improvement, channel attention usually ignores location information that is important for generating spatially selective attention maps, so the Coordinate Attention (CA) [31] module is added after CBAM, it can capture not only cross-channel information but also direction awareness and location awareness information. The CA module encodes the channel relationships and long-range dependencies with precise location information. The structure of CA is shown in Figure 10.

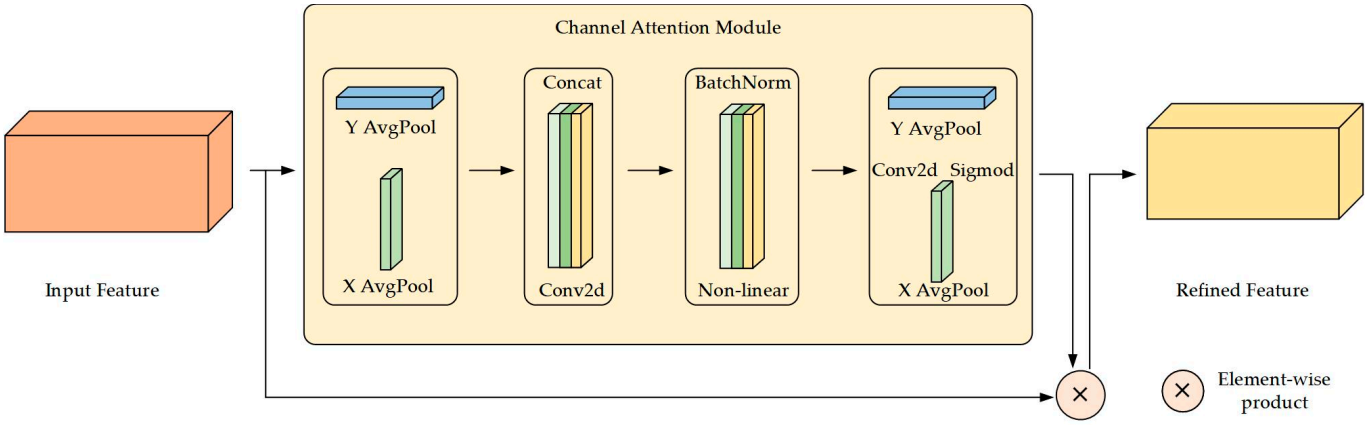


Figure 10. CA structure diagram.

The input feature map F is divided into width and height directions for global averaging pooling to obtain the feature maps in the width and height directions. Then, the feature maps in the width and height directions are stitched together to obtain the global perceptual field. After which they are fed into a convolution module with a shared convolution kernel of 1×1 to reduce their dimension to the original C/r; then the batch normalized feature map F_1 is fed into the Sigmoid activation function to obtain a feature map f in the shape of $1 \times (W + H) \times C/r$. The feature map f is convolved with a 1×1 kernel according to the original height and width to obtain the feature maps F_h and F_w with the same number of channels as the original one, followed by the Sigmoid activation function to obtain the attention weights g^h in the height and width of the feature map and g^w in the width direction, respectively. Then, the final feature maps with attention weights in the width and height directions are obtained by multiplying and weighting the original feature maps.

(3) Improvement of global information acquisition via Transformer.

In complex environments, the local information of trees depends on the global information of their surroundings. However, CNN prefers to acquire local information and is weak in acquiring global information. For this reason, Transformer [32] is introduced to improve the ability to capture global information. Transformer is essentially a model based on self-attention, and the overall structure is macroscopically an encoder–decoder structure, where Encoder uses a multi-headed attention mechanism. The Encoder module structure is shown in Figure 11.

(4) Feature fusion enhancement by using BiFPN structure.

Although CBAM and CA attention mechanisms are added in the feature extraction process to enhance the feature extraction capability for multi-scale targets, as the convolution goes deeper, features of large objects are easily retained and features of small objects are more easily ignored as they go further back. In the case of pine nematode detection work, the feature extraction and fusion ability of pine nematode trees in the imagery is too small due to the UAV flight altitude problem, which is more difficult compared to large-scale targets, so an improved feature pyramid BiFPN structure is used to enhance the feature

fusion ability for multi-scale targets. Its layer structure in YOLOv5 is shown in Figure 12. BiFPN is a weighted bidirectional feature pyramid network proposed in EfficientDet [33] that introduces learnable weights to learn the importance of different input features while iteratively applying top-down and bottom-up multi-scale feature fusion.

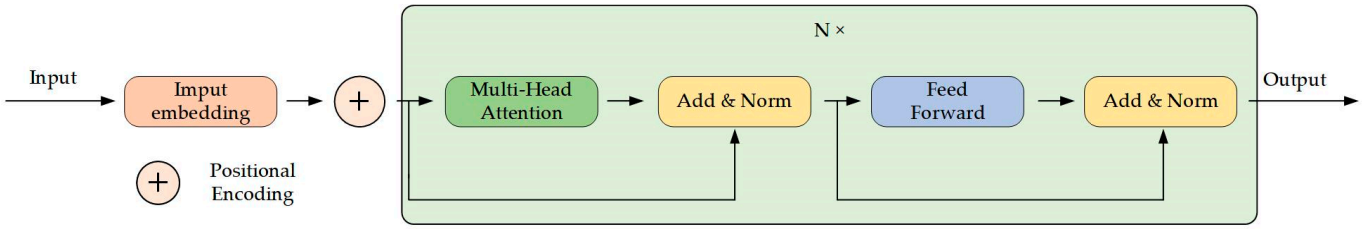


Figure 11. Encoder module structure diagram.

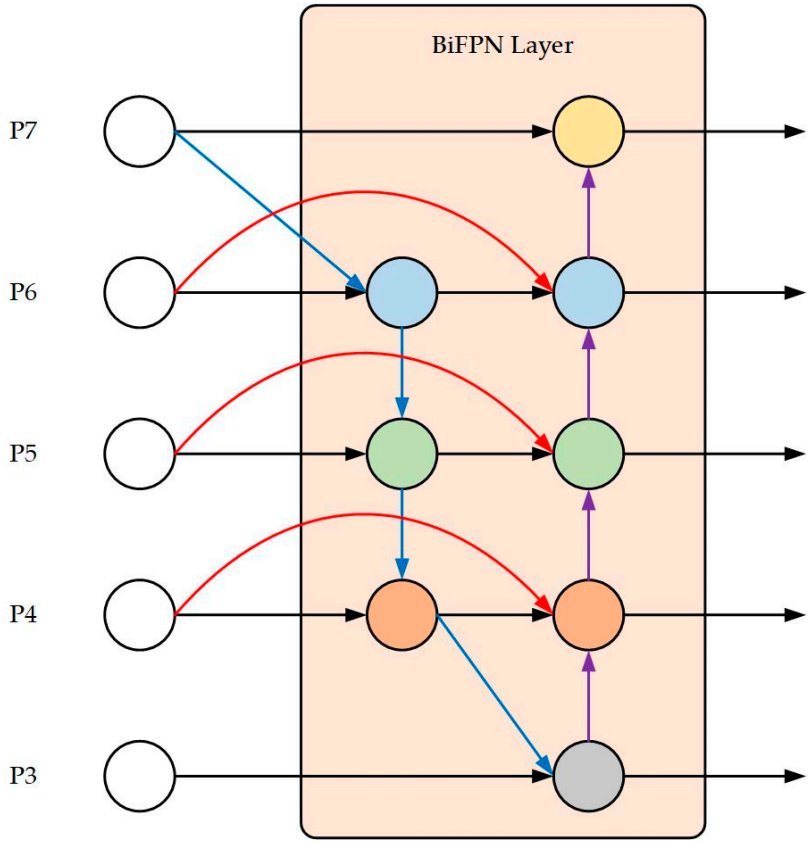


Figure 12. BiFPN Layer structure diagram in YOLOv5. The circles represent features. Black arrows represent the feature output direction; red arrows represent features that have gone through one layer of fusion; blue arrows represent top-down feature fusion; purple arrows represent bottom-up feature fusion.

3. Results and Discussion

3.1. Experimental Environment and Evaluation Index

The Pytorch deep learning framework is used for the model built, and model training and testing are performed under Windows 10 64-bit system. The CPU is AMD Ryzen 7 5800 H with a 3.2 GHz benchmark frequency and 16 GB RAM, and the GPU is NVIDIA GeForce RTX 3060 with 6 GB video memory.

A total of 17,314 images were obtained after pre-processing and augmenting, among which 15,256 images were used for training and 2058 were for validation. The visible, red, near-infrared, red edge, red edge 750 nm, NDVI and NDRE images were combined for training. Stochastic gradient descent (SGD) was used to optimize the network and

speed up training. The learning rate is 0.01, the SGD momentum parameter is 0.9, and the weight decay parameter is 0.0005. A warm-up strategy is used for training, and the training learning rate is 0.0001 within the first three epochs, after which it reverts to the preset initial learning rate. The training input images (visible and multispectral images) are uniformly 512 × 512 pixels, the batch size is 16, and the epoch is 500.

The following evaluation metrics, including mean average precision (mAP), the number of parameters (in MB) and the mean detection time (in s/sheet), are used to examine model performance. AP is the integral of the PR curve with Precision as the horizontal axis and Recall as the vertical axis. Precision P reflects the ability to classify samples of the dataset, Recall R reflects the ability to find the positive sample, and the mAP represents the average AP of all categories. The above indicators are calculated as follows:

$$P = \frac{TP}{TP + FP} \tag{3}$$

$$R = \frac{TP}{TP + FN} \tag{4}$$

$$\mathrm{m}AP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{5}$$

where TP (true positive) and FP (false positive) denote the number of positive samples with correct and incorrect predictions, respectively. FN (false negative) denotes the number of negative samples with incorrect prediction and N represents the number of categories of data. Positive and negative samples are judged by the Intersection over Union (IoU) threshold between the predicted and actual regions. If both IoU exceeds a certain threshold, then it is a positive sample, and vice versa; it is a negative sample. The mAP@ 0.5 is the mean AP of all categories at an IoU threshold of 0.5. The average detection speed is the average time the model takes to detect a single image in the validation set.

3.2. Experimental Results

Extensive experiments are provided in this part to verify the effectiveness of the improved model.

Firstly, this section provides a comparative performance analysis of YOLOv5l and the improved model over different datasets, as well as a comparative analysis of the ablation experiments of different improved schemes on the optimal dataset. Dataset 1 is the original visible images collected by UAV, with a total of 1958 images; dataset 2 is the augmented visible images after image segmentation and expansion, with a total of 9550 images; dataset 3 is an expansion of dataset 2 by adding 5706 augmented multispectral images, with a total of 15,256 images. Table 4 shows the performance of YOLOv5l and the improved model over the three different data sets. Table 5 shows the performance of the ablation experiments, where the evaluation indexes include mAP@0.5, number of parameters (in MB), and average detection time (in s/sheet). Figure 13 shows the comparison of mAP@0.5 with the changes of epochs, and Figure 14 shows the comparison of mAP@0.5 for different improved ablation experiments with the changes of epochs over the optimal data set.

Table 4. Experimental results of comparing the base model with the improved model.

Model	Experiment Name	Precision	Recall	mAP@0.5	Parameters (in MB)	Detection Time (in s/sheet)
Base YOLOv5l	Dataset 1	0.802	0.630	0.643	91.11	0.131
	Dataset 2	0.891	0.844	0.843	91.09	0.125
	Dataset 3	0.921	0.884	0.918	91.12	0.121
Improved YOLOv5l	Dataset 1	0.830	0.657	0.700	46.65	0.061
	Dataset 2	0.924	0.857	0.894	46.75	0.064
	Dataset 3	0.981	0.973	0.987	46.69	0.067

Table 5. Experimental results of different improved ablation of the optimal model.

Model and Experiment Name	Mark Name	Import Module Name				Precision	Recall	mAP@0.5	Parameters (in MB)	Detection Time (in s/sheet)
		GhostNet	CBAM and CA	Trans-former	BiFPN					
Base YOLOv5l + Dataset 3	M1					0.921	0.862	0.918	91.12	0.121
	M2	√				0.898	0.778	0.900	32.00	0.041
	M3	√	√			0.967	0.910	0.944	44.40	0.064
	M4	√	√	√		0.987	0.936	0.973	45.60	0.066
	M5	√	√	√	√	0.981	0.973	0.987	46.69	0.067

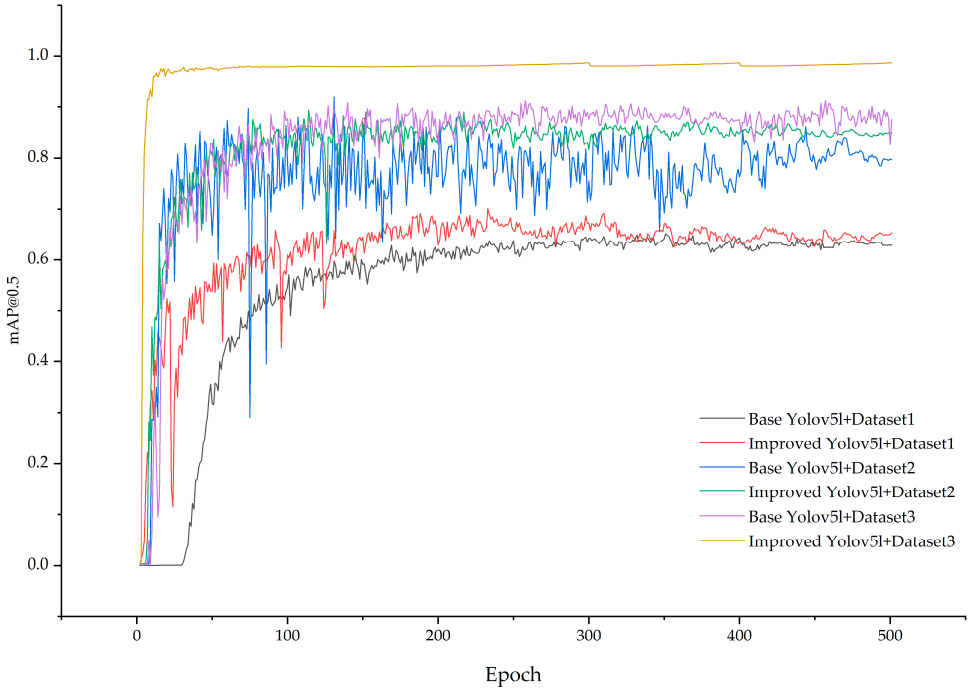


Figure 13. Comparison of mAP@0.5 under different experiments.

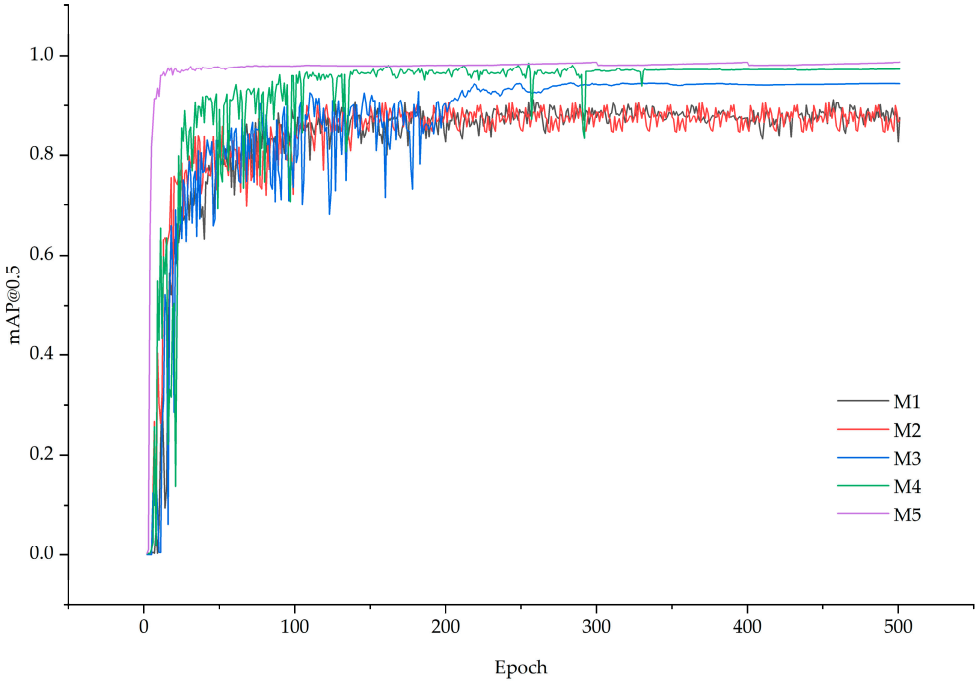


Figure 14. Comparison of mAP@0.5 under different ablation experiments.

From Table 4, it can be concluded that the detection performance of the improved YOLOv5l model is better than the YOLO v5l model under different datasets. mAP@0.5 performance is improved by about 5.7%, 5.1% and 9% over the three data sets, respectively; the number of parameters is reduced by about 1/2, and the average detection time is reduced to 0.06 s/sheet. Therefore, the effectiveness of the improved model can be verified. The detection performance of both the improved YOLOv5l and YOLOv5l models differed significantly over different datasets. Dataset 1 provides the worst detection performance as only original visible images are used. After the image augments to form dataset 2, the performance of both models is improved by about 20% and 19.4%, respectively. Dataset 3 is formed by combining dataset 2 and augmented multispectral images, and the performance of both models is further improved by 5.4% and 9.3%. Therefore, the fusion of visible and multispectral images will significantly improve the detection performance. Among all the experiments, the proposed model has the highest detection accuracy of 0.987, the highest precision of 0.981, and the highest recall of 0.973 on dataset 3.

Then, the ablation experiment is performed on dataset 3 in Table 5. We call the base model M1; M2 is the application of the Ghost module on top of M1, M3 is the application of CBAM and CA on top of M2, M4 is the application of the transformer encoder on top of M3, and M5 is the application of BiFPN structure on top of M4. The Ghost module can greatly reduce the number of model parameters and improve the detection speed by a linear transformation, but at a sacrifice of slight mAP@0.5 loss. By combining CBAM and CA to form a complementary attention mechanism to improve the extraction of important target features, the mAP@0.5 performance is improved by 4.4%, the number of parameters increases to 44.40 MB, at the cost of a slight increase of parameter numbers, and slower detection speed. With almost the same performance in parameter number and detection speed, the addition and Transformer's Encoder and BiFPN can achieve the extraction of global feature information with enhanced feature fusion capability. Thus the precision, recall and mAP@0.5 performance can be finally improved to 98.1%, 97.3% and 98.7%, respectively. Therefore, compared to YOLOv5l, the improved YOLOv5l model can achieve higher detection accuracy and speed with fewer parameters, which verifies the effectiveness of the improvements.

From Figures 13 and 14, we can see that the proposed model provides the best mAP@0.5 performance under different datasets with the increase of epochs. Among these, the proposed model has the fastest and smoothest convergence speed without large fluctuations in dataset 3 and is significantly higher than that for other conditions. Furthermore, under dataset 3, the ablation experiment shows that the improved scheme in each step can enhance the detection performance. The final improved YOLOv5l model achieves the best mAP@0.5 performance with faster convergence and smaller convergence fluctuation. Therefore, the above results show that the fusion of visible images and multispectral graphics can effectively identify infected trees in various environments; meanwhile, the proposed model has obvious advantages in terms of detection accuracy, size and speed.

To further test the performance of the improved YOLOv5l model, we compare the performance of the improved YOLOv5l model with some other classical models, including Base YOLOv5l, Faster R-CNN (Regions Convolutional Neural Networks) [11], YOLOv4 [12], SSD300 (Single Shot MultiBox Detector) [14], YOLOv7 [36] and YOLOX [37]. In experiments, the initial learning rate is 0.01, the epoch number is 500, and the images are scaled to 512 $\times$ 512 pixels. Dataset 3, containing both multispectral and visual images, is used, and the performance is evaluated in terms of mAP@0.5, the number of parameters (in MB), and average detection time (in s/sheet). The results are shown in Table 6.

Table 6. Performance comparison of common target detection models.

Model	Precision	Recall	mAP@0.5	Parameters (in MB)	Detection Time (in s/sheet)
Improved YOLOv5l	0.981	0.973	0.987	46.69	0.064
Base YOLOv5l	0.921	0.862	0.918	91.12	0.121
Faster R-CNN [11]	0.830	0.712	0.775	548.32	1.94
YOLOv4 [12]	0.946	0.749	0.893	256.26	1.02
SSD300 [14]	0.941	0.812	0.940	95.01	1.04
YOLOv7 [36]	0.923	0.937	0.968	149.18	1.05
YOLOX [37]	0.926	0.904	0.948	36.01	1.04

As can be seen, the improved YOLOv5l achieves the optimal performance in terms of mAP@0.5 and detection time, and only the number of parameters is slightly larger than that of YOLOX. Therefore, the improved YOLOv5l model can provide the best detection performance compared to other models. Further, the recognition results of the proposed model and the comparison model in practical applications are shown in Figure 15.

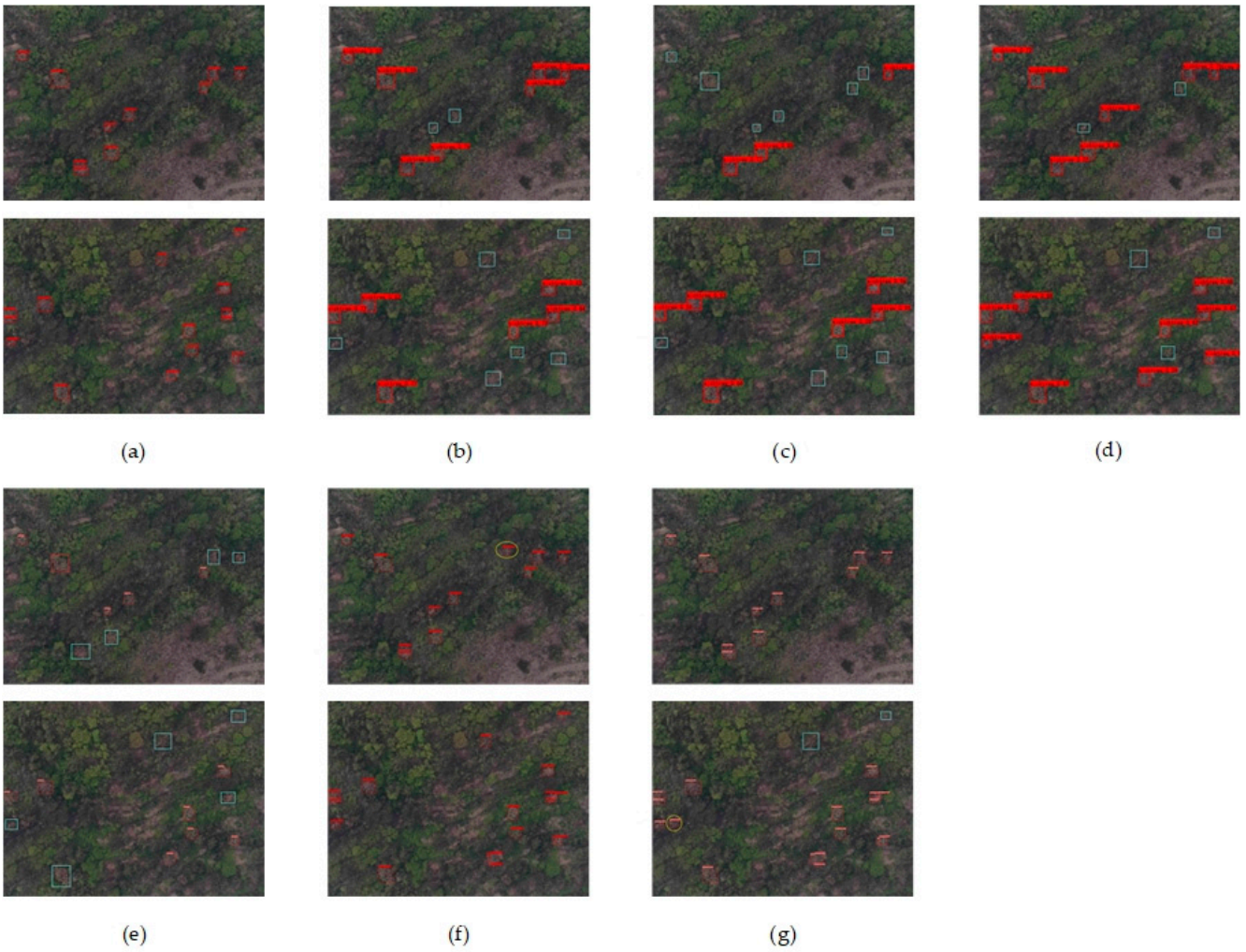


Figure 15. The recognition effect under the real complex background. Red boxes are detected infected trees, light blue boxes are undetected infected trees, and yellow circles are misidentified infected trees. (**a**) Improved YOLOv5l (**b**) Base YOLOv5l (**c**) Faster R-CNN (**d**) YOLOv4 (**e**) SSD300 (**f**) YOLOv7 (**g**) YOLOX.

Figure 15 shows that the improved YOLOv5l has made greater progress in false recognition and missed recognition and has achieved relatively good results for the difficult-to-recognize early and mid-stage disease trees; the base YOLOv5l has a high rate of missed recognition. The comparison algorithms YOLOv7 and YOLOX achieved the second-best result after the improved YOLOv5l, Faster R-CNN recognition was the worst, and SSD300 and YOLOV4 recognition was about the same. YOLOv7 achieves such high accuracy because it has an extended efficient layer aggregation network (E-ELAN), but its model size is much larger than that of the improved YOLOv5l. In contrast, YOLOX achieves the smallest model size; its unique anchor-free approach does not allow it to achieve higher accuracy and faster detection speed.

3.3. Discussion

By extracting and analyzing the reflectance values of multispectral images, we found that diseased trees differed significantly from healthy trees in four bands, red, near-infrared, red edge, and red edge 750 nm, which is consistent with the studies of Iordache et al. [21] and Kim et al. [22] in recent years. Some differences in the bands occur due to different data collection methods and environmental effects. It can also be found that there will be differences between multispectral data and hyperspectral data. The biggest difference is that multispectral images are often composed of several bands, whereas hyperspectral images are a whole range of bands [21,22], which can have an impact on the data analysis.

In disease tree identification, more and more scholars have used deep learning models [6–16], such as the YOLO series. Meanwhile, a single visible light and a single multispectral cannot meet the demand for high accuracy. Therefore, fusion recognition will have more and more applications. Still, the way of multispectral visible fusion recognition will inevitably affect the speed and size of the model [27], and we chose the input fusion that is optimal in terms of speed and accuracy this way.

The current work has a high accuracy in identifying completely yellowed diseased trees, but it is difficult to identify early diseased trees that are not visible to the naked eye or difficult to observe with the naked eye. Multispectral images can solve part of the problem, but they still cannot identify early diseased trees quickly and accurately. Therefore, how to identify diseased trees that do not exhibit PWD characteristics becomes a direction for our future work.

4. Conclusions

The timely identification is critical for the diagnosis and treatment of PWD-infected trees. However, single visual or multispectral imagery-based identification models may suffer from missing and false detection due to the effect of complex environments. By fusing visual or multispectral imagery, we propose a YOLOv5l-based improved detection method, which can make full use of the spatial features of visible images and spectral features of multispectral images. The reflectance difference in multispectral images between PWD-infected trees and healthy trees was analyzed. Four multi-spectral bands (red, near-infrared, red edge and red edge 750 nm) and two vegetation indexes (NDVI and NDRE) were selected as the distinguishable criteria. The YOLOv5l model was improved by using a combination of two attention mechanisms, CBAM and CA, BiFPN structure, and GhostNet and Transformer structure, to enhance the model's accuracy, reduce the number of parameters and improve the detection speed. The results show that the improved detection model has a parametric number of 46.69 MB, a detection speed of 0.064 s/sheet and mAP@0.5 reaches 0.987, which fully demonstrates the effectiveness of the improved YOLOv5l model. The proposed model is expected to provide rapid and accurate detection of PWD-infected trees in complex environments.

Author Contributions: Conceptualization and Methodology, P.L., F.S. and B.Q.; Software and validation, B.Q.; Investigation, B.Q. and X.H.; Resources, W.S., B.D. and S.M.; Data curation, B.Q.; Writing—original draft preparation, B.Q.; Writing—review and editing, P.L., F.S. and B.Q.; Visualization, B.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Shandong Agricultural Science and Technology Funds (Forestry Science and Technology Innovation) Project under Grant 2019LU003, Shandong Provincial Key Research and Development Program of China under Grant 2019GNC106106, Shandong Provincial Natural Science Foundation of China under Grant ZR2019MF026 and Shandong Science and Technology SMEs Innovation Capacity Enhancement Project under Grant 2022TSGC2437.

Data Availability Statement: Not applicable.

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

References

1. Zhao, B.G.; Futai, K.; Sutherland, J.R.; Takeuchi, Y. *Pine Wilt Disease*; Springer: Tokyo, Japan, 2008.
2. Ye, J.R.; Wu, S.Q. Research Progress of Pine Wilt Disease. *For. Pest Dis.* **2022**, *41*, 1–10.
3. Sousa, E.; Vale, F.; Abrantes, I. *Pine Wilt Disease in Europe: Biological Interactions and Integrated Management*; FNAPF: Lisbon, Portugal, 2015.
4. Abelleira, A.; Picoaga, A.; Mansilla, J.P.; Aguin, O. Detection of Bursaphelenchus xylophilus, Causal Agent of Pine Wilt Disease on Pinus Pinaster in Northwestern Spain. *Plant Dis.* **2011**, *95*, 776–776. [CrossRef] [PubMed]
5. Vollenweider, P.; Günthardt-Goerg, M.S. Diagnosis of Abiotic and Biotic Stress Factors Using the Visible Symptoms in Foliage. *Environ. Pollut.* **2006**, *140*, 562–571. [CrossRef]
6. Ren, D.; Peng, Y.; Sun, H.; Yu, M.; Yu, J.; Liu, Z. A Global Multi-Scale Channel Adaptation Network for Pine Wilt Disease Tree Detection on UAV Imagery by Circle Sampling. *Drones* **2022**, *6*, 353. [CrossRef]
7. Zhang, R.R.; Xia, L.; Chen, L.P.; Xie, C.C.; Chen, M.X.; Wang, W.J. Recognition of Wilt Wood Caused by Pine Wilt Nematode Based on U-Net Network and Unmanned Aerial Vehicle Imagery. *Trans. Chin. Soc. Agric. Eng.* **2020**, *36*, 61–68.
8. Liu, W.D.; Tian, H.B.; Xie, J.J.; Zhao, E.T.; Zhang, J.G. Identification Methods for Forest Pest Areas of UAV Aerial Photography Based on Fully Convolutional Networks. *Trans. Chin. Soc. Agric. Mach.* **2019**, *50*, 179–185.
9. Tao, H.; Li, C.; Zhao, D.; Deng, S.; Hu, H.; Xu, X.; Jing, W. Deep Learning-Based Dead Pine Tree Detection from Unmanned Aerial Vehicle Imagery. *Int. J. Remote Sens.* **2020**, *41*, 8238–8255. [CrossRef]
10. Liu, S.C.; Wang, Q.; Tang, Q.; Liu, L.; He, H.Y.; Lu, J.F.; Dai, X.Q. High-resolution Imagery Identification of Trees with Pinewood Nematode Disease Based on Multi-Feature Extraction and Deep Learning with Attention Mechanism. *J. For. Eng.* **2022**, *7*, 177–184.
11. Xu, X.L.; Tao, H.; Li, C.J.; Chen, C.; Guo, H.; Zhou, J.P. Detection and Location of Pine Wilt Disease Induced Dead Pine Trees Based on Faster R-CNN. *Trans. Chin. Soc. Agric. Mach.* **2020**, *51*, 228–236.
12. Huang, L.M.; Wang, Y.X.; Xu, Q.; Liu, Q.H. Recognition of Abnormally Discolored Trees Caused by Pine Wilt Disease Using YOLO Algorithm and UAV Imagery. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 197–203.
13. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
14. Sun, Y.; Zhou, Y.; Yuan, M.S.; Liu, W.P.; Luo, Y.Q.; Zong, S.X. UAV Real-Time Monitoring for Forest Pest Based on Deep Learning. *Trans. Chin. Soc. Agric. Eng.* **2018**, *34*, 74–81.
15. Zhou, Z.D.; Li, R.R.; Ben, Z.Q.; He, B.Z.; Shi, Y.Y.; Dai, W.X. Automatic Identification of *Bursaphelenchus xylophilus* from Remote Sensing Imagery Using Residual Network. *J. For. Eng.* **2022**, *7*, 185–191.
16. Hu, G.; Yin, C.; Wan, M.; Zhang, Y.; Fang, Y. Recognition of Diseased Pinus Trees in UAV Imagery Using Deep Learning and AdaBoost Classifier. *Biosyst. Eng.* **2020**, *194*, 138–151. [CrossRef]
17. Oide, A.H.; Nagasaka, Y.; Tanaka, K. Performance of Machine Learning Algorithms for Fetecting Pine Wilt Disease Infection Using Visible Color Imagery by UAV Remote Sensing. *Remote Sens. Appl. Soc. Environ.* **2022**, *28*, 100869.
18. Li, F.; Liu, Z.; Shen, W.; Wang, Y.; Wang, Y.; Ge, C.; Sun, F.; Lan, P. A Remote Sensing and Airborne Edge-computing Based Detection System for Pine Wilt Disease. *IEEE Access* **2021**, *9*, 66346–66360. [CrossRef]
19. Franke, J.; Menz, G. Multi-Temporal Wheat Disease Detection by Multi-Spectral Remote Sensing. *Precis. Agric.* **2007**, *8*, 161–172. [CrossRef]
20. Zong, S.X.; Bi, H.J. Monitoring Progress and Prospect of Pine Wilt Disease Based on UAV Remote Sensing. *For. Pest Dis.* **2022**, *41*, 45–51.
21. Iordache, M.D.; Mantas, V.; Baltazar, E.; Pauly, K.; Lewyckyj, N. A Machine Learning Approach to Detecting Pine Wilt Disease Using Airborne Spectral Imagery. *Remote Sens.* **2020**, *12*, 2280. [CrossRef]
22. Kim, S.R.; Lee, W.K.; Lim, C.H.; Kim, M.; Kafatos, M.C.; Lee, S.H.; Lee, S.S. Hyperspectral Analysis of Pine Wilt Disease to Determine an Optimal Detection Index. *Forests* **2018**, *9*, 115. [CrossRef]
23. Iordache, M.D.; Mantas, V.; Baltazar, E.; Lewyckyj, N.; Souverijns, N. Application of Random Forest Classification to Detect The Pine Wilt Disease From High Resolution Spectral Images. In Proceedings of the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 4489–4492.
24. Qin, J.; Wang, B.; Wu, Y.; Lu, Q.; Zhu, H. Identifying Pine Wood Nematode Disease Using UAV Imagery and Deep Learning Algorithms. *Remote Sens.* **2021**, *13*, 162. [CrossRef]
25. Yu, R.; Luo, Y.; Li, H.; Yang, L.; Huang, H.; Yu, L.; Ren, L. Three-Dimensional Convolutional Neural Network Model for Early Detection of Pine Wilt Disease Using UAV-Based Hyperspectral Imagery. *Remote Sens.* **2021**, *13*, 4065. [CrossRef]

26. Park, H.G.; Yun, J.P.; Kim, M.Y.; Jeong, S.H. Multichannel Object Detection for Detecting Suspected Trees with Pine Wilt Disease Using Multispectral Drone Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 8350–8358. [CrossRef]
27. Zhou, Y.; Liu, W.; Bi, H.; Chen, R.; Zong, S.; Luo, Y. A Detection Method for Individual Infected Pine Trees with Pine Wilt Disease Based on Deep Learning. *Forests* **2022**, *13*, 1880. [CrossRef]
28. Zhu, X.; Lyu, S.; Wang, X.; Zhao, Q. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-Captured Scenarios. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 2778–2788.
29. Han, K.; Wang, Y.; Xu, C.; Guo, J.; Xu, C.; Wu, E.; Tian, Q. GhostNets on Heterogeneous Devices Via Cheap Operations. *Int. J. Comput. Vis.* **2022**, *130*, 1050–1069. [CrossRef]
30. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
31. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.
32. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *Adv. Neural Inf. Process. Syst. arXiv* **2017**, arXiv:1706.03762.
33. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and Efficient Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.
34. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv* **2015**, arXiv:1511.06434.
35. Al Mansoori, S.; Kunhu, A.; Al Ahmad, H. Automatic Palm Trees Detection from Multispectral UAV Data Using Normalized Difference Vegetation Index and Circular Hough Transform. In Proceedings of the High-Performance Computing in Geoscience and Remote Sensing VIII, Berlin, Germany, 12–13 September 2018; Volume 10792, pp. 11–19.
36. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New-State-of-the-Art for Real-Time Object Detectors. *arXiv* **2022**, arXiv:2207.02696.
37. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding Yolo Series in 2021. *arXiv* **2021**, arXiv:2107.08430.

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.