PINE WILT DISEASE TREE RECOGNITION ON UAV IMAGES VIA SAMPLING THRESHOLD INTERVAL WEIGHTING METHOD AND DOUBLE-HEAD DETECTION

Lu Wang*,** Dong Ren*,** Xiaoran Tian*,** Yunjie Zhang*,** and Deao Hu*

Abstract

The detection method of pine wilt disease tree using UAV images is effective. When using the deep learning method to detect pine wilt disease, the problem that the number of diseased trees is so small compared with non-diseased trees that there is an imbalance between the positive and negative samples. This paper proposes a sampling threshold interval weighting method. The sampler is redesigned, and the IOU interval is divided into batches. The interval of difficult samples is weighted to improve the sampling rate of difficult samples and suppress the simple samples', so as to avoid sampling a large number of easily negative samples in random sampling, which leads to the imbalance of sample difficulty. In addition, a double-head detection is introduced to resolve the problem, that is, the spatial misalignment caused by the common feature parameters of classification and regression in the traditional R-CNN network. To verify the effectiveness of the proposed method, we conducted ablation experiments and comparative experiments on UAV images taken from the city of Yichang. The experimental results showed that the F1 accuracy of the improved Faster R-CNN network reached 91.46%, which could achieve accurate detection of pine wilt disease tree.

Key Words

Pine wilt disease, double-head, sampling strategy, UAV

1. Introduction

Pinewood nematode disease [1], also known as pine wilt, is a devastating pest. Once a healthy pine tree is infected, it will completely die within 1–2 months, and it will also infect nearby pine trees, which can cause the entire pine forest to die within 2–3 years. Therefore, the supervision of pine wood nematode disease is an important subject of my country's forest resource monitoring, and how to efficiently and accurately detect the location of diseased trees is the most important thing.

Domestic monitoring methods for pine wood nematode disease mainly include satellite remote sensing image monitoring, drone image monitoring, and artificial ground exploration. Most of them still use manual visual interpretation methods to identify diseased trees. For example, Lu *et al.* [2] interpreted the orthophotos images collected by UAV according to the crown colour changes of infected pine trees. Manual visual interpretation requires the interpreter to have relevant research experience, and the efficiency is low. Tao *et al.* [3] used the HSV threshold method to assist manual judgement, and Liu *et al.* [4] used the multi-template matching method to automatically identify diseased trees based on UAV images, which improved the recognition efficiency with an accuracy of 83.9%.

With the development of image technology, it has been applied to object recognition [5], [6], image segmentation [7], [8], autonomous driving [9], 3D Positioning [10], information technology [11], [12] and other fields. Researchers have begun to use deep learning methods to identify pine wood nematode disease. Zhou and Yang [13] used a one-stage target recognition algorithm, YOLO, to identify pine wood nematodes. Xu *et al.* [14] used the Faster R-CNN network to identify pine wood nematodes, and the F1 accuracy was 82.42%, which fully demonstrated the feasibility of deep learning in the identification of pine wood nematodes. However, there is still room for improvement in recognition accuracy.

In the experiment, several object detection algorithms are tried to train and test on the data set of pine wood disease, but the results were not good. The main reason is that the features are complex and changeable, the diseased trees are not evenly distributed, and the features of similar colors are easy to be misclassified.

* China Three Gorges University, Yichang 443002, China; e-mail: 276379755@qq.com; 252328196@qq.com; 202008540021049@ctgu.edu.cn; h_yuzhe@foxmail.com

** Hubei Engineering Technology Research Centre for Farmland Environmental Monitoring, China Three Gorges University, Yichang 443002, China; e-mail: rendong5227@163.com
Corresponding author: Dong Ren

Recommended by Anmin Zhu
(DOI: 10.2316/J.2024.206-0916)

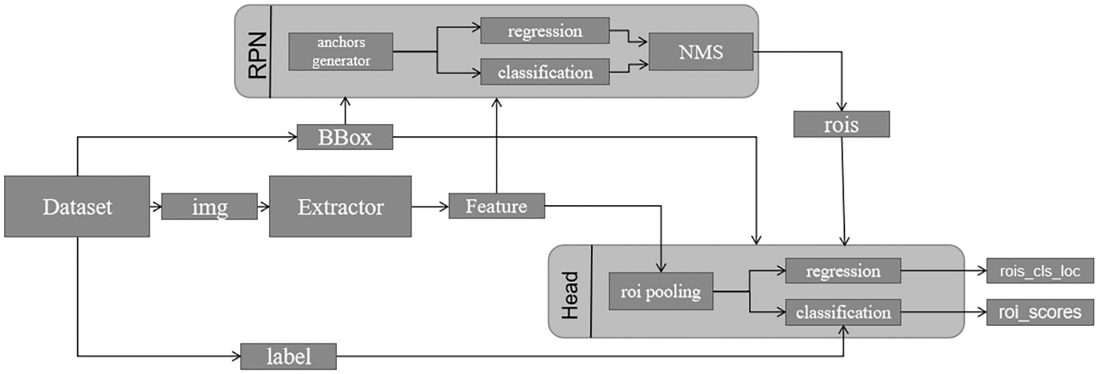

Figure 1. The architecture of Faster R-CNN.

In this paper, a two-stage deep learning object detection algorithm Faster R-CNN is adopted to detect pine wood nematodes. On the one hand, there is an imbalance between the positive and negative samples, we change the training strategy and propose a sampling threshold interval weighting method. We divide the IOU interval [15] into batches and weight the difficult sample intervals to increase the sampling rate of difficult samples, suppress the sampling rate of simple samples, and avoid sampling a large number of negative samples due to random sampling, making sample sampling difficult and easy to get out of balance.

On the other hand, we decouple the detection head on the basis of Faster R-CNN and separate the convolution and fully connected layers of the detection head, so that a single detection head can better focus on the supervision information of the corresponding task.

The remainder of this paper is organised as follows. We introduce the related work of Faster R-CNN, class imbalance, and sampling strategy in the next section. Following that, we give a detailed introduction to the sampling threshold interval weighting we proposed and the double-head [16] method in Section 3. In Section 4, we introduce the data set and design the ablation experiment. At last, we summarise our paper and present our future works in Section 5.

2. Related Work

In this part, the related knowledge will briefly be introduced, including the Faster R-CNN, the class imbalance, and sampling strategy in object detection.

2.1 Faster R-CNN

Faster R-CNN is a commonly used network architecture in the field of target detection. It combines feature extraction, candidate region generation, classification, and location refinement. It is a strong network comprehensive performance. The Faster R-CNN network structure is shown in Fig. 1.

The whole process of Faster R-CNN can be divided into three steps. (a) *Feature extraction*: The image passes through the backbone network and the image features are extracted. (b) *The region proposal frame generation:* A certain amount of region of interest (ROI) is found through the RPN network using the extracted features. (c) *Classification and regression*: Put ROIs and image features into the head, classify these ROIs, determine which category they belong to, and fine-tune these ROIs at the same time.

Faster RCNN realises object detection performance with higher accuracy through two-stage network and RPN. Compared with other one-stage networks, two-stage network is more accurate, especially for high-precision, multi-scale, and small object problems.

2.2 Class Imbalance

The problem of class imbalance [17] refers to the situation where the number of training examples in different class differs greatly. In this kind of problem, some classes will contain more examples, and the model will be brought to the main class in the sample during training, losing its robustness. The methods generally used to solve this problem can be divided into three categories.

1. *Threshold Shift*: Adjust the threshold of the classification according to the actual situation, and then increase the weight of a certain type of sample to achieve the purpose of solving the imbalance of the category.
2. *Sampling Strategy:* There are three methods of selective sampling, including under-sampling, over-sampling, and a combination of two sampling methods. In practise, a combination of three methods may be used to obtain a more robust model.

The over-sampling method does not introduce more data to the model, but repeats the proportion of the positive sample data. The under-sampling method discards part of the positive or negative sample data to achieve a balance of data. But both the under-sampling method and the over-sampling method can easily cause the model to have an over-fitting problem, that is, the model learns the data too thoroughly so that it also learns the characteristics of the noise data.

3. *Data Synthesis:* The data synthesis method is an improvement based on the random oversampling algorithm. The idea is to analyse minority samples and add new samples to the data set by artificially synthesising new samples.

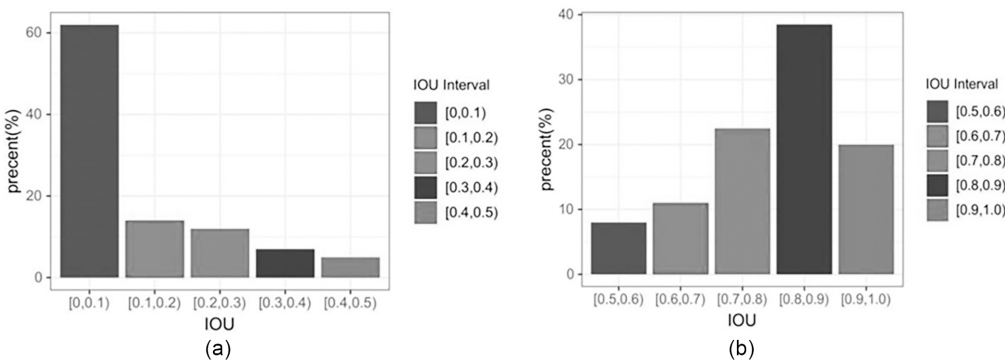

Figure 2. The sampling distribution of Faster R-CNN: (a) The distribution of negative samples; and (b) the distribution of positive samples.

The class imbalance problem is a widespread problem in object detection. If the class imbalance problem in training cannot be solved, it will have a great impact on network training.

2.3 Sampling Strategy in Object Detection

Because the background samples occupy a large part of the training process, the current object detection models will face the problem of extreme imbalance between foreground and background samples [18].

In order to solve this type of problem, the commonly used methods include two types. The first is the hard sampling methods: such as mini-batch bias method, OHEM [19], and IOU-Balanced sampling. These methods select a certain amount of positive samples from all samples and negative samples, only selected samples will be calculated Loss, generally tend to choose some difficult negative samples. A better training effect can be achieved by mining difficult negative examples.

The second is the soft sampling methods, such as Focal Loss [20] and GHM. These methods calculate Loss by selecting all samples but assign different weight values to different samples, such as Focal Loss.

These sampling methods have been proved to be effective in various data sets and experiments, but their hyper-parameter settings and the definition of positive and negative samples cannot be widely promoted in all data sets. However, hard sampling has a great impact on the training time of the network, while the soft sampling method needs to adjust a large number of hyperparameters.

In this paper, we proposed a method based on threshold interval weighting, which can improve the network without consuming additional resources, and increase the sampling rate of difficult samples, reduce the sampling rate of simple samples, thereby improving the sampling quality of network samples.

3. Method

In this section, we present a detailed description of sampling threshold interval weighting and the double-head. The sampling threshold interval weighting method is introduced in Section 3.1 and the double-head method is introduced in Section 3.2.

3.1 Sampling Threshold Interval Weighting Method

This study conducted experiments on 105.312 square kilometers of data. The number of diseased trees is relatively small and the distribution is not concentrated. The area of the target object is very small, and the background occupies most of the positions. During the training process, the number of positive examples is much smaller than the number of negative examples, causing category imbalance. This imbalance is not entirely caused by the data set, but more importantly, it is caused by the existing anchor-based target detection architecture.

According to the point of view in machine learning, when the data is extremely unbalanced, classification will lead to categories with more samples and over-fitting.

This kind of problem needs to be overcome in network training. The IOU threshold for positive and negative samples defined in Faster R-CNN is 0.5. If it is greater than 0.5, it is taken as a positive sample, and if it is less than it is taken as a negative sample. After determining the positive and negative attributes of each sample, random sampling is performed on the positive and negative samples through a random sampler. But random sampling does not distinguish the difficulty of the samples.

In network training, difficult samples contribute the most to the network, and simple negative samples are not helpful to the training of the network. Too many susceptible samples will dominate the network training, resulting in poor training effects. Random sampling in the negative sample set has a high probability of distribution in the negative sample set, which leads to poor network training effects.

The experiment analyses the sampling distribution of Faster R-CNN. As shown in Fig. 2, it can be seen that the random sampling method is used to sample, and there is no difference in the difficulty of sampling samples. The samples mostly gathered at both ends of the sampling interval are called simple samples. The single sample loss function of simple samples is small, but because the number of simple samples is too large, the cumulative loss function

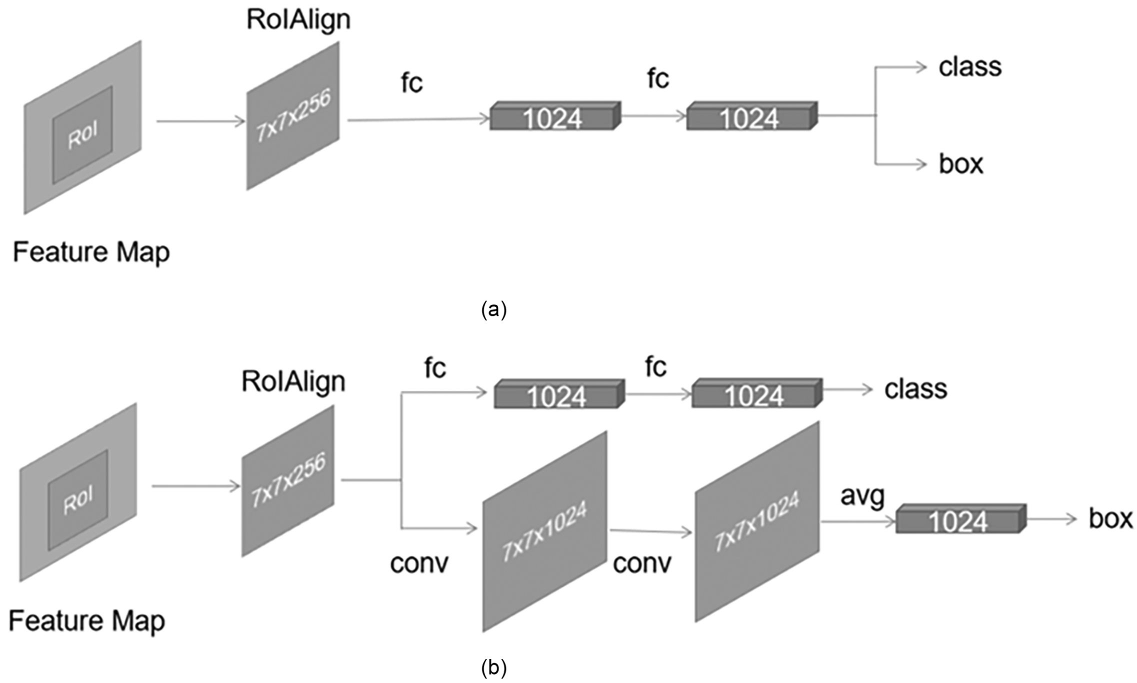

Figure 3. The comparison diagram of the traditional detection and the double-head detection: (a) The structure of R-CNN detector head; and (b) the structure of double-head R-CNN detector head.

will dominate the overall loss function. Because simple samples are easy to classify by the model, the update of the part parameter will not improve the judgement ability of the model, making the entire training process inefficient. Therefore, we need to improve the sampler to suppress the sampling rate of simple samples and to increase the sampling rate of difficult samples.

The experiment proposes a method of weighting the sampling threshold interval and divides the overall sample space into ten parts according to the threshold space of 0.1, and adjust the sampling ratio of interval threshold by σweighting coefficient. The weighting formula for the sampling interval is as follows:

$$S_k = \frac{N}{K} * \sigma + C \quad k \in [0, K) \tag{1}$$

$$\sigma = \begin{cases} \frac{n_{\max}}{n_h} & Hard \\ 1 - \frac{n_h}{n_{\max}} & Easy \end{cases} \tag{2}$$

S_k is the number of candidate samples in different batches, N is the number of samples, K is the number of divided intervals, in the experiment, we set $K = 10$, the entire threshold interval is evenly divided into K intervals, and C is the random sampling from the overall sample space when the overall number of samples is not enough for the N value. $\sigma$is a weighting coefficient, where $n_{\max}$ represents the number of samples in the entire sample space, and n_h represents the number of samples in the current threshold interval. The overall trend of the sample space threshold is U-shaped, and the number of samples in the entire sample space is constant. The number of samples in the difficult sample interval is relatively rare compared to the number of easy samples.

The experiment divides the K intervals into simple sample intervals and difficult sample intervals. The interval [0,0.1), [0.1,0.2), [0.8,0.9), [0.9,1) are the sample interval, and the rest are difficult samples. Using different weighting coefficients σin different sample intervals suppresses the sampling ratio of simple samples and increases the sampling ratio of difficult samples. After the pre-selected boxes are filtered by the sampling network, the remaining pre-selected boxes as sample boxes are sent to the network detection head to identify and posit.

3.2 Double-head Method

The structure of the detection head of the traditional R-CNN network is generally shown in Fig. 3(a), and the structure of the double-head detection network is shown in Fig. 3(b).

In fact, the classification and regression have different requirements for features. The classification requires features of translation and scale in-variance. And the learned features will not change to the change of the target's position. The localisation is a position-sensitive task. The location of the feature is different, and the learned feature for different scales is different. Because the fully connected layer links different parameters to different input elements, it is more suitable for classifying the overall target from the local features. The convolutional layer owned the weight sharing characteristics is used for all input elements. With the same convolution of the kernel convolution, the obtained features have stronger spatial correlation, and it is easier to distinguish the background and object pixels.

This research uses a double-head structure in the traditional R-CNN, separating the fully connected layer and the convolutional layer, classifying on the fully connected layer, and performing regression on the convolutional layer. It can get better results than classification regression on a single head. It keeps the original configuration on the classification branch, still uses two fully linked layers for output, but changes the regression branch to convolution output.

4. Experiments

In this section, we describe the data sets and the parameter setting and demonstrate the effectiveness of the proposed method through comparative experiments.

4.1 Experimental Setup

4.1.1 *Data Sets*

The 1:500 UAV images collected in this study are mainly distributed in Gaobazhou Town, Songmuping Town and other areas around Yichang city.

The data comes from UAV (camera model is a7r2 35 mm camera equipped with RGB three channels), with a total of 105.312 square kilometers of image data. It screens the severely epidemic area as the training area, cropping (1000×1000-pixel size) and marking the training area. A total of 3,064 images of training samples are collected in the training area, which are divided into the training set (2758 photos) and the verifying set according to the ratio of 9:1 collection (306 photos).

4.1.2 *Implementation Details*

The model training uses the stochastic gradient descent (SGD) optimiser to update the parameter $\Delta\theta_{t+1}$ of the round of training. $\Delta\theta_t$ is the parameter of the previous round, γ is the momentum parameter set to 0.9, and $\frac{\partial L(\theta)}{\partial \theta_t}$ is the gradient of the network. L is the learning rate set to 0.0025, as the initial 12 epoch training parameters.

In the neural network training, there are often problems, such as gradient explosion or gradient disappearance, that make the training unable to continue. In this study, a gradient clipping method is added to the training to keep the gradient in a relatively small state. If $||g|| \geq c$, it adjusts the gradient. C refers to the hyperparameter, g refers to the gradient, and $||g||$refers to the norm of the gradient. The gradient clipping guarantees the maximum norm of the gradient vector, even if the loss function of the model is irregular, the gradient clipping also helps the gradient descent to maintain a reasonable range.

$$\Delta\theta_{t+1} = \gamma \times \Delta\theta_t - 1 \times \frac{\partial L(\theta)}{\partial \theta_t} \tag{3}$$

$$g \leftarrow c \cdot g / \|g\| \tag{4}$$

4.1.3 *Evaluation Metrics*

In order to measure the recognition effect of the model, the model needs to be quantitatively evaluated. The evaluation method should use standard and recognised methods to ensure the validity of the experiment. Usually, the accuracy rate P (Precision), the recall rate R (Recall), and the average precision (Average precision) are used as evaluation indicators in target detection.

1. *Accuracy rate and recall rate*

The calculation of precision and recall is determined by the three values of true positive (TP), false positive (FP), and false negative (FN). TP is a positive sample, FP is a positive sample, and FN is a negative sample. The calculation formula of the accuracy rate and the recall rate is:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

The Precision represents the proportion of positive samples in the recognition sample, and the Recall represents the proportion of positive samples in all ground truth in a category.

2. *Average accuracy*

The average accuracy is an important indicator for evaluating the overall accuracy of the model. The calculation formula is:

$$AP = \sum_{k=1}^{N} P(k)\Delta r(k) \tag{7}$$

3. *F1 score*

In the identification of the sick tree, the Precision and the Recall need to be considered comprehensively. The F1 score can comprehensively evaluate the overall performance of the model. The calculation formula is:

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{8}$$

4.2 Experimental Results

At present, the target detection accuracy is still superior in the two-stage network, so the basic model used in the research is selected from the Cascade R-CNN [21] and the Faster R-CNN [22]. The research uses two basic networks for testing, and the best results are as follows.

Table 1
The Comparison Results with the Cascade R-CNN-r50 and Faster R-CNN-r50

Method	Precision%	Recall%	F1%
Cascade R-CNN-r50	79.8	90.8	84.94
Faster R-CNN-r50	82.7	90.6	86.46

Tests show that the accuracy of the Faster R-CNN network is higher than that of the Cascade R-CNN network, besides, the Cascade R-CNN network cascades three detection head parts, if using the dual detection head method, it will make the network very large and not suitable for method promotion. Above these reasons, the

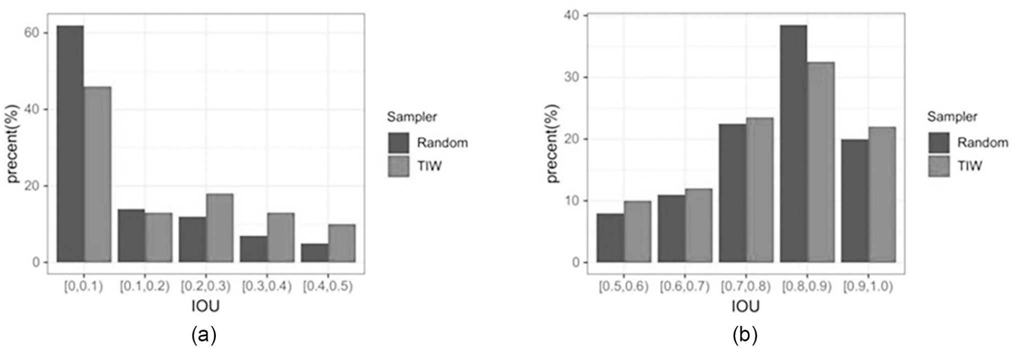

Figure 4. The comparison of the sampling threshold interval weighting method and the random sampling method: (a) The negative samples of the threshold interval weighting sampler and random sampler; and (b) the positive samples of the threshold interval weighting sampler and random sampler.

Table 2
The Comparison Results with the Faster R-CNN-r50-Random and the Faster R-CNN-r50-TIW

Method	Precision%	Recall%	F1%
Faster R-CNN-r50-Random	82.7	90.6	86.46
Faster R-CNN-r50-TIW	85.5	94.5	89.78

Faster R-CNN network is selected for improvement in this experiment.

Through the experimental results, serious misdetection problems are in both Cascade R-CNN and Faster R-CNN, and there will be errors in judging objects with similar pixel colors. It often appears that there are multiple test results for a target, and non-maximum suppression (NMS) cannot be used to remove these test results. It is summarised by two reasons. The first is that the model's own ability is insufficient and the regression classification ability is poor, which leads to a large number of false detections. The second is due to the limitation of the number of samples, the model will detect results with higher confidence in some ambiguities, and it cannot handle these false detection results through the threshold.

4.2.1 *Sampling Threshold Interval Weighting Method*

The sampling threshold interval weighting method is tested on our datasets and is compared with the random sampling method. It is found that by weighting the IOU threshold interval, the sampling of the sampler is selected for the sampling interval, which effectively alleviates the imbalance of the sampling ratio caused by the unbalanced distribution of samples.

It can be seen from Fig. 4 that by weighting the threshold interval, the sampling ratio of the sampler for difficult samples is increased, and the sampling of a large number of simple samples is controlled, making the effect of the entire training process more efficient.

Through experiments, as shown in Table 2, after weighting the threshold interval, the verification accuracy has been improved by 2.8%, and the F1 accuracy has been improved by 3.14%.

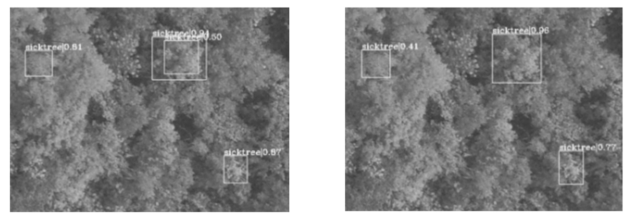

Figure 5. Faster R-CNN experiment comparison, the left is without the double-head structure, and the right is introduced the double-head structure.

Table 3
The Comparison Results with the Faster R-CNN and the Faster R-CNN with Double-head

Method	Precision%	Recall%	F1%
Faster R-CNN-r50	82.7	90.6	86.46
DH-Faster R-CNN-r50	83.9	95.2	89.19

4.2.2 *Double-head Method*

The double-head structure is used to test our datasets. Fig. 5 and Table 3 show the results of the comparative experiments. The classification task and the regression task are decoupled into two branches in the double-head structure, which can make different the branch focuses on the corresponding task, adapting the structure and function to obtain better results.

The experimental results show that the network performance is enhanced by adding the double-head structure, and the redundant frames that the Faster R-CNN network cannot use NMS removal are removed. Besides the classification score of the target sample, the classification score of the uncertain sample and the overall classification accuracy of the network have been improved. Comparing the experimental results without and with double-head in the Faster R-CNN on the pine wood nematode data sets, the F1 value increased by 2.73%.

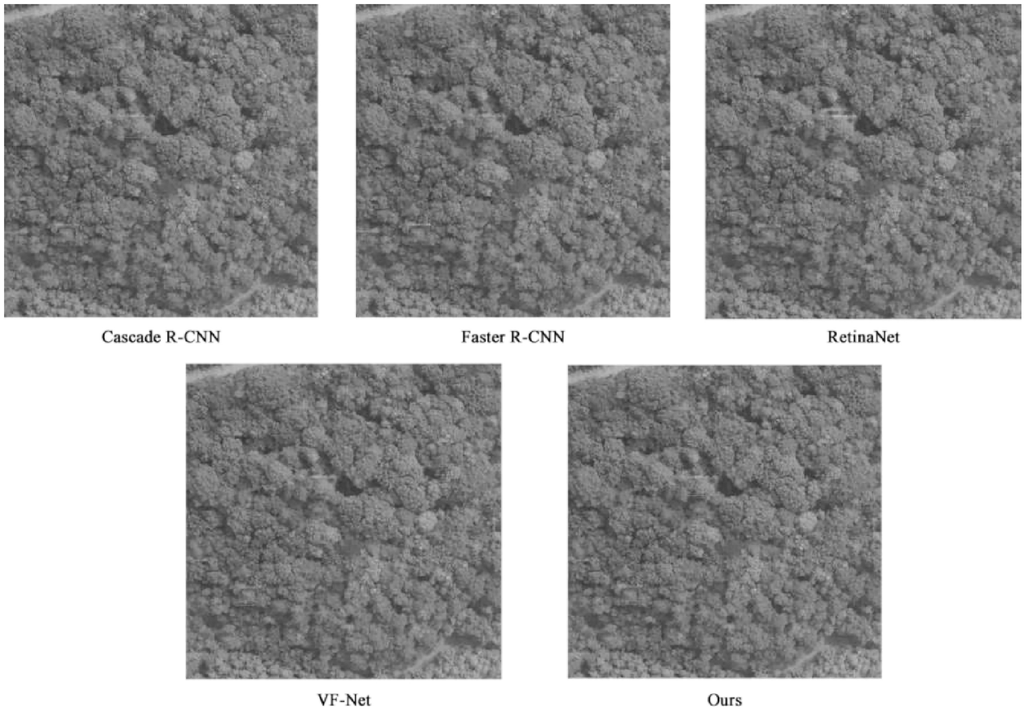

Figure 6. Comparison between our method and other object detection methods.

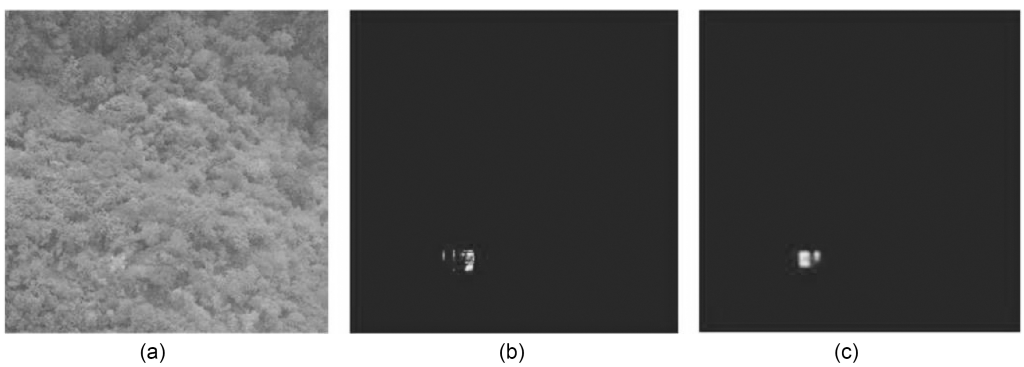

Figure 7. (a) A original map; and (b) and (c) The features in different layers.

However, due to the large range of the datasets, the complex terrain, and the existence of a large number of objects with similar pixel colors, it is easy to cause errors in the judgement of the network category. Besides there are still a large number of negative samples, which will affect the network training.

Based on the above situations, an improved sampler is added for fusion experiments. Experiments show that adding the double-head structure can improve the classification score of the target sample, remove redundant frames that are difficult to remove with NMS, and improve the accuracy of network classification.

In summary, in order to prove the effectiveness of the improved model, the experiment uses other target detection models for experimental comparison, and the model parameters and data parameter settings are the same.

From Table 4 and Fig. 6, it can be seen that Faster R-CNN, Cascade R-CNN, and RetinaNet all have certain recognition accuracy, but there are different problems. Cascade R-CNN has a misclassification situation that identifies bare soil with similar colours as diseased trees. RetinaNet has the problem of redundant frames, and Faster R-CNN also has recognition errors.

Table 4
The Results under Different Algorithms

Method	Precision%	Recall%	F1%
Faster R-CNN-r50	82.7	90.6	86.46
Cascade R-CNN-r50	79.8	90.8	84.94
RetinaNet-r50	81.7	95.0	87.85
VF-Net-r50	80.0	92.7	85.88
Ours	**87.0**	**96.4**	**91.46**

It can be seen that the regression frame of VF-Net cannot fully contain the entire disease tree, and there is a problem that the regression is not accurate enough. Experiments have proved that by introducing the double-head structure to decouple the classification and regression tasks, the two types of tasks can achieve better results and improve the recognition ability of the model.

4.2.3 *Noise Robustness Verification Experiment*

To evaluate the robustness of the proposed model against noise, we have added an experiment. We make the feature maps visualisation. In the visual feature maps, the colour is more red, the more characteristic it is.

Figure 7 is a original map in the test data set and its features in different layers. From the visual maps, the goal boundaries are clear. It proves to be a valid model to evaluate the robustness against noise.

5. Conclusion

In order to solve the problem of low recognition accuracy of the pine wood nematode data set under the influence of complex features, especially for the reason that the number of diseased trees is so small compared with non-diseased trees that there is an imbalance between the positive and negative samples, a sampling threshold interval weighting method is proposed. The sampler is redesigned, and the IOU interval is divided into batches. It can optimise the training effect of the network. Besides a double-headed part of the R-CNN network is added to improve the classification and regression performance of the network.

After the ablation experiments and comparative experiments on UAV images taken from the city of Yichang, the results show that the F1 accuracy of the improved Faster R-CNN network reached 91.45%, which could achieve accurate detection of pine wilt disease and have greatly improved the accuracy of the current mainstream target detection algorithm.

Acknowledgement

This work was supported by the Science and Technology Research and Development Project of Yichang City in Hubei Province of China (Grant no. A22-1-057).

References

[1] M.M. Mota and P.R. Vieira, Pine wilt disease: A worldwide threat to forest ecosystems, *Nematology, 2*(11), 2009, 315–316.

[2] X.J. Lu, J. Wang, and W.G. Yu, Studying on monitoring forest pests and diseases by unmanned aerial vehicle, *Hubei Forestry Science and Technology*, *45*(4), 2016, 30–33. (in Chinese).

[3] H. Tao, C.J. Li, and C.C. Xie, Recognition of red-attack pine trees from UAV imagery based on the HSV threshold method, *Journal of Nanjing Forestry University (Natural Sciences), 43*(3), 2019, 99–106. (in Chinese)

[4] X.L. Liu, D.X. Cheng, and T. Li, Preliminary study on automatic monitoring trees infected by pine wood nematode with high resolution images from unmanned aerial vehicle, *Forest Pest and Disease*, *37*(5), 2018, 16–21. (in Chinese)

[5] J. Wan, H. Xi, J. Zhou, Z.H. Lai, W. Pedrycz, X. Wang, and H. Sun, Robust and precise facial landmark detection by self-calibrated pose attention network, IEEE Transactions on Cybernetics*, 53*(6), 2021, 3546–3560. doi: 10.1109/TCYB.2021.3131569.

[6] J. Wan, Z.H. Lai, J. Li, J. Zhou, and C. Gao, Robust facial landmark detection by multi order multi constraint deep networks, *IEEE Transactions on Neural Networks and Learning Systems, 33*(5), 2022, 2181–2194.

[7] S. Ren, X. Liu, H. Liu, and L. Wang, Cultivated land segmentation of remote sensing image based on PSPNET of attention mechanism, *International Journal of Robotics & Automation, 37*(1), 2022, 11–19.

[8] M. Liu, D. Ren, H. Sun, S.X. Yang, and P. Shao, Orchard areas segmentation in remote sensing images *via* class feature aggregate discriminator, *IEEE Geoscience and Remote Sensing Letters*, *19*, 2022, 1–5.

[9] H. Sun, Y. Zhang, P. Chen, Z.P. Dan, Z. Dan, S. Sun, J. Wan, and W. Li, Scale-free heterogeneous cycleGAN for defogging from a single image for autonomous driving in fog, *Neural Computing and Applications*, *35*, 2023, 3737–3751.

[10] F. Wen, J. Shi, and G. Gui, 3D positioning method for anonymous UAV based on bistatic polarized MIMO radar, *IEEE Internet of Things Journal*, *10*(1), 2023, 815–827.

[11] F. Wen, G. Gui, and H. Gacanin, Compressive sampling framework for 2D-DOA and polarization estimation in mmWave polarized massive MIMO systems, *IEEE Transactions on Wireless Communications,* 22(5), 2023, 3071–3083, doi: 10.1109/TWC.2022.3215965.

[12] M. Yu, Y.Y. Sheng, H. Sun, Y. Zheng, H. Li, and W. Yang, Research on operation and maintenance route planning of offshore wind farm based on multiagent, *Wireless Communications and Mobile Computing*, 2022, 2022, 3278239. doi: https://doi.org/10.1155/2022/3278239.

[13] Z. Zhou and X.T. Yang, Pine wilt disease detection in UAV-captured images, *International Journal of Robotic & Automation, 37*(1), 2022, 37–43.

[14] X.L. Xu, H. Tao, and C.J. Li, Detection and location of pine wilt disease induced dead pine trees based on faster R-CNN, *Transaction of the Chinese Society for Agricultural Machinery, 7*(51), 2020, 228–236. (in Chinese)

[15] S. Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *IEEE Transactions on Pattern Analysis & Machine Intelligence, 39*(6), 2017, 1137–1149.

[16] G. Song, Y. Liu, and X. Wang, Revisiting the sibling head in object detector, 2020, *arXiv:2003.07540*.

[17] K. Oksuz, B.C. Cam, S. Kalkan, and E. Akbas, Imbalance problems in object detection: A review, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(10), 2021, 3388–3415.

[18] J. Chen, Q. Wu, D. Liu, and T. Xu, Foreground-background imbalance problem in deep object detectors: A review, 2020, *arXiv:2006.09238*.

[19] Shrivastava, A. Gupta, and R. Girshick, Training region-based object detectors with online hard example mining, Proc. *Conf. on Computer Vision & Pattern Recognition IEEE Computer Society,* Las Vegas, NV, 2016, 761–769.

[20] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, Focal loss for dense object detection, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 42(2), 2020, 318–327.

[21] Li, X. Yang, and C. Zhang, Rethinking classification and localization for cascade R-CNN, 2019, *arXiv:1907.11914*.

[22] M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, and A. Zisserman, The pascal visual object classes (VOC) challenge, *International Journal of Computer Vision*, *88*(2), 2010, 303–338.

Biographies

Lu Wang is currently pursuing the postgraduate degree with the College of Electrical Engineering & New Energy, China Three Gorges University.

Dong Ren received the Ph.D. degree from Jilin University. He is a Professor with the College of Computer and Information Technology, China Three Gorges University, Yichang, China. His research interests include artificial intelligence, pattern recognition, and 3S technology.

Xiaoran Tian is currently pursuing the postgraduate degree with the College of Computer and Information Technology, China Three Gorges University.

Yunjie Zhang is currently pursuing the postgraduate degree with the College of Electrical Engineering & New Energy, China Three Gorges University.

Deao Hu is currently pursuing the postgraduate degree with the College of Electrical Engineering & New Energy, China Three Gorges University.