forests

MDPI

Article

A Small Target Tea Leaf Disease Detection Model Combined with Transfer Learning

Xianze Yao [1], Haifeng Lin [1,*], Di Bai [2,*] and Hongping Zhou [3]

[1] College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China
[2] College of Information Management, Nanjing Agricultural University, Nanjing 210037, China
[3] College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China
* Correspondence: haifeng.lin@njfu.edu.cn (H.L.); baidi000@njau.edu.cn (D.B.); Tel.: +86-25-8542-7827 (H.L.)

Abstract: Tea cultivation holds significant economic value, yet the leaves of tea plants are frequently susceptible to various pest and disease infestations. Consequently, there is a critical need for research focused on precisely and efficiently detecting these threats to tea crops. The investigation of a model capable of effectively identifying pests and diseases in tea plants is often hindered by challenges, such as limited datasets of pest and disease samples and the small size of detection targets. To address these issues, this study has chosen TLB, a common pest and disease in tea plants, as the primary research subject. The approach involves the application of transfer learning in conjunction with data augmentation as a fundamental methodology. This technique entails transferring knowledge acquired from a comprehensive source data domain to the model, aiming to mitigate the constraints of limited sample sizes. Additionally, to tackle the challenge of detecting small targets, this study incorporates the decoupling detection head TSCODE and integrates the Triplet Attention mechanism into the E-ELAN structure within the backbone to enhance the model's focus on the TLB's small targets and optimize detection accuracy. Furthermore, the model's loss function is optimized based on the Wasserstein distance measure to mitigate issues related to sensitivity in localizing small targets. Experimental results demonstrate that, in comparison to the conventional YOLOv7 tiny model, the proposed model exhibits superior performance on the TLB small sample dataset, with precision increasing by 6.5% to 92.2%, recall by 4.5% to 86.6%, and average precision by 5.8% to 91.5%. This research offers an effective solution for identifying tea pests and diseases, presenting a novel approach to developing a model for detecting such threats in tea cultivation.

Keywords: tealeaf disease detection; transfer learning; TSCODE; Triplet Attention; Wasserstein distance

Citation: Yao, X.; Lin, H.; Bai, D.; Zhou, H. A Small Target Tea Leaf Disease Detection Model Combined with Transfer Learning. *Forests* **2024**, *15*, 591. https://doi.org/10.3390/f15040591

Academic Editors: Viacheslav I. Kharuk and Nikolay Strigul

Received: 20 January 2024
Revised: 1 March 2024
Accepted: 22 March 2024
Published: 25 March 2024

1. Introduction

Tea is an important cash crop in traditional agriculture. The market demand for tea is vast. Global tea production is over USD 1.7 billion annually, while the world tea trade is valued at about USD 9.5 billion [1]. Protecting the growth safety of tea trees and tea quality is essential to promote the development of the tea industry. Tea pests and diseases have a significant impact on tea production, which not only leads to low tea yields but also reduces the quality of tea, thus causing huge economic losses to tea farmers and tea producers [2]. Traditional detection methods include expert identification, molecular biology, and spectroscopy. Traditional manual identification by experts suffers from the problems of long time, high subjectivity, and low accuracy, and inviting experts to conduct detection visits in the field is a costly and labor-intensive task [3]. Molecular biology and spectroscopy tests are more accurate, but their learning and instrumentation costs are high [4]. Therefore, research on the accurate and efficient detection of tea pests and diseases is significant.

With the development of updated iterations of computer vision technology, research on the identification and detection of tea pests and diseases is also being carried out. Image

processing and machine learning methods are beginning to be applied to the detection of tea pests and diseases. These methods are mainly classified by segmentation of the disease spots and extraction of features, and more results have been achieved. In 2018, Md. Selim Hossain [5] proposed a novel support vector machine for a disease recognition system. He and his co-researchers extracted eleven identifying features of two common tea diseases in Bangladesh and uploaded them to the SVM (support vector machine) database for detection and identification. The algorithm not only has faster processing time but also maintains a high level of accuracy. In 2019, Yunyun Sun [6] proposed a new algorithm that combines SLIC (Simple Linear Iterative Clustering) with the SVM, and the results show that this method is very effective in improving the complex background to extract the leaf disease feature maps of tea plants. It can better identify the diseases and pests of tea. In 2020, Xiuguo Zou [7] proposed a spectral reflectance-based method for tea disease and pest identification, which includes a feature selector based on a decision tree and a tea disease identifier based on a random forest. The results of the experiments show that the recognition results of this method are improved in both precision and recall. In 2022, S. Prabu detects tea diseases with the help of the SVM approach in traditional machine learning. The watershed transform algorithm is used for the clear segmentation of color-transformed images, and gradient eigenvalues of tea images are used in multi-class support vector machine classifiers to classify tea diseases. The performance is evaluated, and the model is proven effective and performs better [8]. Considering that the life cycle of plant diseases is closely related to environmental conditions, Zhiyan Liu proposed in the same year to predict tea crop diseases utilizing multivariate linear regression with the help of environmental conditions in the crop field. From the implementation results, the accuracy of disease prediction can be as high as 91% [9].

In machine learning methods for plant disease identification and detection, manual feature extraction is required for diseased leaves. The number of diseased tea leaves in samples is often high, and they share similarities in color, texture, and other characteristics. The accuracy of manual feature extraction greatly influences the performance of plant disease leaf identification and detection in traditional machine learning methods. The process of feature extraction is not only time consuming but also subjective [10]. In recent years, deep learning methods, such as convolutional neural networks, have made rapid development. They are more direct, faster, and free from the constraints of manual feature extraction, which significantly improves the model's learning ability and identification accuracy. In 2019, Gensheng Hu [11] proposed an improved deep convolutional neural network (CNN)-based method for tea disease recognition. A multi-scale feature extraction module was added to the improved deep CNN of the CIFAR10 quick model to improve the ability to automatically extract features from images of different tea diseases. In 2020, Shyamtanu Bhowmik [12] used a convolutional neural network (CNN) model for detecting and recognizing black rot and tea rust diseases of tea trees to achieve high accuracy while maintaining minimal computational complexity and minimal resources. In the same year, Shenghung Lee [13] trained a faster region-based convolutional neural network (Faster R-CNN) to detect the location of disease spots on leaves and identify the cause of the spots. Relatively high overall accuracy was obtained for the identification of seven classes of tea diseases and pests. In 2021, Gen Sheng Hu [14] used the Faster R-CNN framework to detect TLB leaf blades in order to improve the detection performance of fuzzy, occluded, and small pieces of diseased leaves. The detected TLB leaves were inputted into a trained VGG16 network to achieve severity grading and facilitate disease severity analysis. In 2022, He Li [15] proposed a framework for recognizing pest and disease symptoms in tea based on Mask R-CNN, wavelet transform, and F-RNet. The Mask R-CNN model was used to segment disease and insect spots from tea leaves, and then a two-dimensional discrete wavelet transform was used to enhance the features of disease and insect spot images. Finally, the feature images were input into F-RNet for recognition. The experimental results show that the framework can accurately segment and recognize tea leaves' pest and disease symptoms.

However, regarding pest and disease detection in tea leaves, all the studies mentioned above have focused more on identifying multiple pests and diseases. This approach allows for slightly larger training data, as deep learning models often require a large number of data samples to avoid overfitting in target detection methods. However, the use of this method can also lead to unsatisfactory accuracy of detection due to the differences between the number of samples of multiple pests and diseases and their vastly different target characteristics. Moreover, in reality, the occurrence time of tea leaf pests and diseases is limited by climate temperature and geographic location, so the types of pests and diseases that often occur in a certain region and a certain period are limited. However, datasets of tea tree leaf pests and diseases for a specific region and period are very demanding, and few high-quality datasets can meet the requirements because of the effects of shooting angles, lighting conditions, backgrounds, and other factors. Such a situation makes it difficult to obtain a large number of high-quality samples. Tea diseases vary in shape and leaf susceptibility to infection, but their common characteristic is that they are relatively small targets, which usually cause poor detection. There are several main reasons for this. First, existing datasets pay less attention to small targets; second, small targets are more prone to aggregation phenomenon; and third, there are fewer features available for small targets. The diseased portion of tea leaves has another troublesome problem on top of meeting the definition of small targets. That is, the distribution of diseased parts of tea leaves and healthy tea leaves is denser, and it is easy to obscure each other and difficult to distinguish. Due to this, the difficulty of accurate detection is further increased. For the task of target detection in the specific field of tea leaf diseases and pests, how to effectively utilize the small sample dataset and solve the small target detection problem is an urgent research problem.

Tea leaf blight (TLB) stands out as a prevalent affliction among tea plants. It arises from the presence of Colletotrichum Camelliae Massee, a type of imperfect fungi. The onset of tea leaf blight initially manifests through yellow-brown and water-soaked lesions on the leaves. As the condition progresses, these lesions transform into irregular patches characterized by light brown and gray hues. If left unchecked, severe cases can result in leaf loss and the apparent demise of new shoots, thereby diminishing the overall vitality of the plant. We have conducted an in-depth study on tea tree leaf blight as a disease and found that it is a small sample problem in terms of the dataset. This is due to the fact that tea leaf blight occurs less frequently in a given region and period, and the available data on the disease are relatively small. Furthermore, like tea diseases in general, TLB tends to be characterized by small detection targets and a tendency to overlap and accumulate. In this paper, based on the above considerations, we have selected a single species of TLB for our study and proposed a small target detection model for TLB combined with transfer learning. We used data augmentation and transfer learning to address the small sample problem of TLB. Our pre-trained models on the large-scale source data domain are transferred to the TLB dataset, followed by further model optimization for small targets. Since the overlap between diseased and healthy tea leaves is relatively high during the detection process and some of the TLB targets to be recognized are too small, we want to solve these problems to reduce the occurrence of wrong and missed detections. We use a decoupled detector TSCODE in the detection head part, add a Triplet Attention mechanism to the E-ELAN (extended efficient layer aggregation network) structure, and also introduce a small target detection evaluation method based on Wasserstein distance. All of them effectively improved the model's ability to recognize small targets of tea diseases. The experimental results on the TLB small sample dataset show that our model can better solve the small sample and small target detection problems with higher accuracy and robustness relative to the traditional YOLOv7 tiny model.

In the following structure of the paper, we first give a comprehensive introduction to our dataset in Section 2. After that, the three sections based on transfer learning, model construction, and model evaluation methods are elaborated. In Section 3, there are four pieces of content. Firstly, we explain the environment and some parameters for model

training. We select the source domain for transfer learning and show the result comparison, followed by a series of comparative experiments to compare the transfer learning and the impact of several modules on the model. Finally, we show some comparative results for the experimental results. In Section 4, based on the results of this paper, we discuss and think about the research ideas in detail. In Section 5, we briefly summarize the whole research work.

2. Materials and Methods

2.1. Dataset

2.1.1. Data Acquisition and Annotation

The TLB images needed for the datasets in this paper were all taken independently over six months. In order to make the images taken closer to the real use case and avoid bias or error when training the model, we have taken the following measures. First, when shooting, considering the situation that TLBs will burst locally in reality, some sample pictures with dense and messy TLB distribution are shown in Figure 1b. Second, some pictures were intentionally taken when the aggregation was poor to better simulate the presence of low-quality photos in the dataset in the real situation, as shown in Figure 1c, and also to further increase the training difficulty so that the detection model obtained in the end has a more generalization ability. Third, in addition to the case of dense tea tree leaves, pictures with a background of dead leaves were also taken, as shown in Figure 1d. This not only highlights the foreground more but also simulates the case of taking pictures from different angles in a formal situation. Third, all photos were taken at a distance of 20–30 cm from the target, which is close to the realistic distance and can effectively limit the pixel size of the diseased tea tree leaf portion to the definition of a small target, which is convenient for the subsequent study. In the end, a total of 182 images with TLB were taken, with pixels of 1322 × 992 or 1323 × 992. Another difficulty with datasets with small targets is the accuracy of the labeling, which is even more difficult in a situation as dense and easily shaded as a tea bush. The labeling tasks in this paper were performed manually with the assistance of plant disease experts.

(**a**) (**b**)

Figure 1. *Cont.*

(**c**) (**d**)

Figure 1. Some representative samples of our dataset where the red boxes represent the labeling of tea leaf blight: (**a**) general TLB sample; (**b**) accumulated TLB sample; (**c**) TLB sample out of focus; (**d**) TLB sample with dead leaves background.

2.1.2. Data Augmentation

The number of samples in the dataset remained small after we simulated as many samples as possible obtained in different situations under the established conditions. Such small samples are more likely to lead to overfitting problems in subsequent training. For this reason, we extend the number of samples by data augmentation, i.e., using a specific method to generate new samples with the same distribution as the original samples, which is a relatively direct and effective method [16]. In this study, we randomly adjusted the sample batches' hue, saturation, and luminance before dividing the dataset. Moreover, we performed a certain amount of scaling and flipping operations in order to expand the number of samples(Figure 2). On this basis, we also appropriately use some panning and image blending techniques, but considering that the number of annotations in the TLB part of the original sample is limited and very small, the probability of calling these two methods is set to be relatively small because there is a situation that the annotations will be covered by the panning and image blending. The formula for a simple generation is shown in Formula (1), where (x_i, y_i) represents the original sample and f represents a generation function. The above changes are all within a certain range to ensure that the actual distribution of the transformed image and the original image do not differ too much to not affect the final generalization performance of the model.

$$D\prime_T = (f(x_i), y_i) \quad (1)$$

(**a**) (**b**)

Figure 2. *Cont.*

(**c**) (**d**)

Figure 2. Some representative samples of our dataset: (**a**) TLB target-intensive sample; (**b**) TLB sample accumulation; (**c**) TLB sample out of focus; (**d**) TLB sample with dead leaves background.

For a neural network model, a sample after one simple change is a new sample, which can be very effective in improving the generalized detection ability of the model [17]. In addition to the basic offline enhancement, in order to further expand the information content of the samples, we also use the mosaic method to enhance them online during the training process. The process of the method is shown in Figure 3. At each training, we randomly select batch-size pictures from the dataset. First, enhance them with certain simple operations and then randomly arrange and splice them to form a new picture. Repeat the batch-size times and then feed the newly generated batch-size pictures to the model. Such an approach enriches the dataset while indirectly increasing the number of small targets and the amount of information. Mixing four images with different semantic and localization information allows the model to detect targets beyond the regular context, which also increases the robustness of the model to some extent.

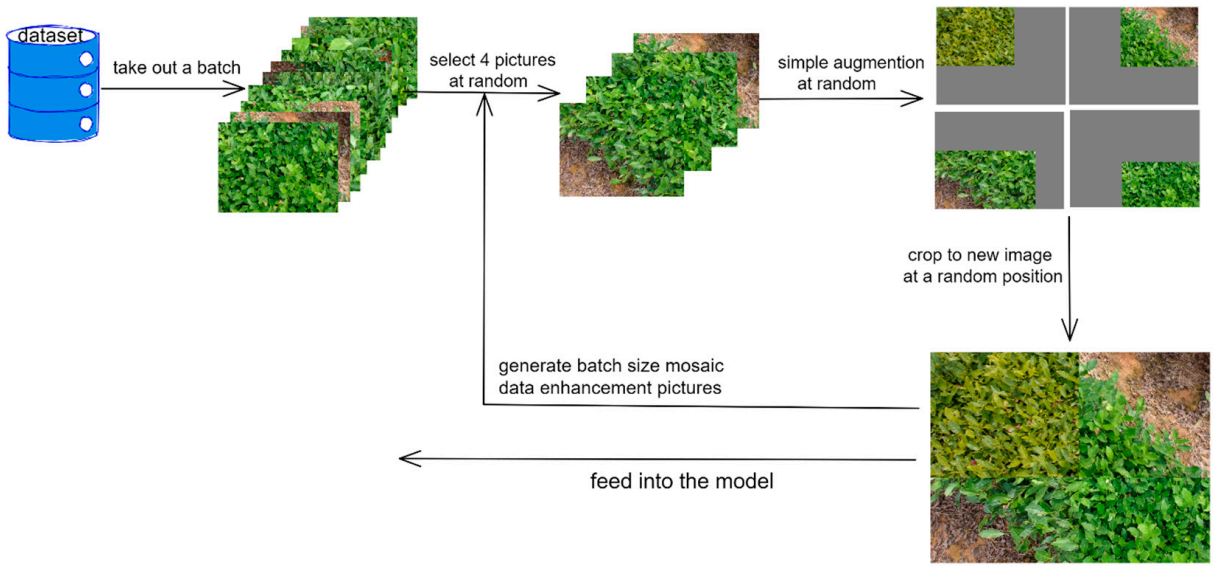


Figure 3. The process of mosaic augmentation.

2.2. Transfer Learning

Transfer learning is a machine learning approach that aims to solve the problem of transferring knowledge from a source domain task to improve the performance of the target task when the amount of data in the target task is insufficient [18]. When solving deep learning tasks, a large amount of labeled data is very expensive and time consuming, and the requirement of high-quality and large sample size datasets often makes the results

of traditional models on small samples unsatisfactory. This is where the advantages of transfer learning come into play. It has been widely used in the field of artificial intelligence and machine learning. For example, in the field of short-term load forecasting in smart grids, Dabeeruddin Syed has noted that there may not be a sufficient amount of historical data available to support research efforts at newly installed distribution nodes [19]. They attempted to introduce the concept of transfer learning. Experimental data indicate that transfer learning can lead to more accurate predictions in low data availability. In our paper, considering the difficulty of collecting TLB single-class datasets in real life, we introduce transfer learning as a method to optimize the TLB detection problem under small sample conditions.

2.2.1. Model-Based Transfer Learning

For the small sample problem, just using data augmentation may not be able to obtain a model performance that has a better performance. The effects of the introduction of transfer learning in disease recognition in a variety of crops have been shown to be significantly beneficial [20,21], so we considered the introduction of transfer learning as a method to further address the small samples in our study on TLB. Transfer learning can be basically categorized into four basic approaches: sample-based transfer, model-based transfer, feature-based transfer, and relationship-based transfer [22]. The more common method of model-based transfer learning is used in this study. That is, we transfer the parameters of the model trained on the source domain to the model on our target domain for new training, i.e., the pre-train and fine-tune paradigm. Note that the model architecture used for pre-training and post-transfer training is the same. In this way, our target domain model will learn the target domain based on some knowledge about the source domain until a high-performance decision function for the target task can be obtained, as shown in Figure 4. Specifically, in this study, after first training on the source dataset and obtaining a training result weight, we then use this weight as our pre-training weight, i.e., a priori knowledge to train on the TLB dataset and finally obtain a suitable result weight for the TLB, i.e., we obtain a detection model suitable for TLB. The pre-training weights can be said to summarize the model's knowledge for the upstream task. After the transfer, the trained model for TLB will try to identify the disease targets in the previous dataset based on the a priori knowledge, and on the basis of which it will re-learn the knowledge related to the detection of TLB and focus on the downstream task more. The effectiveness of this transfer learning approach is further elaborated in Section 3.3.

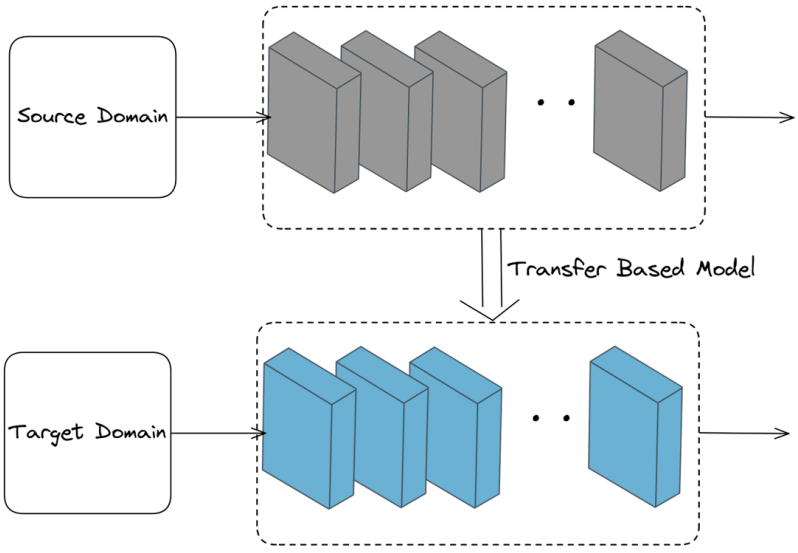


Figure 4. Model-based transfer learning.

2.2.2. Source Domain

The concept of transfer learning is believed to have originated from the theory of transfer generalization proposed by psychologist C. H. Judd, who proposed that transfer learning is the process of the generalization of experience. According to this theory, the prerequisite for transfer is that there should be a connection between two learning tasks [16]. In transfer learning, this connection means that the source and target domains and the source and target tasks should be similar and have a certain degree of transferability. Nowadays, there is no standard answer for the quantitative research of specific transferability. More of them are based on a priori knowledge and then combined with experiments for quantitative assessment of transferability. In this paper, we hope to test the TLB dataset for single-category small targets. The selection of source domains is crucial and is largely related to the final effect of transfer learning, and unsuitable source data domains may have negative transfer [22]. Considering the maturity of the prior knowledge, we selected two source data domains for transfer learning attempts, tea leaf shoots and apple tree leaf disease datasets, and the source tasks are both detection tasks shown in Figure 5. The same and different parts of the two source and target TLB data domains are specifically summarized below to exemplify the achievability of choosing these two datasets as source domains for comparison. Both the tea leaf shoots dataset and TLB in this paper were taken outdoors, with many tea leaves in the background, and the size of the shoots is a small target, like the diseased part of TLB. Still, the tea leaf shoots are not as contrasty as the diseased part. The normal part of TLB is due to the similarity in color and immature tea leaves, and the tea leaf shoots are more densely populated than TLB, which may also be an influential factor affecting the transfer of knowledge. The difference between the diseased dataset of apple leaves and the TLB dataset is that it does not have a large number of healthy leaves in the background like the TLB dataset in this paper; it only has a small number of healthy leaves or even none, and the foreground back scene is obvious. However, since the size and color of the diseased parts of the plant are similar and send out the background influence, this may be beneficial for the effect of knowledge transfer. We relabeled the sample targets of multiple TLB-like diseases of apples as disease tags and used the ability to identify them for transfer. Due to the superior training results on the apple leaf dataset, we finally chose the apple leaf disease dataset as the source domain, and more specific comparison data and procedures are shown in Section 3.2.

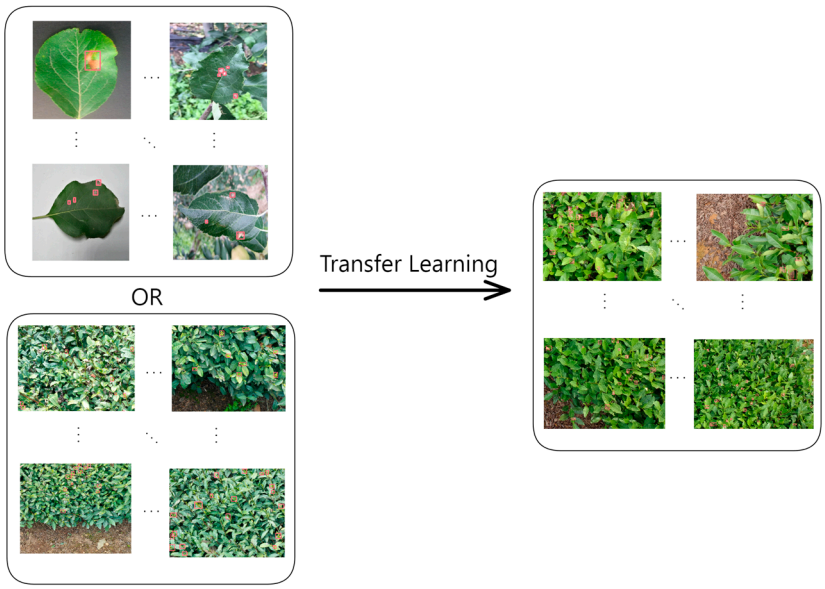


Figure 5. Choose an appropriate source domain to transfer. Note that the red boxes indicate the labeling for the detection part of the sample.

2.3. Model Construction

2.3.1. Summary of Our Model

Our model was built with reference to YOLOv7 [23]. The model, which is shown in Figure 6, consists of three main parts: the backbone feature extraction network, the neck feature fusion network, and the target detection head. The backbone feature network is used to extract feature information from the input image, the neck is a fusion feature fusion network that combines features from other relevant layers with those extracted by the backbone network, and the target detection head uses the output of the neck feature fusion network to detect the target. The introduction of the TSCODE decoupling header, the Triplet Attention module, and the NWD metric are all considered to better overcome the difficulty of recognizing small targets in TLB. The model is optimized for monitoring task decoupling, feature extraction, and loss function metrics, respectively, so that the model can better identify small targets in TLB. The benefits of each of them for the current task, and our considerations and thoughts when choosing them will be elaborated on in the subsequent introductions of each. The specific ablation experiments and comparison test results are presented in Sections 3.3 and 3.4.

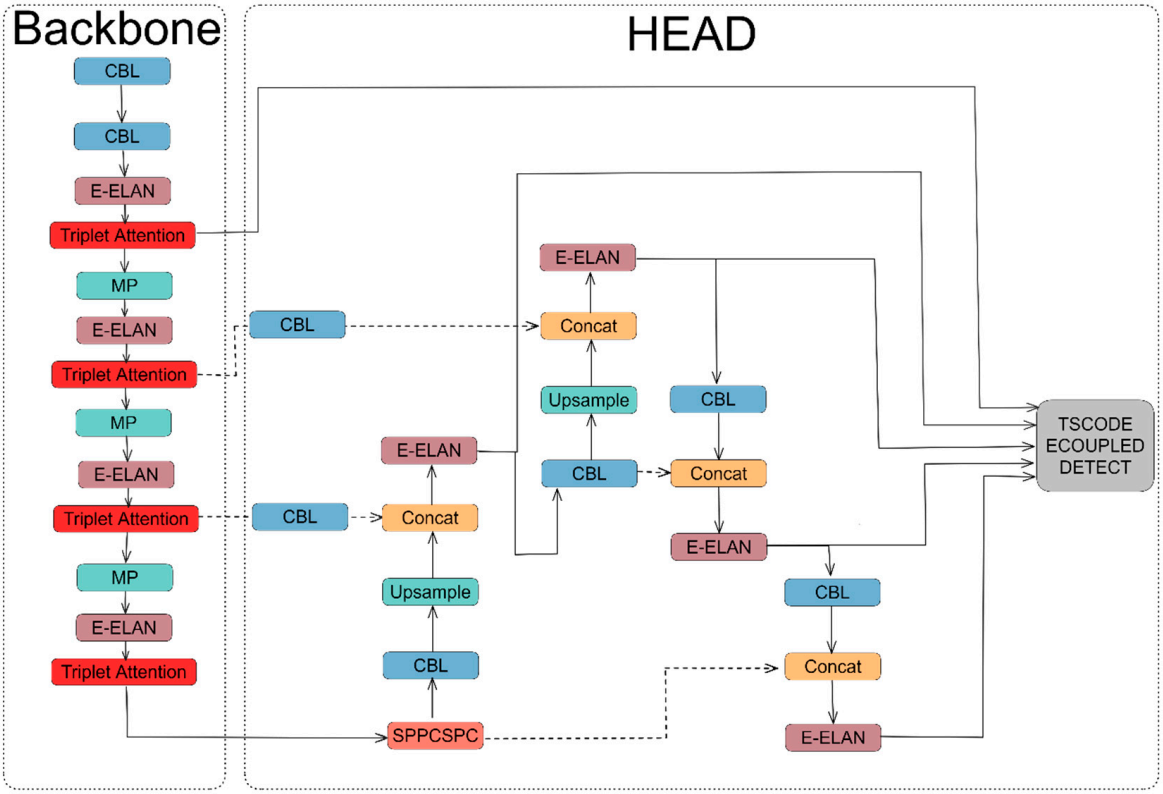


Figure 6. Structure of our model.

2.3.2. Basic Module

The E-ELAN module in our model undergoes tandem-based deflation, which means it reduces the network width and depth through the reduction in the number of CBL (Conv + BN + LeakyReLU) modules, i.e., reducing the number of branches in the E-ELAN, as shown in Figure 7. The overall modular design approach of an extended efficient layer aggregation network continued.

We continue to apply the idea of multi-branch module stacking combined with the SPP (Spatial Pyramid Pooling) [24] structure after the backbone network, which effectively separates important contextual features and increases the sensory field based on the fact that it can also accept any size of feature map input. Compared to SPPCSPC in the original version, the number of CBLs before the three parallel maximal pooling layers is reduced

from 3 to 1 in the tiny version, and the number of CBL modules between two Concat layers is reduced from 2 to 1, as shown in Figure 8, which reduces a large number of parameters and computations.

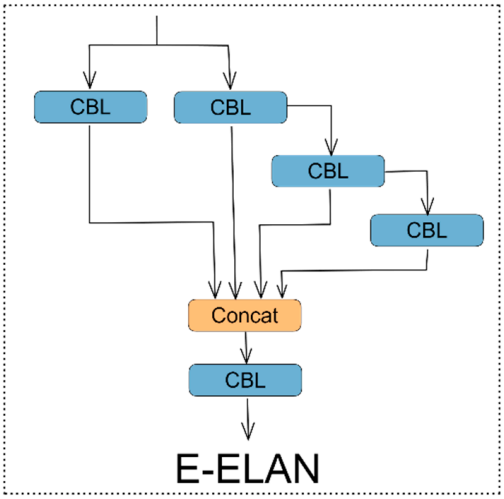


Figure 7. Structure of the E-ELAN module.

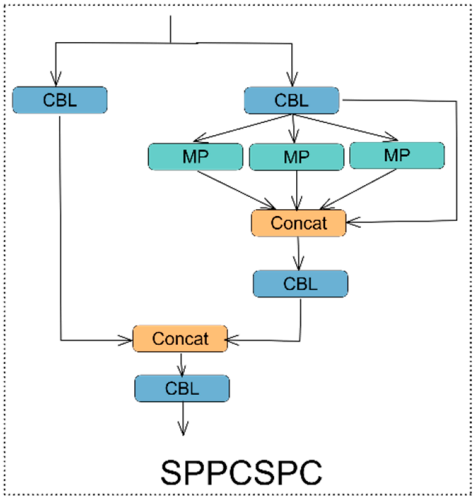


Figure 8. Structure of the SPPCSPC module.

The neck network architecture of our basic model utilizes the PAFPN (Path Aggregation Pyramid Network) structure. The PAFPN consists of two main parts: a top-down feature fusion network and a bottom-up feature extraction network. Each feature layer contains feature information of different sizes, and by fusing the low-sampling layer with the high-sampling layer, it realizes the feature processing of multiple scales while retaining the high-resolution information, preserving the detailed information of the image, and improving the model's detection ability for objects of different sizes [25,26].

2.3.3. TSCODE

The idea of single-stage algorithms is to turn the target detection problem into a regression problem, where probabilistic prediction of the regression bounding box is made directly after feature extraction and feature fusion. Our basic model uses a shared detection head common in the YOLO lineage for the classification and localization tasks. Despite the inclusion of several training tricks, the problem still exists that the coupling of the two subtasks to each other affects the recognition of the network to some extent, as there is a kind of spatial misalignment between the two tasks [27].

In this paper, the small target characteristic of TLB and various occlusions in the dataset will exacerbate the conflict between the two classification tasks. At the bottom network, there is little semantic information, but TLB boundary features are well preserved. At the deep network, there is rich semantic information, but due to the small target nature of TLB, certain boundary information may have been lost. Traditional decoupling heads used in YOLOx [28], for example, tend to be parameter decoupled, i.e., they still share the use of the same feature inputs but can use separate parameters for each of the two tasks. Considering such a problem, we introduce TSCODE [29], a decoupling head that combines multiple scales. TSCODE was proposed by Jiayuan Zhuang et al. Its overall structure is shown in Figure 9.

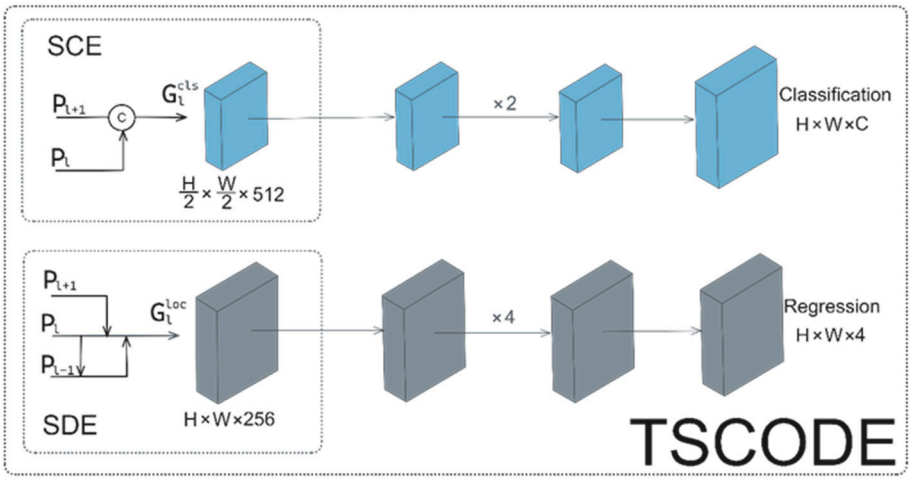


Figure 9. The overall structure of the TSCODE detector.

For localization, a task reliant on richer texture and boundary details, the SCE (Semantic Context Encoding for Classification) branch of TSCODE can incorporate shallow information into the local layer and complement it with deeper high-level semantic information. For classification, a task demanding richer semantic information, the DPE (Detail-Preserving Encoding for Localization) branch of TSCODE merges features from deeper layers. It embeds them into the current feature map to enhance the classification effect (Figure 10).

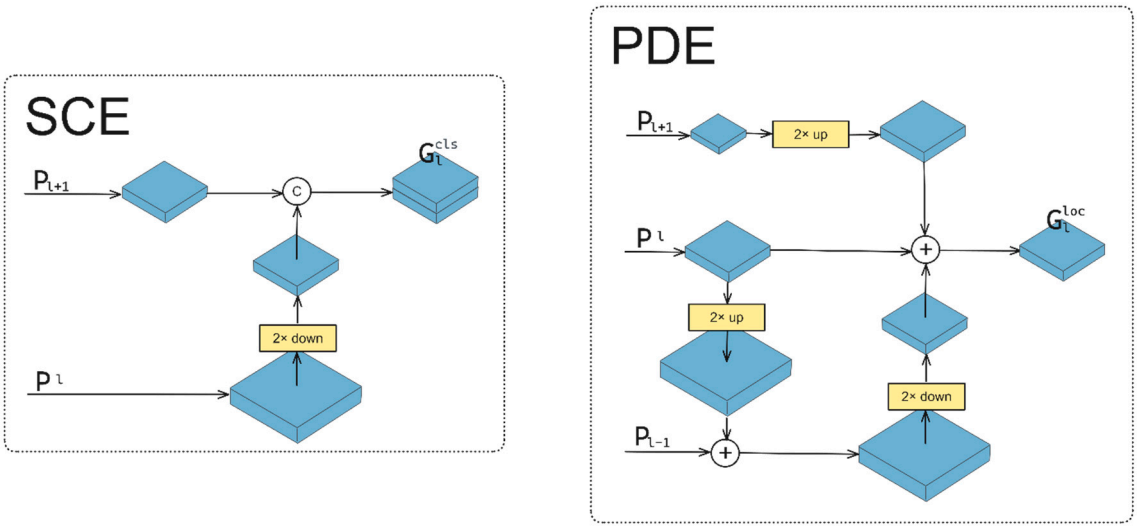


Figure 10. The detailed structure of the TSCODE detector.

In the TLB dataset of this paper, i.e., with more occlusion and dense small targets, the way of going further to obtain the feature maps suitable for specific tasks according to different tasks combined with the PAFPN structure alleviates the contradiction between the classification and localization tasks and reduces the phenomenon of omission and faulty detection, which is commonly found in the monitoring of small targets. In addition, the multi-scale feature fusion approach also alleviates the situation of those samples with large feature loss or even feature disappearance in downsampling to a certain extent, which can further improve the recognition generalization ability of the model.

2.3.4. Triplet Attention

Due to the small size of the TLB target, it has little information of its own. Moreover, the TLB samples have both scattered and accumulated scenarios, which makes it more difficult for the model to create a targeted treatment for the target. Therefore, we consider introducing Triplet Attention after the E-ELAN module in the backbone part of the model, which can extract the critical information of the target more efficiently and comprehensively and also help the model to focus on the crucial areas of the TLB target, which helps to retain and convey the critical information better in the whole network, thus improving the detection performance of small TLB targets.

Triplet Attention [30] is a method that captures cross-dimensional interactions through a three-branch structure, thereby enabling the computation of attention weights. For the input of a tensor, i.e., the three dimensions, channel (C), height (H), and width (W), the three branches establish the dependencies between (C, W), (C, H), and (H, W), respectively, to take into account the exchange of information between different dimensions. They are maximizing the most critical information. The structure of this method is shown in Figure 11.

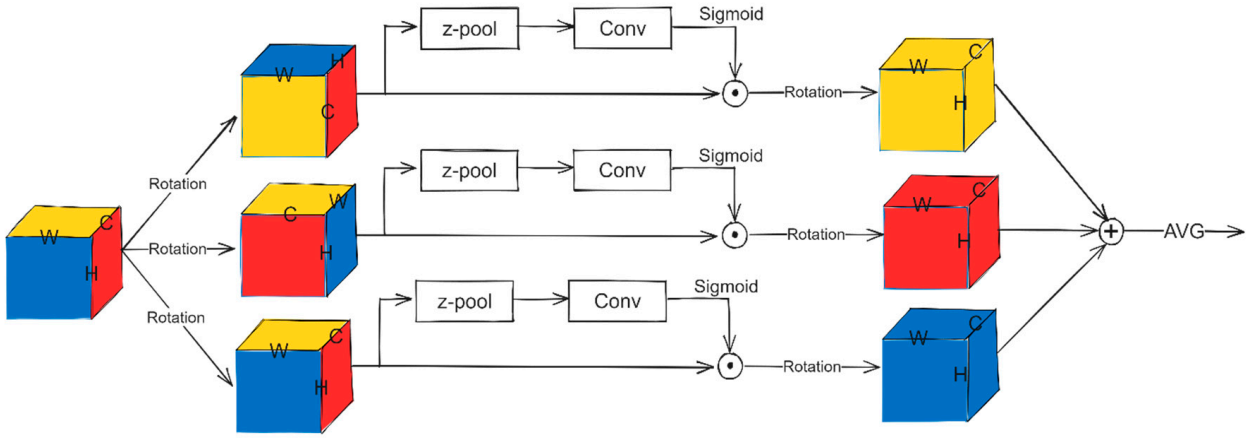


Figure 11. Structure of the Triplet Attention method.

To realize the cross-dimensional interaction, the ternary attention mechanism performs a rotation operation on the input tensor in each branch. Then, it uses a Z-pool layer and a convolutional layer to generate the attention weights. In the first branch, the tensor is rotated by the W dimension of the tensor to obtain a new tensor $X \in R_{H \times C \times W}$, which captures the relationship between the channel dimension C and the spatial dimension W. In the second branch, the tensor is optionally fitted by the H dimension of the tensor to obtain the new tensor $X \in R_{W \times H \times C}$, which captures the relationship between the channel dimension C and the spatial dimension H. In the third branch, the tensor is not rotated, i.e., similar to the traditional spatial attention mechanism, to capture the information between the spatial dimensions W and H. The Z-pool layer is a dimensionality reduction operation, where dimensions that are not needed to capture the dependencies are downgraded to 1. Then, the result of average pooling and maximal pooling is performed by splicing to pool the feature information in that dimension. Then, each branch undergoes another

convolution and then passes through the sigmoid activation function to obtain the weights of the attention machine for each branch. The weights of each branch are multiplied by the corresponding tensor and then rotated back to the C × H × W dimensional representation. The tensors of each branch are averaged to obtain the final tensor.

Due to the introduction of the Triplet Attention mechanism in the backbone part, more detailed information can be obtained in the target area of the network concentration, which can improve the feature extraction ability of the model and effectively reduce the interference of invalid targets. The weight information captured from multiple dimensions can better improve the detection effect on the small targets of the TLB of concern and achieve the purpose of improving the overall detection effect of the model.

2.3.5. Loss Function Based on Normalized Wasserstein Distance

TLB small targets are difficult to detect because the number of pixels they occupy in the feature map is particularly small. They are very sensitive to positional deviation, especially in IoU-based metrics [31]. This sensitivity easily leads to the positional deviation of small targets, which will cause performance degradation in the parts of the small target label assignment, non-maximum suppression, and loss function. The final localization detection of small targets is poor, and the recognition model for small targets is often difficult to converge. Jinwang Wang [32] proposed a new Wasserstein distance-based small target detection evaluation method, i.e., considering the fact that the object pixels and background pixels in BBox tend to be concentrated on its center and boundaries. The BBox is modeled to a 2D Gaussian distribution so that the center pixel of the BBox has the highest weight, and the weight of the pixels decreases from the center of gravity to the boundary. For a BBox, its interior elliptic equation can be expressed as follows:

$$\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_x^2}=1 \tag{2}$$

(μ_x, μ_y) are the coordinates of the center of the ellipse. σ_x, σ_y are the lengths of the semiaxes along the x and y axes, so that $\mu_x = cx$, $\mu_y = cy$, $\sigma_x = w/2$, and $\sigma_y = h/2$. The probability density function of its two-dimensional Gaussian distribution is as follows:

$$f(x|\mu,\Sigma)=\frac{\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}{2\pi|\Sigma|^{\frac{1}{2}}} \tag{3}$$

x,μ,Σ denote the two-dimensional coordinates (x,y), the mean vector, and its covariance matrix, respectively. At this point, the BBox can be modeled as a two-dimensional Gaussian distribution where it denotes the center coordinates, width, and height.

$$\mu=\begin{bmatrix}cx\\cy\end{bmatrix}, \Sigma=\begin{bmatrix}\frac{w^2}{4} & 0\\0 & \frac{h^2}{4}\end{bmatrix} \tag{4}$$

Thus, the similarity between BBox can be converted into the distribution distance between two Gaussian distributions. Wasserstein distance is used to compute two two-dimensional Gaussian distributions, and the distribution distance is as below. The subscripts 'a' and 'b' represent the respective two BBoxes.

$$W_2^2(\mathcal{N}_a,\mathcal{N}_b)=\left\|\left(\left[cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2}\right]^T, \left[cx_b, cy_b, \frac{w_b}{2}, \frac{h_b}{2}\right]^T\right)\right\|_2^2 \tag{5}$$

Since it is a distance metric and cannot be used directly as a similarity metric, it is normalized in exponential form to obtain the final NWD new metric, where C is a constant related to the specific dataset as follows:

$$NWD(\mathcal{N}_a, \mathcal{N}_b) = \exp\left(-\frac{\sqrt{W_2^2(\mathcal{N}_a, \mathcal{N}_b)}}{C}\right) \tag{6}$$

Since Wasserstein distance can better measure the geometric similarity between two distributions and can handle the correspondence between them more stably, the introduction of the Wasserstein distance measure in the loss calculation of the bounding box can better focus on the geometric similarity between the predicted bounding box and the target bounding box, instead of the positional and dimensional differences, and focus more on the localization effect in the task. As a result, it can help our model focus more on the localization effect in the task. The pairwise invariance in the scenario of two disjoint or overlapping bounding boxes and the smoothness of the positional offset can reduce the sensitivity to the positional deviation of the small target identification. Considering that the offset rotation mosaic and other methods are randomly added to the data enhancement in this paper, it can make the identification of small targets more robust in some special cases. The combination of the Wasserstein distance metric and the CIoU metric can form a more integrated and comprehensive loss function, which can better guide the model to improve the identification and localization ability in the detection of small targets.

2.4. Evaluation Method

The evaluation metric used in this paper is the average precision (mAP@0.5) under the IoU threshold of 0.5 since there is only one sample category in this paper, i.e., the mAP value is consistent with the AP value. Furthermore, we also use the F1-score to measure the accuracy and recall capability of the model comprehensively. There are also precision and recall, the formulas for which are shown as follows:

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

$$AP = \int_0^1 P(r)dr \tag{9}$$

$$F_1 - Score = \left(\frac{\mathrm{R}^{-1} + P^{-1}}{2}\right)^{-1} \tag{10}$$

In these equations, TP denotes the number of TLBs labeled and detected as such, and FP denotes the number of misdetections labeled non-TLBs but detected as TLBs; in other words, it denotes the number of false detections. Similarly, FN indicates the number of TLBs labeled TLBs but not detected, i.e., the number of missed sightings. Indicator P reflects the accuracy of the detection of TLBs, and indicator R reflects whether all TLBs were detected. AP is the area under the P-R curve and indicates the average accuracy of TLB detection.

3. Results

3.1. Training Environment and Parameters

In this section, Table 1 shows some of the language environments and the hardware and software environments during the experimental process. Table 2 shows the number of training sets, validation sets, and test sets for the source and target domains, which have all been augmented after the 7:1.5:1.5 division of the two data. Table 3 and Figure 12 show some results for parameter selection. Tables 4 and 5 show the training parameters of the model before and after transfer learning respectively.

Table 1. Experimental conditions.

Experimental Environment	Details
Programming language	Python 3.8.10
Operating system	Windows 11
Deep learning framework	Pytorch 1.12.1 + CUDA 11.7 + cuDNN 8.6.0
GPU	NVIDIA GeForce RTX 3080 [1]

[1] The manufacturer of the graphic card is ASUS in Nanjing, China.

Table 2. Details of the two datasets.

Source Domain	Train	Val	Test	All
Apple leaf disease	469	103	101	673
Tea leaf blight	126	28	28	182
Apple leaf disease (augmentation)	2345	515	505	3365
Tea leaf blight (augmentation)	2181	479	470	3130

Table 3. Comparison of model performance for parameter selection.

Optimizer	Learning Rate	$AP\%_{0.5}\%$
SGD	0.01	85.3
	0.005	85.1
	0.001	78.4
Adam	0.01	83.3
	0.005	84.5
	0.001	85.0

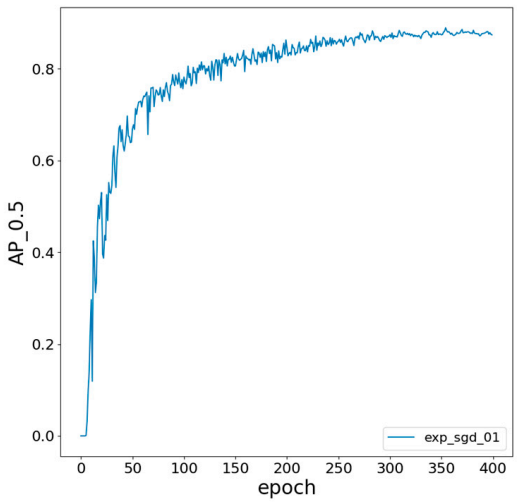


Figure 12. Model performance using SGD and setting the learning rate at 0.01.

Table 4. Training parameters of the apple leaf disease detection model.

Training Parameters	Details
Epochs	300
Batch size	16
Img size	640 × 640
Initial learning rate	0.01
Optimization algorithm	SGD
Pre-training weights file	None

Table 5. Training parameters of the tea leaf blight detection model.

Training Parameters	Details
Epochs	300
Batch size	16
Img size	640 × 640
Initial learning rate	0.01
Optimization algorithm	SGD
Pre-training weights file	The best.pt obtained by the apple leaves disease detection model

The tea leaf blight (augmentation) dataset does not include the original samples before augmentation, i.e., the original tea leaf blight samples are not visible to the model trained after data augmentation. Thus, in the visualization of the detection effect in this paper, in order to fit the real scene, we use all the sample images in tea leaf blight and 182 images as the test set of the model so that the inference results obtained take into account the model's generalization ability in the real scene.

In deep learning, optimizers and learning rates impact model performance. SGD and Adam are classical optimizers in model training. The basic idea of SGD is to continuously adjust the parameters of the model through gradient descent to minimize the model's loss function. Adam's basic idea is to adjust the parameters of the model by maintaining the first and second moments of the model's gradient and the square of the gradient. Therefore, to achieve better performance of the model, we compared the detection performance of the model under different learning rates using the SGD optimizer and the Adam optimizer, respectively. The choice of learning rates in the comparison is based on empirical reference. The results of the comparison experiments are shown in Table 3. Furthermore, we set the iteration parameter to 400 in the comparison experiments to discover the influence between different epochs which is presented in Figure 12.

Due to hardware limitations, we set the batch size to 16 and the image size to 640 × 640. The initial learning rate and optimizer selection are based on the table above. With the optimizer set to SGD and the initial learning rate at 0.01, the model's performance tends to be relatively higher. Performance improvement of the model using SGD and setting the learning rate at 0.01 becomes very marginal after 300 iterations, as indicated in Figure 12. To save computational resources, we chose to set the epochs to 300 for subsequent experiments. The specific experimental parameters during transfer learning are shown in Tables 4 and 5.

3.2. Transfer Learning Source Domain Selection

The core problem of transfer learning is to find the similarity between the two domains. Once this similarity is found, it can be rationally utilized to perform the transfer learning task well. Still, suppose the similarity between the two domains found is not rational or does not exist. In that case, it is impossible to perform the existing task, meaning there will be a negative transfer. For the source domain selection problem of transfer learning, existing research methods, such as the Selective Adversarial Network [26] proposed based on adversarial learning, are generally complicated. In our study, we propose a hypothesis. We first assume that the pre-trained model can recognize TLBs to some extent, i.e., the knowledge drawn from the source domain is already helpful for recognizing TLBs without transferring. The one with better recognition ability is the source domain that is more suitable for transfer learning of TLBs. We default that all the labels marked by TLBs are invisible to the source domain, and then the detection task is performed directly on its test set. The detection results are shown in Table 6, which indicates that the detection task of apple leaf disease can already recognize TLBs in a certain amount, i.e., the knowledge it has acquired on its original task is helpful for detecting TLBs to a certain extent. Meanwhile, the knowledge from the detection task of tea shoots is almost not helpful for TLB detection. Then, we use the pre-trained weights of the two alternative source domains to transfer

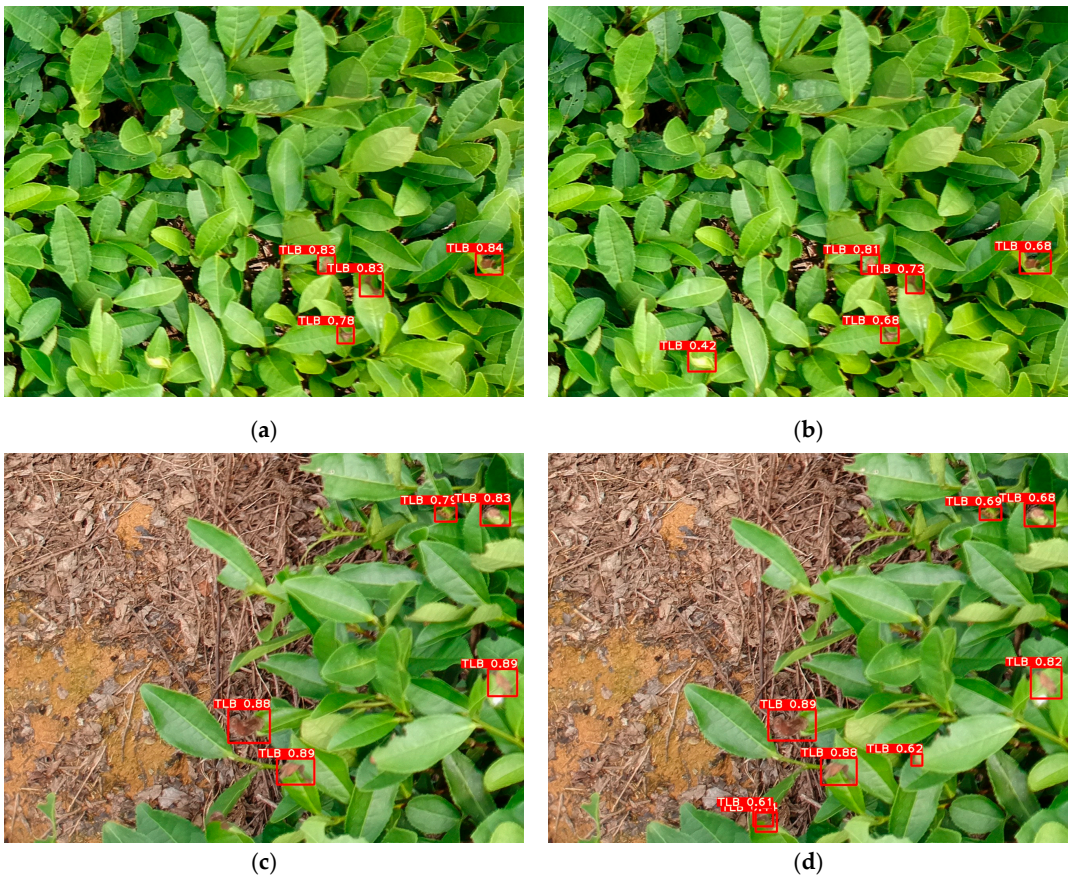
on the model and then train it again. The final data metrics and recognition results we obtained also prove the reasonableness of this hypothesis, as shown in Table 7. Observing the training results, we also have certain findings that although the gap between the weights of the models trained on the two source domains is large on the test set, the gap between the training results as pre-training weights is not very large. The detection plots of the same images after training on the apple leaf disease dataset and the tea shoot dataset, respectively, are given in Figure 13 to visualize the disparity in effect between the different source domains. Although the detection effect of the apple leaf detection model for TLB is better than that of the tea shoot detection model, the situation of missed detection and false detection still exists, so it triggers further optimization of the TLB detection model in the next step. In summary, we finally chose the apple leaf disease set as the source domain for this study.

Table 6. Testing results of tea leaf blight directly using source domain pre-trained models.

Source Domain	Precision	Recall	$AP_{0.5}$
Apple leaf disease	0.182	0.154	0.0738
Tea shoots	1.47×10^{-4}	9.52×10^{-4}	1.32×10^{-6}

Table 7. Training results of transfer learning based on the source domain.

Source Domain	Precision	Recall	$AP_{0.5}$
Apple leaf disease	0.886	0.836	0.876
Tea shoots	0.764	0.799	0.813



(**a**) (**b**)

(**c**) (**d**)

Figure 13. Detection results of the same tea leaf blight pictures on different source domains: (**a**,**c**) using apple leaf disease as the source domain and (**b**,**d**) using tea shoots as the source domain.

3.3. Ablation Experiments

To better test our proposed model's performance and verify each improvement method's necessity, we conducted ablation experiments. Based on transfer learning, i.e., using the weights obtained by training the model on the source domain as the initial weights for training, we added one improvement model method at each step of the ablation experiments to validate the improvement effect of each improvement method and the combination of different methods.

The experimental results of the ablation experiments are given in Table 8. TL denotes the introduction of transfer learning, TS denotes the incorporated TSCODE detection header, TA denotes the incorporated Triplet Attention module, and NWD denotes the use of a loss metric based on the Normalized Gaussian Wasserstein Distance. Baseline refers to experiments using the YOLOv7 tiny model directly after dataset augmentation.

Table 8. Results of ablation experiments.

Model	Precision%	Recall%	$AP_{0.5}$%	F1-Score%
YOLOv7 tiny (baseline)	85.7	82.1	85.7	83.4
YOLOv7 tiny (TL)	86.8	84.2	87.9	85.5
YOLOv7 tiny (TL) + TS	87.6	85.8	89.1	86.7
YOLOv7 tiny (TL) + TA	87.6	84.8	88.7	86.2
YOLOv7 tiny (TL) + NWD	90.9	83.2	89.2	86.9
YOLOv7 tiny (TL) + TS + TA	90.8	84.3	89.9	87.4
YOLOv7 tiny (TL) + TS + NWD	89.9	**87.0**	89.9	88.4
YOLOv7 tiny (TL) + TA + NWD	90.1	86.3	89.9	88.2
YOLOv7 tiny (TL) + TS + TA + NWD (ours)	**92.2**	86.6	**91.5**	**89.3**

The bold represents the best result during the ablation experiments.

Although YOLOv7 tiny is one of the few lightweight models with the best detection and recognition results, its precision and recall values are still relatively low. After the introduction of transfer learning, the model has a certain degree of improvement compared to the baseline: precision increased by 1.1%, recall increased by 2.1%, AP increased by 2.2%, and F1-score increased by 2.5%. This indicates that the knowledge drawn from the source domain of apple leaf disease has played a helpful role in recognizing and detecting TLB, and the introduction of transfer learning is effective.

Improvement on the detection head, decoupling the original coupled detection head, and using the TSCODE detection head for recognition based on different tasks have improved the classification and localization ability of the model. The model has improved its precision by 1.9%, recall by 3.7%, AP by 3.4%, and F1-score by 3.3% compared to the baseline. The way TSCODE acquires the feature maps suitable for a specific task eases the contradiction between the classification and localization tasks and reduces the common phenomenon of missed and wrong detection in TLB small target detection.

Adding the Triplet Attention mechanism to the E-ELAN structure of the backbone improves the model's precision by 2.1%, recall by 2.7%, AP by 3.0%, and F1-score by 3.3%. Due to the introduction of Triplet Attention, the weight information captured from multiple dimensions can better enhance the focus on small TLB targets and make the target area in the network set to obtain more detailed information, which can improve the overall detection effect of the model.

The new NWD-based loss function metric improves precision by 5.2%, recall by 2.1%, AP by 4.5%, and F1-score by 4.2%. This suggests that the new metric can indeed be effective in reducing the sensitivity of small target localization locations and improving the classification ability to a certain extent at the same time. Including the Wasserstein distance metric reduces the positional sensitivity to TLB small targets to better guide the model's TLB small target detection ability during training.

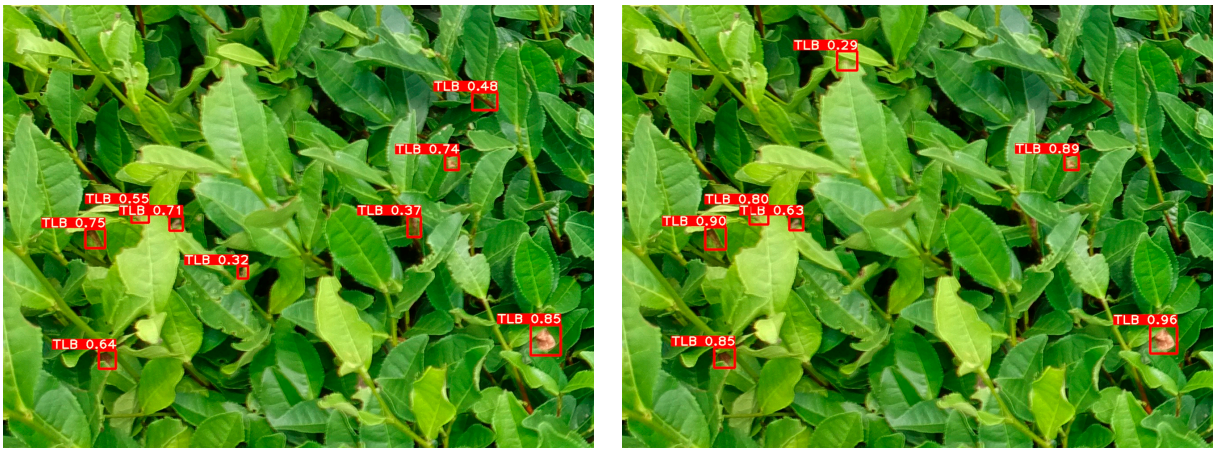
Next, we further improve the model by trying different fusions of the three improved approaches to test the model's performance. According to the results of the ablation experiments, it can be found that the combination of these several improvement methods has a certain degree of improvement in the AP compared with a single improvement,

which reaches 89.9%. Still, some fusion methods lead to a slight decrease in precision and an increase in recall. In contrast, some have the opposite effect, i.e., precision is increased and recall is decreased, which suggests that the three improvement methods have different focuses on the performance enhancement of the model. This shows that the three improvement methods focus on the performance improvement of the model differently. We finally integrate all three improvement methods to obtain our final improved model, which achieves 92.2% in precision, 86.6% in recall, 91.5% in AP, and 89.3 in F1-score, which are 6.5%, 4.5%, 5.8%, and 7.1% higher than the baseline, respectively, and show better performance. This shows that our improvement approach is necessary for TLB detection capability enhancement, which can effectively reduce the occurrence of false and missed detections.

3.4. Comparison

We randomly selected a few more characteristic samples in the dataset for the detection task, compared them with the actual detection capabilities of our proposed model and YOLOv7 tiny, and presented the results in a visual way to show the differences between the two. In the following, the left image is the detection result of the baseline and the right image is the detection result of our model.

Figure 14 shows the detection scenario of TLB in general, and from the recognition results, we can find that our model is better than the baseline in recognition; both are wrongly detected and missed in this sample, but our model performs better in recognition confidence.

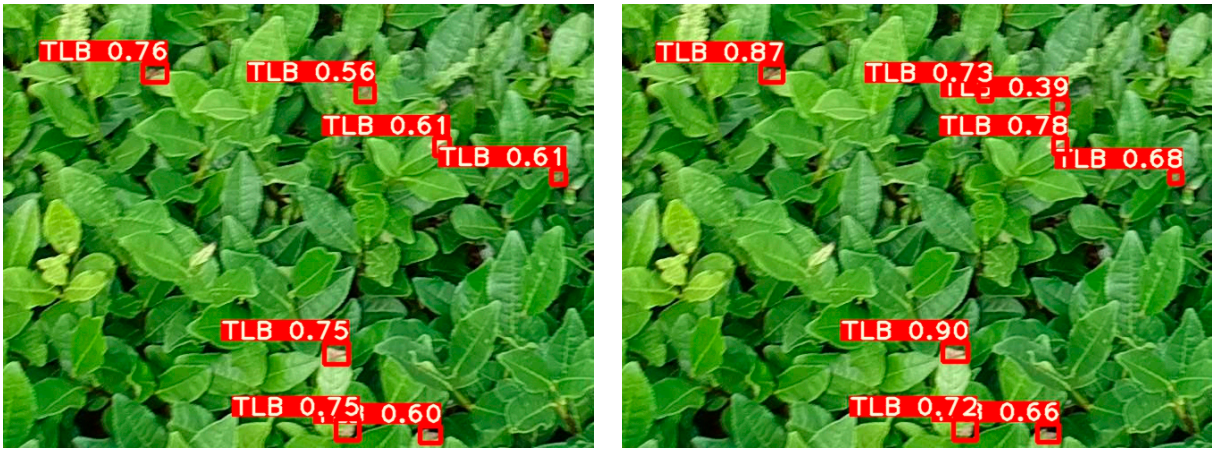


(**a**) (**b**)

Figure 14. Detection of general scenarios. (**a**) YOLOv7's detection results. (**b**) Our model's detection results.

Figure 15 shows the detection scene when the shot is out of focus; the baseline model has a missed detection, but our model recognizes it, and under the same TLB small target recognition, our model has a smaller recognition frame, the localization effect is more accurate, and the detection confidence of our model is significantly better than that of the baseline model.

Figure 16 shows the detection scenario under the influence of the background of dead leaves, and the baseline model has several misdetections. However, none of them have a high confidence level, which is a disturbance to the detection results. Our model not only recognizes two TLBs in the sample with high confidence but also does not have one false detection. The baseline model is significantly worse than our model in the case of dead leaf background.



(**a**) (**b**)

Figure 15. Detection of out-of-focus scenarios. (**a**) YOLOv7's detection results. (**b**) Our model's detection results.

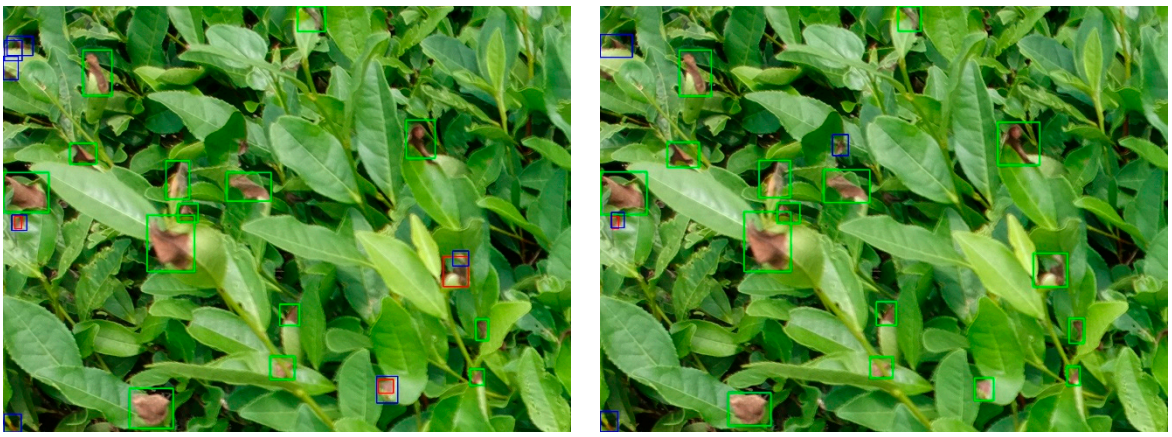
(**a**) (**b**)

Figure 16. Detection of background interference with dead leaves. (**a**) YOLOv7's detection results. (**b**) Our model's detection results.

In the scenario where TLB is denser, the direct output of the resulting effect is more chaotic and complex, and the performance difference between the models cannot be compared well. We compare and contrast the recognition detection frame and the sample detection frame to re-visualize the recognition effect; the green recognition frame displays the correctly detected target, i.e., TP, the blue recognition frame displays the missed target, FN, and the red recognition frame displays the wrongly detected target, FP. After randomly selecting a sample, the visualization of this approach based on a confidence level of 0.45 is shown in Figure 17. Our model has fewer false detections and fewer missed detections than the baseline model in this scenario, and the model has better TLB detection performance.

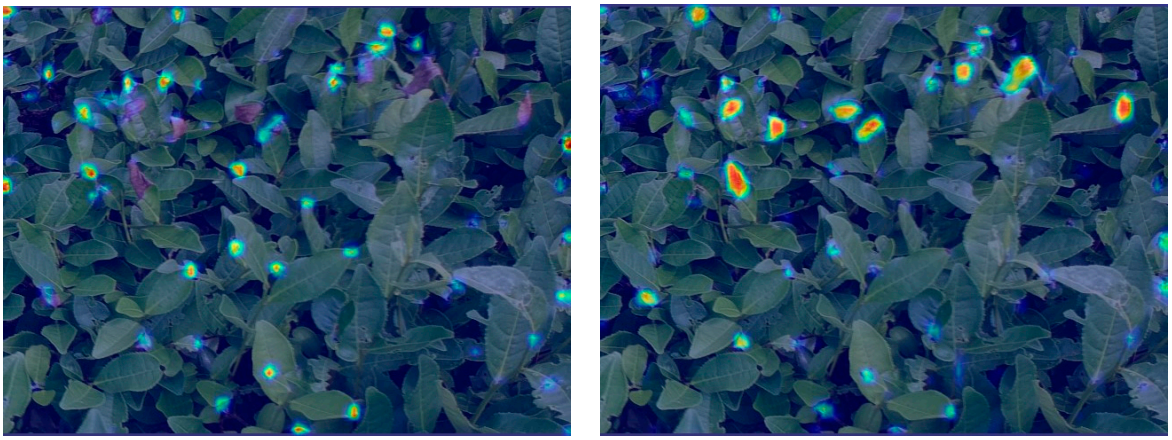
Finally, after introducing the three improvements, we use the grad-cam approach to visualize the feature maps output after the first E-ELAN module in the head section of the baseline model and the head section of our model. Since this layer belongs to the shallower layer of the head section, it retains more information about the small target features, which allows us to compare the difference in the tendency of the models in small target detection more directly. Figure 18 is the visualization effect of a randomly selected sample. In the feature map of this layer, our model can focus more on the target to be identified in the relatively larger target, which is concentrated, and the center of the weight tends to be higher. In the case of TLB, the target itself is very small, which can ensure

a certain amount of target localization and classification accuracy and is less susceptible to some environmental interference. The baseline's model feature map of the TLB target basic baseline model seems to treat TLB targets equally. Still, the size of the heat map of the area attached to the target is relatively small, which brings more challenges to the model's localization and classification ability and the model's ability to be affected by the interference to a certain extent, which is also verified by the experimental results, and the baseline model's robustness of the generalization ability is a little bit poorer.



(**a**) (**b**)

Figure 17. Detection of target-intensive scenarios where green boxes represent TP, blue boxes represent FN and red boxed represent FP. (**a**) YOLOv7's detection results. (**b**) Our model's detection results.



(**a**) (**b**)

Figure 18. Feature maps visualized using heat maps. (**a**) YOLOv7's feature map. (**b**) Our model's feature map.

At the same time, we compared our model with some mainstream detection models to better reflect its advantages. They include SSD, RetinaNet, and Efficientdet, which are also single-stage algorithms like our model. In addition, we have also added comparisons of two-stage algorithms, like Faster-RCNN. The comparison of model performance on the validation set is shown in Figure 19.

From the recall values of SSD, RetianNet, and EfficientDet, it can be seen that the positioning of TLB small targets is very difficult, which is why this article believes in the significance of TLB small target recognition. Although the performance of Faster R-CNN and YOLOv5s is relatively good, it still cannot reach a satisfactory level. The overall comparison shows that our model has certain advantages in recall, precision, and

AP0.5 performance. This shows that our model can ensure good recognition of TLB small targets when the dataset is small.

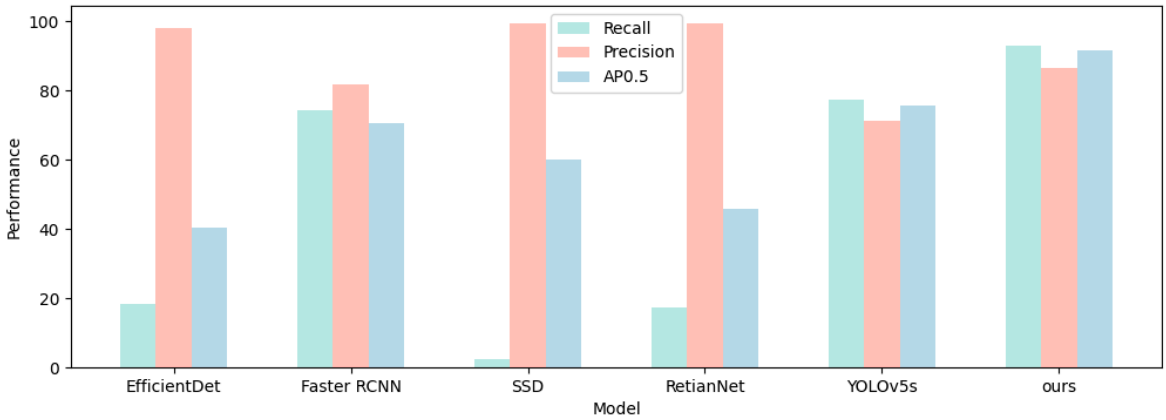


Figure 19. Model performance comparison.

4. Discussion

The research idea of this paper is based on two main aspects: one is to improve and optimize the small sample problem based on the scarcity of datasets related to tea pests and diseases, and the other is to improve and optimize the problem with relatively tiny TLB objectives.

For the small sample problem, which is a relatively common situation for the tea leaf disease detection task, such a situation will greatly affect the generalization ability of the model in deep learning and is prone to overfitting. Therefore, in this paper, firstly, on the basis of the original dataset, offline and online data enhancement are combined, making the TLB dataset samples as first as possible. In the training strategy, transfer learning is introduced. In transfer learning, the selection of a source domain has a great relationship with the model's performance after transfer, so we compare two source domains before transfer One is apple leaf disease detection with different goals and similar tasks, and the other is tea shoot detection with similar goals and different tasks. The way source domains are selected in general transfer learning tends to be more complex and is another difficult challenge. A hypothesis is proposed in this paper, which means going for direct validation on the validation with model weights obtained directly from training on different source domains. The dataset that works better is the more suitable source domain. This is also verified in the subsequent learning effect after transfer. Apple leaves have better results, but we also found an interesting phenomenon in our experiments, i.e., the directly validated tea shoot source domains have improved results after transfer, which may be more similar to the distribution of TLBs and tea shoot. That is, although there is not much information drawn from the source domain that can directly identify TLB, its model also learns the distribution of this type to some extent, which gives us a new direction to draw on for future source domain selection strategies. Based on this finding, we looked for relevant transfer learning studies. We found that when this situation occurs, we can consider using partial domain adaptation, i.e., all source domains that can be helpful to the target domain are included in the scope of transfer learning. An example is Selective Adversarial Networks (SANs) [33] proposed by Zhangjie Cao. They avoid negative transfer by separating the samples of non-shared categories in the source domain and, at the same time, promoting positive transfer by maximally matching the distribution of samples in the shared category space. With this transfer approach, we can then try to learn all the knowledge embedded in the source domains of tea shoots and apple tree leaves.

For the small target problem, we mainly optimize and improve the network structure. The idea of YOLO is to turn the target detection problem into a regression problem and directly predict the probability of the regression bounding box after feature extraction and feature fusion, which ensures the model's lightness to a certain extent, but when

encountering problems such as the small target, the model's detection performance is often not enough to support practical applications. Therefore, we utilize a TSCODE detection head for fused context decomposition tasks in TLB detection, and the experimental results show that such a detection head with different levels of feature maps for different tasks with fused contexts is very helpful in improving the model performance. To further improve the model's performance, we try to optimize the feature extraction for small targets by adding Triplet Attention in the E-ELAN structure of the backbone part of the model. The amount of information on small targets is relatively small in the samples. Triplet Attention extracts the sample information from three dimensions together, and its introduction can maximize the information of small samples, making the network more focused on the key regions of small targets and improving the detection accuracy of small targets. Considering that the normal tea in the background of TLB is also denser and will have some impact on the detection of TLB, the introduction of an effective attention mechanism can help to inhibit the interference of the background on the small targets, which makes the network pay more attention to the information of the small targets and thus reduces the possibility of misdetection; the improvement of this situation is obvious in the visualization of the results. The results show that the detection recall is relatively low compared to precision. The reason may be that the target of TLB will be more sensitive to its localization position because it occupies fewer pixels in the sample and is smaller, so having a detection frame with a good regression effect is not easy. Therefore, we add a new metric on the loss function, i.e., based on Wasserstein distance, to carry out the positional difference metric between the detection frames. After optimizing and updating the model by incorporating the three improvements, the model has greatly improved its ability to locate and identify small targets. In the visualization of the heat map of the feature map, we can clearly find that the model has a preference to focus on the larger TLB targets among the same small targets, which also shows the imbalance between classification and localization in the detection task. This also shows the imbalance between classification and localization in the detection task. There is also a new idea for TLB small target detection, i.e., we can categorize the small targets that are already difficult to detect into harder and easier to detect and have a preference for detecting targets with high priority to improve the performance of the model. While reviewing papers on small object detection, I found that a strategy proposed by Chang Xu [31] aligns closely with the approach mentioned above. Their research revealed that there is often a bias towards larger objects due to IoU threshold-based and center sampling strategies, similar to our tendency to focus on relatively larger objects within small object detection in this paper. This presents a scale imbalance issue. To mitigate such concerns, the authors introduced a Region-based Feature Learning and Assignment (RFLA) strategy based on receptive fields to achieve balanced learning for small objects. This finding shares similarities with our founds above.

Of course, there are some limitations in the research process of this paper. First, the dataset in this paper only involves one tea disease; when multiple diseases or pests are involved, different pests and diseases may have similar recognition features, which poses a greater challenge to the model performance and is likely to have problems, such as sample imbalance. In addition, tea samples may vary over time, which may lead to misdetection and faulty detection. At the same time, the light quality of the samples is also an important factor; the photo samples taken in this paper are basically under normal light, which is not guaranteed in specific cases. The possible impact of the bias of the dataset on the generalization ability of the model, which we have also not elaborated in depth, could serve as an entry point for our subsequent research on the small sample problem. Secondly, this paper mainly focuses on optimizing the model by improving its performance. It does not consider the lightweightness of the model, i.e., it does not consider the deployment possibilities and improves its classification and localization capabilities purely from the model perspective. Therefore, in the subsequent design and optimization process of the model, it is necessary to focus on how to deal with these issues in order to improve the performance and usefulness of the model further. In the future, our further research will

continue based on the continuation of the research ideas in this paper, i.e., the small sample problem and the small target problem, in addition to the actual deployment and application of the model, which will also be included in the research scope.

In terms of the small sample problem, on the basis of some limitations existing in our study, the different approaches to transfer learning and the impact of different transfer layers on the model could serve as entry points for our future research. Different types of crops may require different transfer methods as well. Considering that some studies have proved that meta-learning can still maintain good recognition accuracy under small samples [34], we can introduce the idea of meta-learning to optimize the tea leaf disease detection model. For the problem of scarcity of tea leaf disease and pest datasets, we can use oversampling or replicating multiple copies [35] for the further enhancement of sample images. In addition, we consider the use of GAN [36] to expand the dataset further and enhance the generalization ability of the model. At the same time, GAN can effectively eliminate this dataset bias during the training of models with low-quality samples, and such a research idea can also be included in consideration of subsequent research [37].

For small-objective problems, the reinforcement learning approach, which has been prevalent in many other fields, can be borrowed to overcome and optimize small-objective problems [38]. It can also be combined with traditional image processing methods based on the distribution law of tea diseases on infrared grayscale images [39] to perform model training. In addition, we are concerned about the research content of hyperspectral technology in tea science-related aspects. Still, more often, we use hyperspectral technology in combination with traditional machine learning [40] or in combination with traditional image processing techniques [41] for a variety of classifications or quality detection. Combining it with deep learning may be able to deal with small target problems more effectively.

Regarding practical deployment application, the YOLO series of algorithms have demonstrated relatively good deployment effectiveness across multiple areas [42–44], which is what we aim to achieve in our further research. Thanks to the modular design ideas borrowed from YOLO in this paper, our algorithm can also be very scalable due to the plug-and-play nature of some modules. In subsequent deployments, we will try to replace the backbone with less computationally demanding and lightweight network structures, such as shufflenetv2, mobilenetv2, and ghostnet. These approaches have been implemented in other fields [45,46]. We can consider deploying the model in a UAV and using some optimization algorithms for route deployment to better detect tea health [47]. In addition, with the rise of the Internet of Things (IoT) and edge computing, we also hope to build a smart Internet of Things (IoT) hardware system, including the collection of tea leaf images by a high-definition zoom camera in a cluster structure and the deployment of a detection model through the edge computing nodes at the head of the cluster to realize the detection of tea leaf diseases [48]. As deployment and practical application issues are further explored, the ability of deployment and the deployment performance of the model become inevitable considerations. The choice of training methods and model usage may also be subject to certain limitations, with research potentially leaning towards lightweight and recognition speed. This is also one of the reasons why these two aspects were not introduced in this paper's evaluation methods.

At the same time, the idea of our research can be applied to other tea leaf pests and diseases and the detection of multiple tea leaf pests and diseases. With proper source domain selection and problem-specific structural optimization, the model can be made to perform better. A comprehensive tea leaf detection model with better adaptability is our ultimate goal.

5. Conclusions

Tea is an important cash crop in traditional agriculture. Protecting the growth of tea trees and the quality of tea leaves is crucial to the development of the tea industry. However, tea pests and diseases have a significant impact on tea yield and quality, leading to an

average annual reduction in production of about 20%. Therefore, studying the accurate detection of tea pests and diseases is of great significance.

We selected tea leaf blight (TLB), a single species of disease, for our study, and after an in-depth study of it, we found that TLB has a small sample problem with respect to the dataset. To address this issue, we employ data augmentation and transfer learning. We transferred models pre-trained on the large-scale source data domain to the TLB dataset, followed by further model optimization for small targets.

During the detection process, due to the high overlap between diseased and healthy tea leaves and the small size of some of the TLB targets to be recognized, we worked on solving these small target problems to reduce the occurrence of false and missed detections. We utilize a decoupling detection head TSCODE in the detection head section and introduce Triplet Attention on the original E-ELAN structure. In addition, a small target detection evaluation method based on Wasserstein distance is introduced, significantly improving the model's ability to recognize small targets of tea diseases.

The experimental results on the TLB small sample dataset show that our model can better cope with the small sample and small target detection problems with higher accuracy and robustness compared to the traditional YOLOv7 tiny model. Overall, this study successfully optimizes the detection model, which provides an effective solution for the accurate identification of tea tree leaf blight and an idea for building an effective detection model for tea tree leaf pests and diseases.

Author Contributions: X.Y. was responsible for program design and drafting the initial manuscript. H.Z. assisted with data collection and analysis. H.L. and D.B. designed the project and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Jiangsu Modern Agricultural Machinery Equipment and Technology Demonstration and Promotion Project (NJ2021-19) and the Nanjing Modern Agricultural Machinery Equipment and Technological Innovation Demonstration Projects (NJ [2022]09).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. International Tea Market: Market Situation, Prospects and Emerging Issues. Available online: https://www.fao.org/3/cc0238en/cc0238en.pdf (accessed on 8 September 2023).
2. Xue, Z.; Xu, R.; Bai, D.; Lin, H. YOLO-tea: A tea disease detection model improved by YOLOv5. *Forests* **2023**, *14*, 415. [CrossRef]
3. Bao, W.; Fan, T.; Hu, G.; Liang, D.; Li, H. Detection and identification of tea leaf diseases based on AX-RetinaNet. *Sci. Rep.* **2022**, *12*, 2183. [CrossRef]
4. Chen, J.; Liu, Q.; Gao, L. Visual tea leaf disease recognition using a convolutional neural network model. *Symmetry* **2019**, *11*, 343. [CrossRef]
5. Hossain, M.S.; Mou, R.M.; Hasan, M.M.; Chakraborty, S.; Razzak, M.A. Recognition and Detection of Tea Leaf's Diseases Using Support Vector Machine. In Proceedings of the IEEE 14th International Colloquium on Signal Processing and Its Applications (CSPA), Penang, Malaysia, 9–10 March 2018; pp. 150–154.
6. Sun, Y.; Jiang, Z.; Zhang, L.; Dong, W.; Rao, Y. SLIC_SVM based leaf diseases saliency map extraction of tea plant. *Comput. Electron. Agric.* **2019**, *157*, 102–109. [CrossRef]
7. Zou, X.; Ren, Q.; Cao, H.; Qian, Y.; Zhang, S. Identification of Tea Diseases Based on Spectral Reflectance and Machine Learning. *J. Inf. Process. Syst.* **2020**, *16*, 435–446. [CrossRef]
8. Prabu, S.; Bapu, B.T.; Sridhar, S.; Nagaraju, V. Tea plant leaf disease identification using hybrid filter and support vector machine classifier technique. In *Recent Advances in Internet of Things and Machine Learning: Real-World Applications*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 117–128.
9. Liu, Z.; Bashir, R.N.; Iqbal, S.; Shahid, M.M.A.; Tausif, M.; Umer, Q. Internet of Things (IoT) and machine learning model of plant disease prediction–blister blight for tea plant. *IEEE Access* **2022**, *10*, 44934–44944. [CrossRef]
10. Hu, G.; Wu, H.; Zhang, Y.; Wan, M. A low shot learning method for tea leaf's disease identification. *Comput. Electron. Agric.* **2019**, *163*, 104852. [CrossRef]
11. Hu, G.; Yang, X.; Zhang, Y.; Wan, M. Identification of tea leaf diseases by using an improved deep convolutional neural network. *Sustain. Comput.-Inform. Syst.* **2019**, *24*, 100353. [CrossRef]

12. Bhowmik, S.; Talukdar, A.K.; Sarma, K.K. Detection of Disease in Tea Leaves Using Convolution Neural Network. In Proceedings of the Advanced Communication Technologies and Signal Processing (IEEE ACTS), Silchar, India, 4–6 December 2020.
13. Lee, S.-H.; Lin, S.-R.; Chen, S.-F. Identification of tea foliar diseases and pest damage under practical field conditions using a convolutional neural network. *Plant Pathol.* **2020**, *69*, 1731–1739. [CrossRef]
14. Hu, G.; Wang, H.; Zhang, Y.; Wan, M. Detection and severity analysis of tea leaf blight based on deep learning. *Comput. Electr. Eng.* **2021**, *90*, 107023. [CrossRef]
15. Li, H.; Shi, H.; Du, A.; Mao, Y.; Fan, K.; Wang, Y.; Shen, Y.; Wang, S.; Xu, X.; Tian, L.; et al. Symptom recognition of disease and insect damage based on Mask R-CNN, wavelet transform, and F-RNet. *Front. Plant Sci.* **2022**, *13*, 922797. [CrossRef] [PubMed]
16. Yang, J.; Guo, X.; Li, Y.; Marinello, F.; Ercisli, S.; Zhang, Z. A survey of few-shot learning in smart agriculture: Developments, applications, and challenges. *Plant Methods* **2022**, *18*, 28. [CrossRef] [PubMed]
17. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
18. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
19. Syed, D.; Zainab, A.; Refaat, S.S.; Abu-Rub, H.; Bouhali, O.; Ghrayeb, A.; Houchati, M.; Bañales, S. Inductive Transfer and Deep Neural Network Learning-Based Cross-Model Method for Short-Term Load Forecasting in Smarts Grids Méthode de modèle croisé basée sur le transfert inductif et l'apprentissage par réseau neuronal profond pour la prévision de la charge à court terme dans les réseaux intelligents. *IEEE Can. J. Electr. Comput. Eng.* **2023**, *46*, 157–169.
20. Jiang, Z.; Dong, Z.; Jiang, W.; Yang, Y. Recognition of rice leaf diseases and wheat leaf diseases based on multi-task deep transfer learning. *Comput. Electron. Agric.* **2021**, *186*, 106184. [CrossRef]
21. Paymode, A.S.; Malode, V.B. Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG. *Artif. Intell. Agric.* **2022**, *6*, 23–33. [CrossRef]
22. Zhang, W.; Deng, L.; Zhang, L.; Wu, D. A Survey on Negative Transfer. *IEEE-CAA J. Autom. Sin.* **2023**, *10*, 305–329. [CrossRef]
23. 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. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
24. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef]
25. Feng, Y.; Wang, X.; Xin, Y.; Zhang, B.; Liu, J.; Mao, M.; Xu, S.; Zhang, B.; Han, S. Effective feature enhancement and model ensemble strategies in tiny object detection. In Proceedings of the Computer Vision–ECCV 2020 Workshops, Glasgow, UK, 23–28 August 2020; Proceedings, Part V 16. Springer: Berlin/Heidelberg, Germany, 2020; pp. 324–330.
26. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
27. Song, G.; Liu, Y.; Wang, X. Revisiting the sibling head in object detector. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11563–11572.
28. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
29. Zhuang, J.; Qin, Z.; Yu, H.; Chen, X. Task-Specific Context Decoupling for Object Detection. *arXiv* **2023**, arXiv:2303.01047.
30. Misra, D.; Nalamada, T.; Arasanipalai, A.U.; Hou, Q. Rotate to Attend: Convolutional Triplet Attention Module. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 5–9 January 2021; pp. 3138–3147.
31. Xu, C.; Wang, J.; Yang, W.; Yu, H.; Yu, L.; Xia, G.-S. RFLA: Gaussian receptive field based label assignment for tiny object detection. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 526–543.
32. Wang, J.; Xu, C.; Yang, W.; Yu, L. A normalized Gaussian Wasserstein distance for tiny object detection. *arXiv* **2021**, arXiv:2110.13389.
33. Cao, Z.; Long, M.; Wang, J.; Jordan, M.I. Partial Transfer Learning with Selective Adversarial Networks. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 2724–2732.
34. Wu, X.; Deng, H.; Wang, Q.; Lei, L.; Gao, Y.; Hao, G. Meta-learning shows great potential in plant disease recognition under few available samples. *Plant J.* **2023**, *114*, 767–782. [CrossRef]
35. Kisantal, M.; Wojna, Z.; Murawski, J.; Naruniec, J.; Cho, K. Augmentation for small object detection. *arXiv* **2019**, arXiv:1902.07296.
36. Lin, H.; Lin, J.; Wang, F. An innovative machine learning model for supply chain management. *J. Innov. Knowl.* **2022**, *7*, 100276. [CrossRef]
37. Hu, G.; Ye, R.; Wan, M.; Bao, W.; Zhang, Y.; Zeng, W. Detection of Tea Leaf Blight in Low-Resolution UAV Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *62*, 1–18. [CrossRef]
38. Lin, H.; Qian, J.; Di, B. Learning for Adaptive Multi-Copy Relaying in Vehicular Delay Tolerant Network. *IEEE Trans. Intell. Transp. Syst.* **2023**, 1–10. [CrossRef]
39. Yang, N.; Yuan, M.; Wang, P.; Zhang, R.; Sun, J.; Mao, H. Tea diseases detection based on fast infrared thermal image processing technology. *J. Sci. Food Agric.* **2019**, *99*, 3459–3466. [CrossRef]
40. Yuan, L.; Yan, P.; Han, W.; Huang, Y.; Wang, B.; Zhang, J.; Zhang, H.; Bao, Z. Detection of anthracnose in tea plants based on hyperspectral imaging. *Comput. Electron. Agric.* **2019**, *167*, 105039. [CrossRef]

41. Zhao, X.; Zhang, J.; Huang, Y.; Tian, Y.; Yuan, L. Detection and discrimination of disease and insect stress of tea plants using hyperspectral imaging combined with wavelet analysis. *Comput. Electron. Agric.* **2022**, *193*, 106717. [CrossRef]
42. Li, S.; Wang, S.; Wang, P. A small object detection algorithm for traffic signs based on improved YOLOv7. *Sensors* **2023**, *23*, 7145. [CrossRef]
43. Ma, L.; Zhao, L.; Wang, Z.; Zhang, J.; Chen, G. Detection and Counting of Small Target Apples under Complicated Environments by Using Improved YOLOv7-tiny. *Agronomy* **2023**, *13*, 1419. [CrossRef]
44. Zainab, A.; Syed, D. Deployment of deep learning models on resource-deficient devices for object detection. In Proceedings of the 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), Doha, Qatar, 2–5 February 2020; pp. 73–78.
45. Cao, J.; Bao, W.; Shang, H.; Yuan, M.; Cheng, Q. GCL-YOLO: A GhostConv-based lightweight yolo network for UAV small object detection. *Remote Sens.* **2023**, *15*, 4932. [CrossRef]
46. Yan, B.; Fan, P.; Lei, X.; Liu, Z.; Yang, F. A real-time apple targets detection method for picking robot based on improved YOLOv5. *Remote Sens.* **2021**, *13*, 1619. [CrossRef]
47. Lin, H.; Han, Y.; Cai, W.; Jin, B. Traffic Signal Optimization Based on Fuzzy Control and Differential Evolution Algorithm. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 8555–8566. [CrossRef]
48. Xie, S.; Wang, C.; Wang, C.; Lin, Y.; Dong, X. Online Identification Method of Tea Diseases in Complex Natural Environments. *IEEE Open J. Comput. Soc.* **2023**, *4*, 62–71. [CrossRef]