agriculture

MDPI

Article

PL-DINO: An Improved Transformer-Based Method for Plant Leaf Disease Detection

Wei Li [1,*], Lizhou Zhu [2] and Jun Liu [3]

[1] School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China
[2] College of Software Engineering, Southeast University, Suzhou 215123, China
[3] Institute of Agricultural Facilities and Equipment, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China
* Correspondence: li-wei@seu.edu.cn

Abstract: Agriculture is important for ecology. The early detection and treatment of agricultural crop diseases are meaningful and challenging tasks in agriculture. Currently, the identification of plant diseases relies on manual detection, which has the disadvantages of long operation time and low efficiency, ultimately impacting the crop yield and quality. To overcome these disadvantages, we propose a new object detection method named "Plant Leaf Detection transformer with Improved deNoising anchOr boxes (PL-DINO)". This method incorporates a Convolutional Block Attention Module (CBAM) into the ResNet50 backbone network. With the assistance of the CBAM block, the representative features can be effectively extracted from leaf images. Next, an EQualization Loss (EQL) is employed to address the problem of class imbalance in the relevant datasets. The proposed PL-DINO is evaluated using the publicly available PlantDoc dataset. Experimental results demonstrate the superiority of PL-DINO over the related advanced approaches. Specifically, PL-DINO achieves a mean average precision of 70.3%, surpassing conventional object detection algorithms such as Faster R-CNN and YOLOv7 for leaf disease detection in natural environments. In brief, PL-DINO offers a practical technology for smart agriculture and ecological monitoring.

Keywords: leaf disease detection; PL-DINO; convolutional block attention module; equalization loss; crop

Citation: Li, W.; Zhu, L.; Liu, J. PL-DINO: An Improved Transformer-Based Method for Plant Leaf Disease Detection. *Agriculture* **2024**, *14*, 691. https://doi.org/10.3390/agriculture14050691

Academic Editor: Francesco Marinello

Received: 4 March 2024
Revised: 22 April 2024
Accepted: 24 April 2024
Published: 28 April 2024

1. Introduction

In natural environments, crops such as tomatoes, potatoes, and rice are vulnerable to various diseases. Disease detection and prevention play a crucial role in minimizing economic loss and continuously enhancing crop yield and quality [1–3]. Traditionally, the identification of crop diseases is heavily reliant on manual detection and expert experiences. However, this paradigm has drawbacks such as subjectivity, low accuracy, inefficiency, and high cost [4].

With the technological revolution in artificial intelligence, the use of intelligent technology in agriculture has garnered increasing attention [5]. Many image classification methods have been proposed and implemented in agricultural practice, such as crop monitoring, pest and disease detection, and robotic operation. As plant leaves straightforwardly reflect plant species, growth status, and health conditions, detecting and identifying plant leaves hold significant value for plant science research, crop management, and the control of pests and diseases [6–8].

In agriculture, the procedures for plant leaf disease detection include extracting features from images and learning the classifier for disease detection. The related methods can further be categorized into the following: (1) two-stage algorithms, such as Faster Region-based Convolutional Neural Network (Faster R-CNN) [9] and Cascade R-CNN [10]; (2) one-stage algorithms, such as You Only Look Once version 5 (YOLOv5) [11] and Single Shot multibox Detector (SSD) [12]. Generally, two-stage detection algorithms employ a

Region Proposal Network to generate candidate regions. One-stage detection algorithms transform the detection tasks into end-to-end regression tasks. One-stage algorithms have simple structures and, thus, take relatively small inference time. By contrast, two-stage algorithms have large parameter amounts and computational overheads, but they can usually obtain higher accuracy than one-stage algorithms [13].

In recent years, deep learning models have been widely used to detect plant leaf diseases. Jiang et al. [14] constructed an apple leaf disease dataset and developed a real-time SSD that contains a module of rainbow concatenation to enhance the performance of small object recognition. Liu et al. [15] presented an image pyramid to enhance the performance of a tomato disease detection scheme based on YOLOv3. Li et al. [16] improved the YOLOv5 algorithm, incorporating the modules Cross Stage Partial Network, Feature Pyramid Network, and Non-Maximum Suppression for vegetable disease detection. These methods are efficient and lightweight for plant disease detection at the sacrifice of efficacy. Qi et al. [17] constructed a SE-YOLOv5 network to learn the features for tomato virus disease detection in natural backgrounds. Zhu et al. [18] designed YOLOv5-based Apple-Net, which adds the coordinate attention module after creating a cross stage partial structure to strengthen feature competence and uses the feature enhancement module in the network neck to further improve the capability of learned features. These works rely on the attention mechanism to enhance the learned feature discriminability, but they are incompetent for very complex leaf disease images. Zhang et al. [19] devised a Faster R-CNN algorithm with multi-feature fusion that can identify leaf occlusions for soybean leaf disease detection. Wang et al. [20] conceived a Faster R-CNN technique to reduce the missing and incorrect detection results of densely distributed sweet potato leaves. Zhou et al. [21] integrated Faster R-CNN with K-Means clustering to overcome the impact of image noise interference on rice disease detection. Zhang et al. [22] developed a rice spike detection Faster R-CNN with four optimization strategies to identify multiple developmental stages of rice spikes. Pan et al. [23] applied Faster R-CNN to accurately locate strawberry leaf scorch disease whilst depressing the noisy interference of a complex background. These efforts focus on addressing problems such as leaf occlusion and noise interference but neglect the impact of data distribution on classification. Nowadays, transformer-based detectors have gained increasing attention due to their remarkable detection performances in various tasks. Among them, DEtection TRansformer (DETR) and its variant algorithms have achieved inspiring performances for different object detection tasks. Zhang et al. [24] improved DETR by fusing multi-scale features to cope with the problems of leaf overlap, multiple disease types, and complex backgrounds for paddy disease detection. Dananjayan et al. [25] utilized Deformable DETR with ResNet50 to deal with similar problems in citrus leaf disease detection.

It is challenging to classify visual data with a long-tailed distribution. Generally, long-tailed class imbalance easily causes the classification model to overemphasize the head classes with a large amount of data in training, whilst overlooking the tail classes with a limited sample size. Existing methods to address this problem can mainly be categorized into three types: class re-balancing, information augmentation, and module enhancement [26]. Among these methods, class re-balancing is the most capable and practical [26]. Class re-balancing methods can further be divided into re-sampling and re-weighting.

Re-sampling models aim to balance the probabilities of selected classes by means of random over-sampling or under-sampling. Kang et al. [27] put forward a decoupled learning network by applying four sampling strategies and four types of classifiers. This network decouples representation and classification for handling the imbalanced classification problem of long-tailed data. Zhou et al. [28] brought forward a unified Bilateral Branch Network, which jointly performs feature learning and classifier learning, to learn universal patterns from original data distributions and simultaneously attribute importance to the tail data in an adaptive manner. However, the performance of re-sampling models on head classes may degrade due to under-sampling. On the other hand, these model may overfit the tail classes due to over-sampling as well.

Re-weighting balances the classes by assigning suitable weights to different classes. Lin et al. [29] utilized focal loss, which increases the weights of hard-to-classify samples, to handle the problem of foreground–background class imbalance. Nevertheless, focal loss seems incompetent in addressing the imbalance among foreground classes in long-tailed scenarios [30]. Cui et al. [31] presented a framework with class-balanced loss, which introduces a weighting factor that is inversely proportional to the effective number of samples per class. However, this method fails to effectively use the distribution information of sample data. Cao et al. [32] introduced a Label-Distribution-Aware Margin (LDAM) loss, which recommends a larger margin for the minority classes than the majority ones, to improve method generalization performance. However, this method requires the combination of LADM loss and a deferred re-balancing training procedure, which largely increases the inference cost.

While past research has made some progress, detecting plant leaf diseases in complex agricultural environments still remains challenging. First, images captured in real-world scenarios usually contain complicated foreground variations and noisy background clutter, which bring great difficulties for feature extraction. Second, different categories of plant leaf diseases are highly similar. Third, plant leaves with different disease classes frequently suffer from the severe problem of class imbalance. These problems are entangled with each other, forming a bottleneck for the issue of plant leaf disease detection. Nevertheless, nearly no studies have handled this bottleneck adequately in the literature.

To fill this research gap, we propose a novel transformer-based model named "Plant Leaf Detection transformer with Improved deNoising anchOr boxes (PL-DINO)". PL-DINO introduces a Convolutional Block Attention Module (CBAM) into the backbone network of DINO to learn the channel-wise and global image features. In addition, PL-DINO adopts a re-weighting loss function to balance the weights of different classes in classification. The main contributions of this work are summarized as follows.

(1) We design a new leaf disease detection method, PL-DINO, based on the attention mechanism. By using the CBAM, PL-DINO can efficiently exploit both channel-wise and spatial features that discriminate leaf diseases.

(2) In PL-DINO, we take advantage of the loss function EQualization Loss (EQL) based on re-weighting to address the problem of imbalanced classification. EQL can mitigate the severe suppression of head category samples on the tail category ones in training, thereby effectively resolving the long-tailed distribution problem.

(3) The proposed PL-DINO is evaluated using the public benchmark dataset PlantDoc. The experimental results demonstrate the superiority of our method over the related state-of-the-art approaches for leaf disease detection in the complex scenarios of natural environments.

2. Materials and Methods

2.1. Datasets

The PlantDoc dataset [33] is broadly used for visual plant leaf disease detection. This dataset contains 2598 images of 13 plant species belonging to 27 classes: 10 healthy classes and 17 disease classes. The images are taken in the natural environments, suffering from complex foreground variations, noisy background clutters, and severe class imbalance. In PlantDoc, the training set contains 2120 images, and each testing and validation set contains 239 images. The dataset statistics are supplied in Figure 1. Additionally, Figure 2 shows some leaf image examples from the PlantDoc dataset.

The Microsoft Common Objects in Context (MS COCO) dataset [34] is a widely used benchmark dataset for object detection. MS COCO contains 328,000 images from 82 common object classes in the scenes of daily life, such as people, animals, vehicles, plants, and electronic devices.

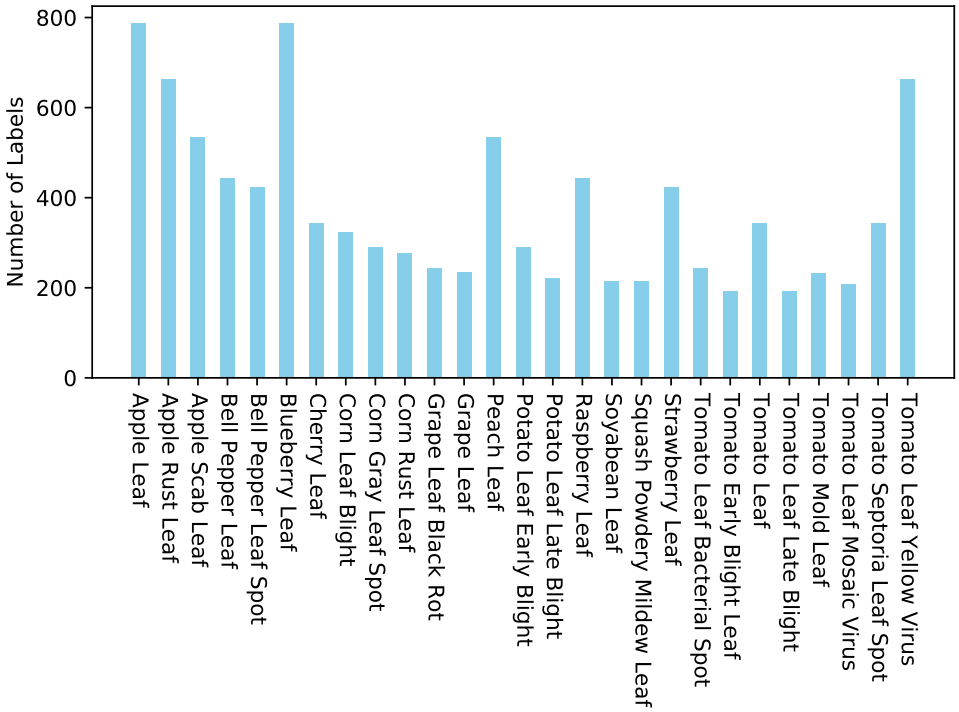


Figure 1. Statistics of PlantDoc dataset.

(**a**) Apple Scab Leaf

(**b**) Tomato Septoria Leaf Spot

(**c**) Corn Leaf Blight

(**d**) Potato Leaf Late Blight

Figure 2. Leaf image examples from PlantDoc dataset.

2.2. DETR with Improved De-Noising Anchor Boxes

DETR is an end-to-end object detection model based on a transformer [35]. More specifically, DETR converts the object detection problem into a prediction task [36].

For object detection models like DETR, the backbone typically refers to the basic CNN structure used for image feature extraction. DETR transforms the feature map output from the backbone into one-dimensional sequential data. Before the sequential data pass into the encoder of DETR, the sequential data are added with positional encoding. Then, DETR

utilizes the attention mechanism to process the sequential data and positional information. For training, DETR resorts to a specialized loss function named "Hungarian loss" [35].

DINO contains a CNN backbone, a multi-layer transformer encoder, a multi-layer transformer decoder, and multiple prediction heads [37]. The main differences between DINO and DETR lie in the contrastive de-noising training, the mixed query selection technique and the "look forward twice" scheme. The scheme of plant leaf disease detection based on DINO is illustrated in Figure 3.

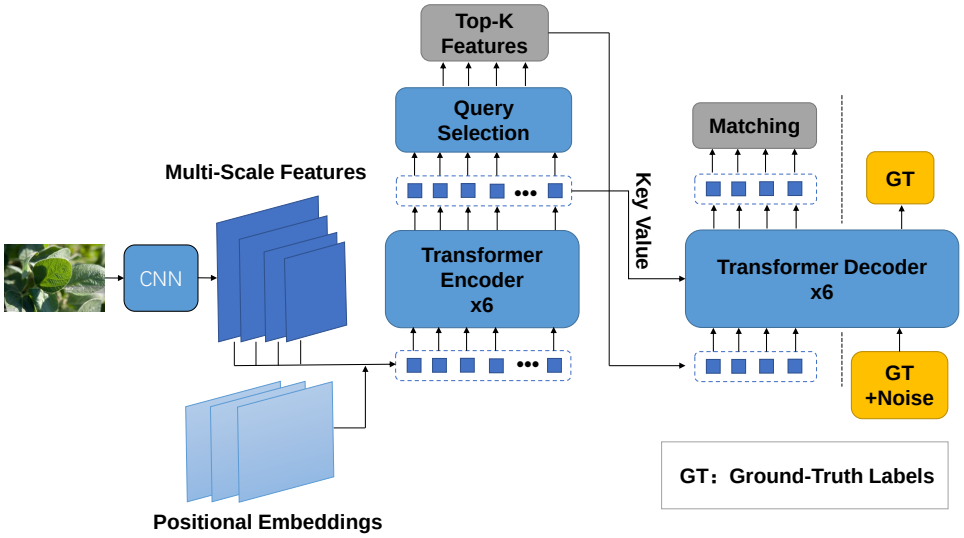


Figure 3. Scheme of DINO.

2.3. Framework of PL-DINO

Our proposed method, PL-DINO, integrates the CBAM [38] into the backbone network of DINO to improve the discriminability of leaf disease features. Specifically, the backbone network consists of a Residual Network (ResNet50) with four convolutional layers, denoted as Layer 1, Layer 2, Layer 3, and Layer 4. In PL-DINO, we place the CBAM module after Layer 4 to extract more discriminative leaf disease features. At the same time, we adopt EQL, a category-balanced loss, to mitigate the suppression of head categories on tail ones [39]. The overall framework of PL-DINO is exhibited in Figure 4.

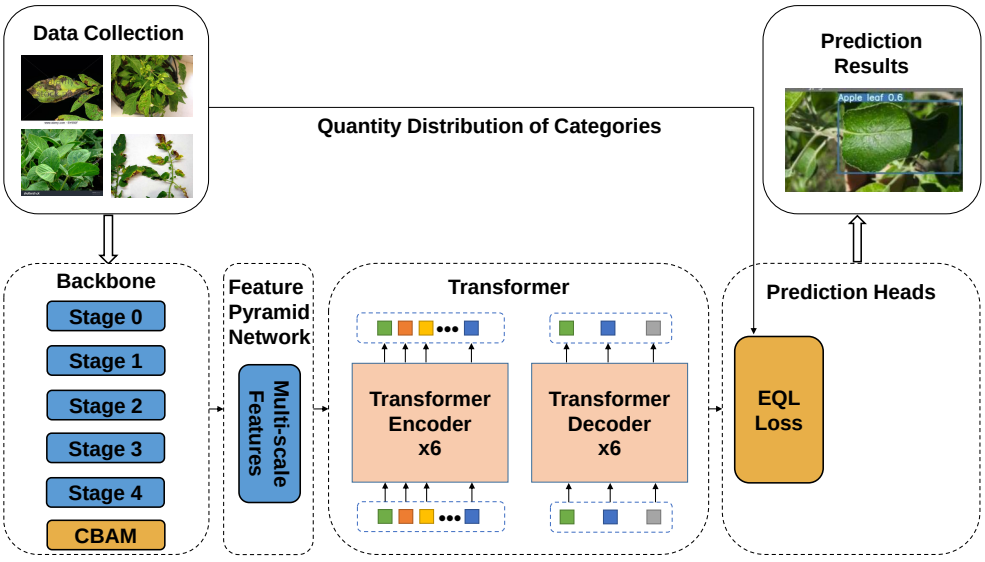


Figure 4. Overall framework of PL-DINO.

2.3.1. Convolutional Block Attention Module

CBAM [38] comprises two key components: the Channel Attention Module (CAM) and the Spatial Attention Module (SAM). CAM adaptively refines the channel-wise features by exploiting the interdependencies among different channels of a feature map. SAM refines the spatial features in the informative spatial regions of the target object.

By combining both channel-wise and spatial attention mechanisms, CBAM enables CNNs to better capture the discriminative features. The structure of CBAM is exhibited in Figure 5.

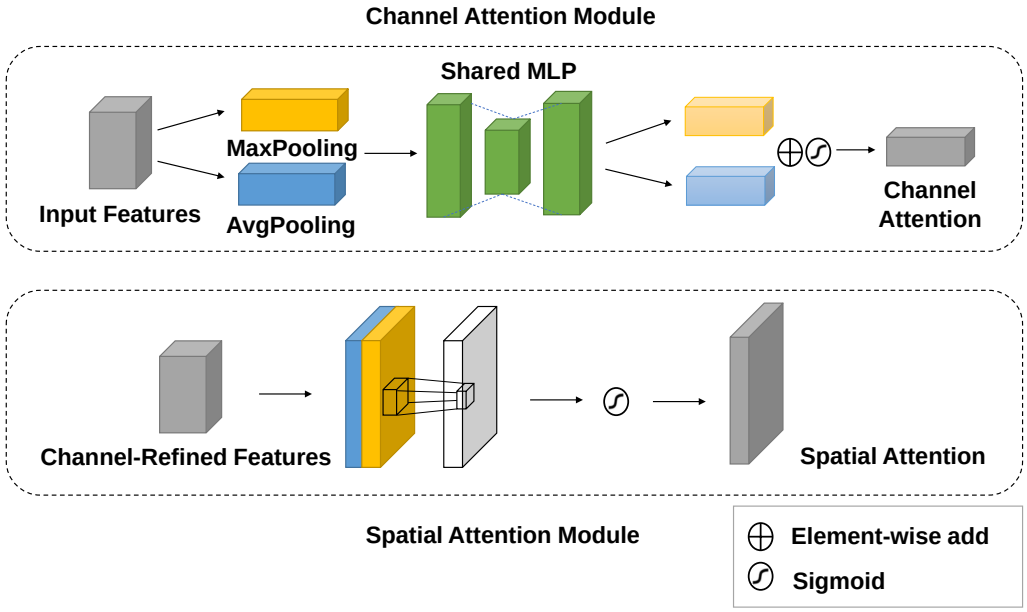


Figure 5. Channel attention and spatial attention components of CBAM.

The Squeeze-and-Excitation (SE) module takes the two critical steps to refine the feature map within CNNs [40]. First, the "squeeze" step compresses the channel-wise information through global average pooling. The shape of the input feature map is squeezed from $W \times H \times C$ into $1 \times 1 \times C$, where W denotes the width, H denotes the height, and C denotes the channel number of the feature map. Second, in the "Excitation" step, Multi-Layer Perceptron (MLP) explores the interdependencies among these channels and generates the channel-specific weights. Subsequently, the feature map is improved by these weights. The structure of the SE module is displayed in Figure 6.

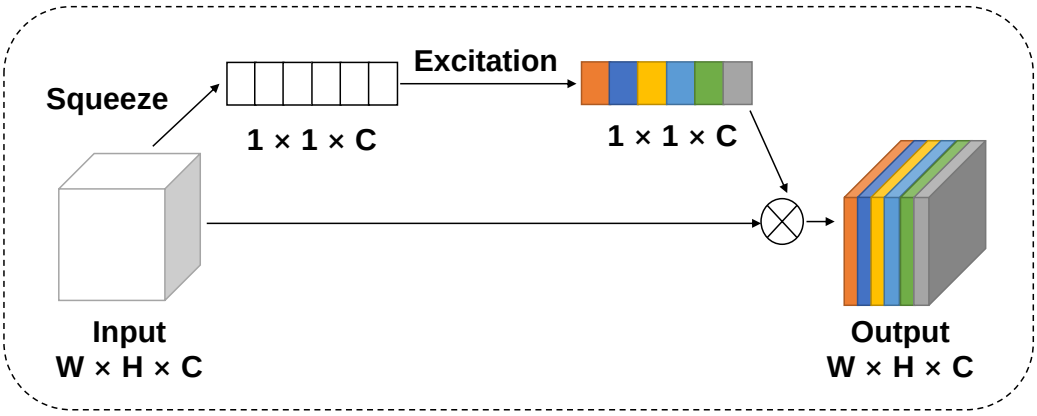


Figure 6. Squeeze-and-Excitation module.

For the SE module, the global average pooling in the channel direction inevitably fails to explore the spatial information. To account for this gap, we integrate CBAM into the

PL-DINO architecture. In comparison to the SE module, CBAM is beneficial for PL-DINO to learn both channel-wise and spatial features.

2.3.2. Loss Function for Imbalanced Classification

Data samples collected in natural environments often exhibit a long-tailed distribution. The long-tailed distribution problem presents a big challenge for plant leaf disease detection models, because the information of minority classes can easily be overwhelmed by the majority ones, resulting in the decline of overall classification performance. In the literature, two fundamental methods are commonly used to address this problem: re-sampling and re-weighting. In essence, re-sampling balances the class distribution by adaptively adjusting the sampling frequency according to the class size. However, re-sampling has weakness in repeated learning on tail classes and insufficient learning on the majority ones. In contrast, re-weighting can counteract the adverse effect of a long-tailed distribution by assigning lower weights to the head classes and higher weights to the tail ones. Hence, in PL-DINO, we suggest adopting Binary Cross-Entropy (BCE) as the loss function:

$$L_{\mathrm{BCE}} = -\sum_{j=1}^{C} y_j \log \hat{p}_j \tag{1}$$

where

$$\hat{p}_j = \begin{cases} p_j & \text{if } y_j = 1 \\ 1 - p_j & \text{otherwise} \end{cases} \tag{2}$$

$$y_j = \begin{cases} 1 & \text{if } j = c \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where p_j denotes the sigmoid of network output, y_j denotes the ground-truth label, and C denotes the number of classes. Only when j is the ground-truth class c is y_j equal to 1; otherwise, it is equal to 0.

Based on BCE, EQL is formulated as follows:

$$L_{\mathrm{EQL}} = -\sum_{j=1}^{C} w_j \log \hat{p}_j \tag{4}$$

$$w_j = 1 - E(r)T_\lambda(f_j)(1 - y_j) \tag{5}$$

where

$$E(r) = \begin{cases} 1 & \text{foreground} \\ 0 & \text{background} \end{cases} \tag{6}$$

$$T_\lambda(f_j) = \begin{cases} 1 - f_j/\lambda & \text{if } f_j < \lambda \\ 0 & \text{if } f_j > \lambda \end{cases} \tag{7}$$

where f_j denotes the proportion of the size of class j to the sample size of the whole dataset, λ denotes the threshold to distinguish the tail classes from the head ones, w_j denotes the weight of each term, and $E(r)$ is used to distinguish whether r is the background or not.

To address the class imbalance problem, PL-DINO resorts to the EQL loss function. This loss function can mitigate the influence of negative samples on the minority classes, thereby enabling the model to treat all classes in a relatively fair manner.

3. Results

3.1. Experimental Settings

We conduct experiments on a PC with an Intel Core i7 CPU, NVIDIA RTX 2080Ti, 12 GB GPU, and 16 GB RAM. We implement PL-DINO using the MMDetection Toolbox based on the PyTorch deep learning framework in the Python 3.8 environment. The proposed model is pre-trained using all data samples from MS COCO and fine-tuned and tested using PlantDoc. Note that the training and testing data are totally independent.

We carry out five-fold cross-validation for method evaluation. The methods are evaluated using the following metrics: mAP@0.5, Precision, Recall, and F1-score. Here, mAP@0.5 stands for the mean average precision calculated at the intersection over the union threshold of 0.5.

The hyperparameters for training PL-DINO are provided as follows. Learning rate: 0.0001; weight decay: 0.0001; batch size: 2; epoch number: 24; optimizer: AdamW; λ: 0.03; learning rate decay: 0.1. Specifically, we suggest performing learning rate decay at epochs 11 and 23.

The loss-epoch curve and mAP-epoch curve of PL-DINO are illustrated in Figure 7 and Figure 8, respectively. From these two figures, we can observe that both the training loss and mAP of PL-DINO gradually converge to a stable value in the end, which confirms the efficacy of the training process of our method.

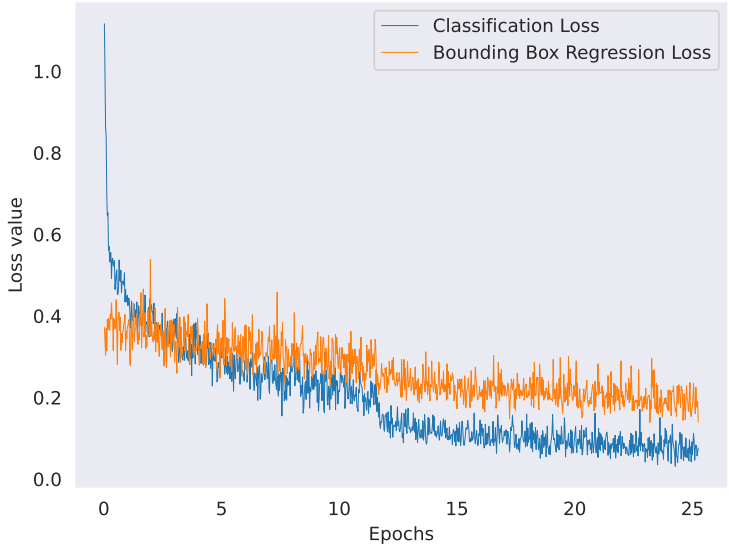


Figure 7. Training loss of PL-DINO.

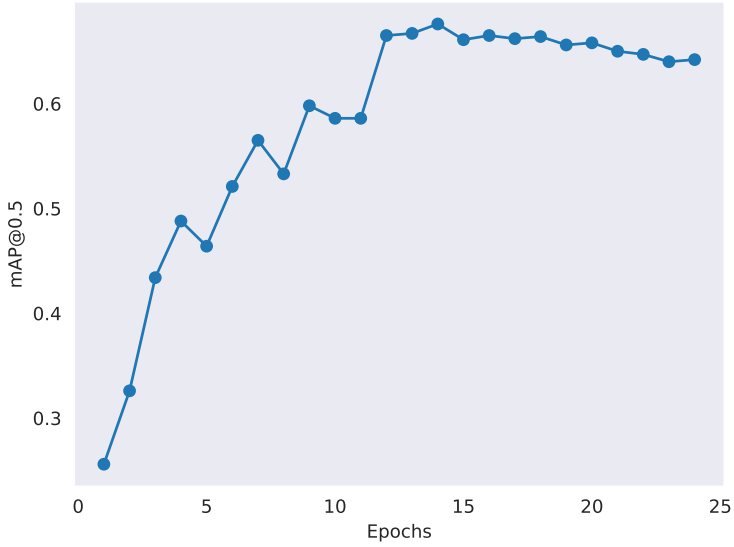


Figure 8. Training mAP of PL-DINO.

3.2. Method Comparison

We compare PL-DINO with the related state-of-the-art approaches, including Faster R-CNN [9], YOLOv5s [16], YOLOv5m [17], YOLOv7 [41], and DETR [35]. The experimental results are reported in Table 1. The results show that our proposed method achieves the

best performance overall. In comparison to the compared approaches, the globally oriented attention mechanism is one primary contributing factor for the high detection rate achieved by our method. The attention mechanism can help our method explore both the global semantic and positional information in an image and capture the correlation information between these informational aspects, thus forming the discriminative representation of the entire plant leaf image. For intuitiveness, the Precision–Recall curve of PL-DINO is visualized in Figure 9.

Table 1. Comparison of PL-DINO with related state-of-the-art approaches.

Methods	Precision (%)	Recall (%)	F1-Score (%)	mAP@0.5 (%)
Faster R-CNN	38.0	61.4	42.3	45.6
YOLOv5s	50.2	57.6	53.5	52.6
YOLOv5m	56.6	60.8	56.1	55.4
YOLOv7	57.9	69.7	62.0	67.0
DETR	40.7	67.1	46.3	48.9
PL-DINO	**62.9**	**75.0**	**63.2**	**70.3**

Note: the highest performance under each evaluation metric is highlighted in bold.

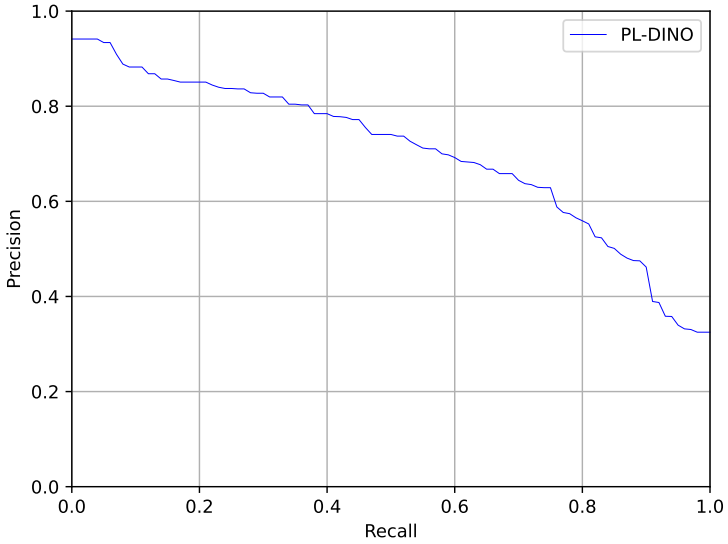


Figure 9. Precision–Recall curve of PL-DINO.

3.3. Model Ablation

We further carry out an ablation study on the modules of CBAM and EQL in PL-DINO. The ablation results have been reported in Table 2. From the table, we can find that, with the CBAM module, the mAP@0.5 of DINO has climbed by 1.0%, and with the EQL loss, the mAP@0.5 of DINO has risen by 1.4%. Hence, these two components indeed play important roles in our method.

Table 2. Ablation of CBAM and EQL modules in PL-DINO.

Models	mAP@0.5 (%)	Precision (%)	Recall (%)
Baseline	67.2	63.4	64.0
Baseline + CBAM	68.2	63.1	69.0
Baseline + EQL	68.6	**65.6**	64.0
Baseline + CBAM + EQL	**70.3**	62.9	**75.0**

Note: the highest performance under each evaluation metric is highlighted in bold.

Moreover, we evaluate different combinations of attention modules and backbones in PL-DINO. The evaluated attention modules include SE [40], Efficient Channel Attention (ECA) [42], and CBAM [38]; the evaluated backbones include ResNet50 [43] and ResNeSt50 [44].

As exhibited in Table 3, when the backbone network is ResNet50, the performances of DINO with each attention module are slightly superior or equivalent to ResNet50, and among all the attention modules, the CBAM module brings the greatest performance improvement to DINO.

Table 3. Ablation of backbone networks and attention modules in DINO.

Backbone Network	Attention Module	mAP@0.5 (%)	Precision (%)	Recall (%)
ResNet50	-	67.2	**63.4**	64.0
ResNet50	SE	66.7	62.4	63.0
ResNet50	ECA	66.1	59.1	69.0
ResNet50	CBAM	**68.2**	63.1	**69.0**
ResNeSt50	-	64.3	60.3	66.0
ResNeSt50	SE	65.1	64.2	63.0
ResNeSt50	ECA	63.8	61.5	62.0
ResNeSt50	CBAM	64.9	59.1	68.0

Note: the highest performance under each evaluation metric is highlighted in bold.

3.4. Computational Expense

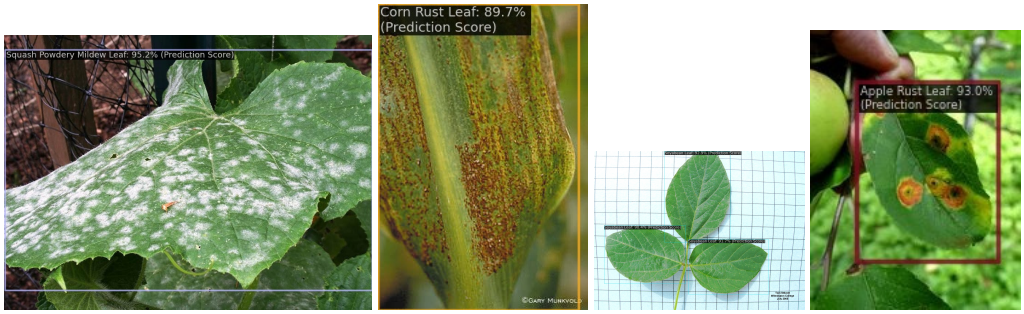
The training durations of PL-DINO, Faster R-CNN, YOLOv5, YOLOv7, and DETR are recorded in Table 4. From the table, we can see that PL-DINO takes the second shortest training time. However, its detection performance is significantly higher than Faster R-CNN, according to Table 1. Therefore, we can agree that PL-DINO performs the best among all the evaluated methods if considering aspects of both effectiveness and efficiency, both of which are important and indispensable for practice.

Table 4. Method training durations.

Models	Time (h:min:s)
Faster R-CNN	4:40:14
YOLOv5	9:03:25
YOLOv7	8:46:33
DETR	12:10:47
PL-DINO	5:43:05

3.5. Result Visualization

Some results of PL-DINO are visualized in Figure 10. From the figure, we can see that the leaves with diseases can be accurately detected with high confidence scores. These results intuitively demonstrate that PL-DINO is not only robust to foreground variations and background noises but also capable of handling the imbalanced class distribution in plant leaf disease detection.



(**a**) Squash Powdery Mildew Leaf (**b**) Corn Rust Leaf (**c**) Soy Bean Leaf (**d**) Apple Rust Leaf

Figure 10. Examples of detection results by PL-DINO.

Additionally, Figure 11 presents a normalized confusion matrix of PL-DINO's classification rates for all leaf disease classes. From the confusion matrix, we can find that, as expected, PL-DINO obtains the highest classification rates for the majority classes of blueberry leaf, peach leaf, raspberry leaf, strawberry leaf, and tomato leaf yellow virus. Even so, PL-DINO also acquires not-so-bad classification rates for the minority classes of apple rust leaf, corn gray leaf spot, corn rust leaf, grape leaf black rot, and grape leaf. These results straightforwardly show the strong ability of the method to handle the class imbalance problem.

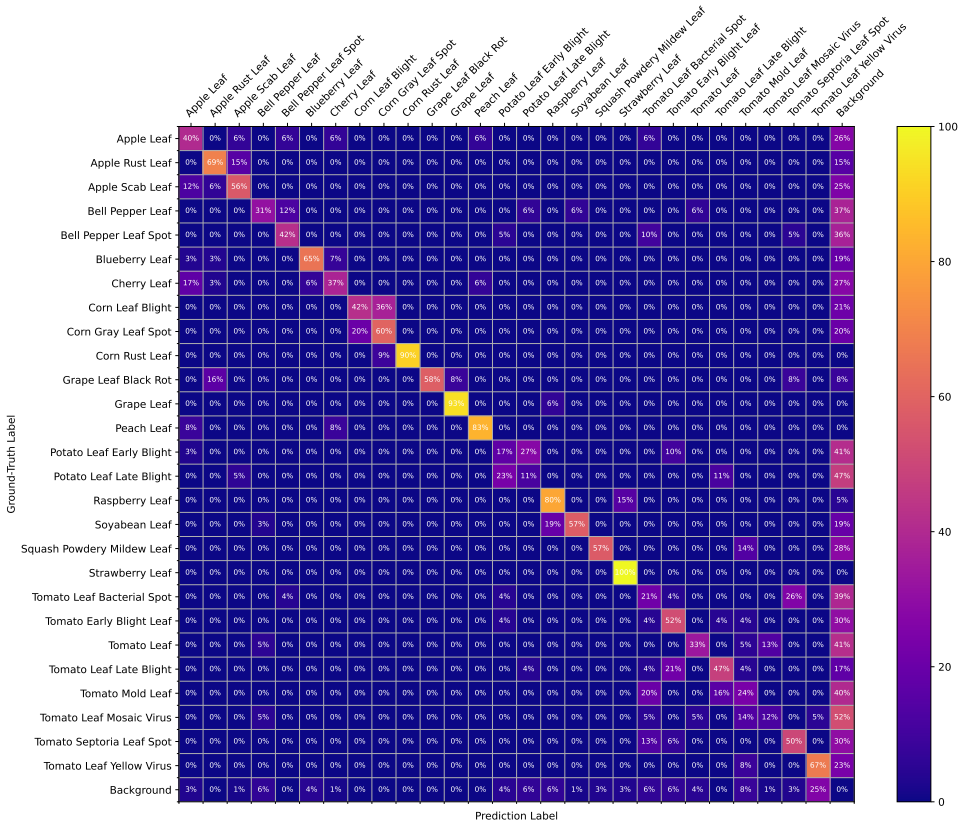


Figure 11. Normalized confusion matrix of classification rates of PL-DINO for all leaf disease classes.

4. Discussion

Detecting plant leaf diseases in natural environments is fundamental and important for smart agriculture and automated ecological monitoring systems. Plant leaf images in real-world scenarios undergo complicated foreground variations, noisy background clutters, and imbalanced class distribution. These obstacles pose significant challenges to plant leaf disease detection. To cope with these challenges, we have proposed a new method, PL-DINO. Compared to the traditional networks, Faster R-CNN, YOLOv5, and YOLOv7, even if our method PL-DINO does not contain any manually designed anchor and non-maximal suppression, it still can obtain higher precision and recall performances, benefiting from the long-range dependence information captured by the transformer structure. Furthermore, compared with DETR and DINO with a transformer structure, PL-DINO still performs better, which is attributed to its advantages from the CBAM and EQL loss modules. As experimentally demonstrated, PL-DINO behaves effectively in plant leaf disease detection and outperforms the related state-of-the-art approaches on widely-used benchmark datasets.

Although PL-DINO shows encouraging performance in this study, there are still some limitations left to be addressed. The attention mechanism adopted in this study places great emphasis on global image information, potentially resulting in a loss of local detailed information. So, our method may not be suitable for situations where the disease patterns are overly small and densely distributed in one single leaf image. In the future, we plan to

design a new multi-scale feature fusion network on the basis of PL-DINO for the purpose of jointly exploiting both global and local useful information from images for plant leaf disease detection. Additionally, as our method employs a transformer structure, the number of model parameters is substantial, potentially affecting its real-time applicability in agriculture. So, in the future, it will so be a valuable direction to develop a lightweight version of PL-DINO to reduce the model's complexity without compromising its performance.

5. Conclusions

In conclusion, PL-DINO offers an effective technology for smart agriculture and ecological monitoring. Its promising performance in leaf disease detection presents as a new opportunity for enhancing agricultural production and improving ecosystem health monitoring.

Author Contributions: Conceptualization, W.L.; Methodology, W.L. and L.Z.; Software, W.L. and L.Z.; Validation, W.L. and L.Z.; Formal analysis, W.L., L.Z. and J.L.; Investigation, W.L., L.Z. and J.L.; Resources, W.L. and J.L.; Data curation, W.L., L.Z. and J.L.; Writing—original draft, W.L. and L.Z.; Writing—review & editing, W.L. and L.Z.; Visualization, W.L. and L.Z.; Supervision, W.L.; Project administration, W.L.; Funding acquisition, W.L. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data are available from the corresponding author on reasonable request.

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

References

1. Xu, Y.; Li, J.; Wan, J. Agriculture and crop science in China: Innovation and sustainability. *Crop J.* **2017**, *5*, 95–99. [CrossRef]
2. Shill, A.; Rahman, M.A. Plant disease detection based on YOLOv3 and YOLOv4. In Proceedings of the International Conference on Automation, Control and Mechatronics for Industry 4.0, IEEE, Rajshahi, Bangladesh, 8–9 July 2021; pp. 1–6.
3. Atila, Ü; Uçar, M.; Akyol, K.; Uçar, E. Plant leaf disease classification using EfficientNet deep learning model. *Ecol. Inform.* **2021**, *61*, 101182 . [CrossRef]
4. Bai, Y.; Hou, F.; Fan, X.; Lin, W.; Lu, J.; Zhou, J.; Fan, D.; Li, L. A lightweight pest detection model for drones based on transformer and super-resolution sampling techniques. *Agriculture* **2023**, *13*, 1812. [CrossRef]
5. Yu, M.; Ma, X.; Guan, H. Recognition method of soybean leaf diseases using residual neural network based on transfer learning. *Ecol. Inform.* **2023**, *76*, 102096. [CrossRef]
6. Cheng, S.; Cheng, H.; Yang, R.; Zhou, J.; Li, Z.; Shi, B.; Lee, M.; Ma, Q. A high performance wheat disease detection based on position information. *Plants* **2023**, *12*, 1191. [CrossRef] [PubMed]
7. Liu, Y.; Liu, J.; Cheng, W.; Chen, Z.; Zhou, J.; Cheng, H.; Lv, C. A high-precision plant disease detection method based on a dynamic pruning gate friendly to low-computing platforms. *Plants* **2023**, *12*, 2073. [CrossRef] [PubMed]
8. Wang, D.; Wang, J.; Li, W.; Guan, P. T-CNN: Trilinear convolutional neural networks model for visual detection of plant diseases. *Comput. Electron. Agric.* **2021**, *190*, 106468. [CrossRef]
9. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
10. Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving into high quality object detection. In Proceedings of the Conference on Computer Vision and Pattern Recognition, IEEE, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162.
11. Zhu, X.; Lyu, S.; Wang, X.; Zhao, Q. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In Proceedings of the International Conference on Computer Vision Workshops, IEEE, Montreal, BC, Canada, 11–17 October 2021; pp. 2778–2788.
12. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In *Proceedings of the European Conference on Computer Vision*; Springer: Amsterdam, The Netherlands, 2016; pp. 21–37.
13. Li, W.; Zhu, T.; Li, X.; Dong, J.; Liu, J. Recommending advanced deep learning models for efficient insect pest detection. *Agriculture* **2022**, *12*, 1065. [CrossRef]
14. Jiang, P.; Chen, Y.; Liu, B.; He, D.; Liang, C. Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. *IEEE Access* **2019**, *7*, 59069–59080. [CrossRef]
15. Liu, J.; Wang, X. Tomato diseases and pests detection based on improved YOlO V3 convolutional neural network. *Front. Plant Sci.* **2020**, *11*, 898. [CrossRef] [PubMed]
16. Li, J.; Qiao, Y.; Liu, S.; Zhang, J.; Yang, Z.; Wang, M. An improved YOLOv5-based vegetable disease detection method. *Comput. Electron. Agric.* **2022**, *202*, 107345. [CrossRef]

17. Qi, J.; Liu, X.; Liu, K.; Xu, F.; Guo, H.; Tian, X.; Li, M.; Bao, Z.; Li, Y. An improved YOLOv5 model based on visual attention mechanism: Application to recognition of tomato virus disease. *Comput. Electron. Agric.* **2022**, *194*, 106780. [CrossRef]
18. Zhu, R.; Zou, H.; Li, Z.; Ni, R. Apple-Net: A model based on improved YOLOv5 to detect the apple leaf diseases. *Plants* **2023**, *12*, 169. [CrossRef] [PubMed]
19. Zhang, K.; Wu, Q.; Chen, Y. Detecting soybean leaf disease from synthetic image using multi-feature fusion faster R-CNN. *Comput. Electron. Agric.* **2021**, *183*, 106064. [CrossRef]
20. Wang, M.; Fu, B.; Fan, J.; Wang, Y.; Zhang, L.; Xia, C. Sweet potato leaf detection in a natural scene based on faster R-CNN with a visual attention mechanism and DIoU-NMS. *Ecol. Inform.* **2023**, *73*, 101931. [CrossRef]
21. Zhou, G.; Zhang, W.; Chen, A.; He, M.; Ma, X. Rapid detection of rice disease based on FCM-KM and faster R-CNN fusion. *IEEE Access* **2019**, *7*, 143190–143206. [CrossRef]
22. Zhang, Y.; Xiao, D.; Liu, Y.; Wu, H. An algorithm for automatic identification of multiple developmental stages of rice spikes based on improved Faster R-CNN. *Crop J.* **2022**, *10*, 1323–1333. [CrossRef]
23. Pan, J.; Xia, L.; Wu, Q.; Guo, Y.; Chen, Y.; Tian, X. Automatic strawberry leaf scorch severity estimation via faster R-CNN and few-shot learning. *Ecol. Inform.* **2022**, *70*, 101706. [CrossRef]
24. Zhang, X.; Dong, H.; Gong, L.; Cheng, X.; Ge, Z.; Guo, L. Multiple paddy disease recognition methods based on deformable transformer attention mechanism in complex scenarios. *Int. J. Comput. Appl.* **2023**, *45*, 660–672. [CrossRef]
25. Dananjayan, S.; Tang, Y.; Zhuang, J.; Hou, C.; Luo, S. Assessment of state-of-the-art deep learning based citrus disease detection techniques using annotated optical leaf images. *Comput. Electron. Agric.* **2022**, *193*, 106658. [CrossRef]
26. Zhang, Y.; Kang, B.; Hooi, B.; Yan, S.; Feng, J. Deep long-tailed learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 10795–10816. [CrossRef]
27. Kang, B.; Xie, S.; Rohrbach, M.; Yan, Z.; Gordo, A.; Feng, J.; Kalantidis, Y. Decoupling representation and classifier for long-tailed recognition. In Proceedings of the International Conference on Learning Representations, ICLR, Addis Ababa, Ethiopia, 26–30 April 2020; pp. 1–16.
28. Zhou, B.; Cui, Q.; Wei, X.S.; Chen, Z.M. BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition. In Proceedings of the Conference on Computer Vision and Pattern Recognition, IEEE, Seattle, WA, USA, 13–19 June 2020; pp. 9719–9728.
29. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the International Conference on Computer Vision, IEEE, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
30. Li, B.; Yao, Y.; Tan, J.; Zhang, G.; Yu, F.; Lu, J.; Luo, Y. Equalized focal loss for dense long-tailed object detection. In Proceedings of the Conference on Computer Vision and Pattern Recognition, IEEE, New Orleans, LA, USA, 18–24 June 2022; pp. 6990–6999.
31. Cui, Y.; Jia, M.; Lin, T.Y.; Song, Y.; Belongie, S. Class-balanced loss based on effective number of samples. In Proceedings of the Conference on Computer Vision and Pattern Recognition, IEEE, Long Beach, CA, USA, 15–20 June 2019; pp. 9268–9277.
32. Cao, K.; Wei, C.; Gaidon, A.; Arechiga, N.; Ma, T. Learning imbalanced datasets with label-distribution-aware margin loss. In Proceedings of the Advances in Neural Information Processing Systems, NeurIPS, Vancouver, BC, Canada, 8–14 December 2019; pp. 1567–1578.
33. Singh, D.; Jain, N.; Jain, P.; Kayal, P.; Kumawat, S.; Batra, N. PlantDoc: A dataset for visual plant disease detection. In Proceedings of the ACM India Joint International Conference on Data Science and Management of Data, Hyderabad, India, 5–7 January 2020; ACM: New York, NY, USA, 2020; pp. 249–253.
34. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.
35. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-end object detection with transformers. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 213–229.
36. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable DETR: Deformable transformers for end-to-end object detection. In Proceedings of the International Conference on Learning Representations, ICLR, Vienna, Austria, 3–7 May 2021; pp. 1–16.
37. Zhang, H.; Li, F.; Liu, S.; Zhang, L.; Su, H.; Zhu, J.; Ni, L.M.; Shum, H.Y. DINO: DETR with improved denoising anchor boxes for end-to-end object detection. *arXiv* **2022**, arXiv:2203.03605.
38. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–19.
39. Tan, J.; Wang, C.; Li, B.; Li, Q.; Ouyang, W.; Yin, C.; Yan, J. Equalization loss for long-tailed object recognition. In Proceedings of the Conference on Computer Vision and Pattern Recognition, IEEE, Seattle, WA, USA, 13–19 June 2020; pp. 11662–11671.
40. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation networks. In Proceedings of the Conference on Computer Vision and Pattern Recognition, IEEE, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
41. 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 Conference on Computer Vision and Pattern Recognition, IEEE, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
42. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the Conference on Computer Vision and Pattern Recognition, IEEE, Seattle, WA, USA, 13–19 June 2020; pp. 11531–11539.

43. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Conference on Computer Vision and Pattern Recognition, IEEE, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
44. Zhang, H.; Wu, C.; Zhang, Z.; Zhu, Y.; Lin, H.; Zhang, Z.; Sun, Y.; He, T.; Mueller, J.; Manmatha, R.; et al. ResNeSt: Split-attention networks. In Proceedings of the Conference on Computer Vision and Pattern Recognition, IEEE, New Orleans, LA, USA, 18–24 June 2022; pp. 2736–2746.