ENGINEERING Agriculture

Engineering Journals

RESEARCH ARTICLE

2026 / Volume 13 / Issue 4 / 26669

GDE-YOLO: a robust and accurate method for real-time tea leaf disease detection in complex plantation environments

Lei GUO[1,2], Hongbin CHEN[1,2], Zheming CUI[1,2], Zhenhui LAI[1,2], Weiliang HUANG[1,2], Tianliang LIN[3], Jufei WANG[1,2], Shuhe ZHENG (✉)[1,2], Wuxiong WENG (✉)[1,2]

1 College of Mechanical and Electrical Engineering, Fujian Agriculture and Forestry University, Fuzhou 350002, China.
2 Fujian University Engineering Research Center for Modern Agricultural Equipment, Fujian Agriculture and Forestry University, Fuzhou 350002, China.
3 Fujian Key Laboratory of Green Intelligent Drive and Transmission for Mobile Machinery, Xiamen 361021, China.

KEYWORDS

Attention mechanism, deep learning, precision agriculture, tea leaf diseases, YOLOv8

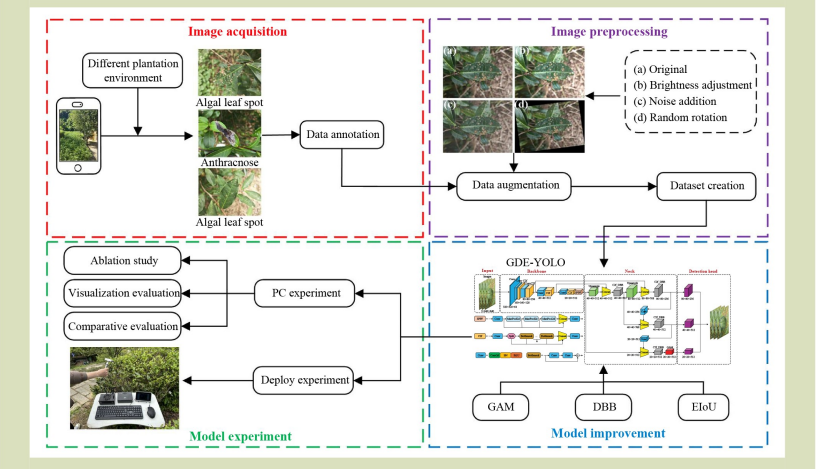
HIGHLIGHTS

- A dataset of tea leaf diseases containing multiple planting environments and categories was constructed.
- A robust and accurate method for real-time tea leaf diseases detection based on GDE-YOLO was proposed.
- GDE-YOLO achieved remarkable 91.7% precision in tea leaf diseases recognition, which was 3.1% higher than YOLOv8n.
- Edge device deployment on NVIDIA Jetson Orin Nano was achieved and the field experiments exhibited superior performance.

Received July 2, 2025;
Accepted October 13, 2025.

Correspondences: zsh@fafu.edu.cn, wwx@fafu.edu.cn

GRAPHICAL ABSTRACT



ABSTRACT

To address the inherent problems of high labor costs and poor efficiency of current visual diagnosis methods for tea leaf diseases, this study proposes a GDE-YOLO-based real-time detection method for tea leaf diseases identification in complex tea plantation environments. The proposed architecture integrates three key enhancements: (1) combination of the neck network with a global attention mechanism (GAM), (2) optimization of the C2f module through a diverse branch block (DBB), and (3) replacement of a complete intersection over union loss function with an efficient intersection over union loss function, collectively improving recognition accuracy and speed. Experimental validation demonstrates that GDE-YOLO achieved 91.7% precision (3.1% higher than YOLOv8n) across different tea plantation scenarios and disease types, with specific improvements of 0.7% for tea anthracnose and 12.4% for tea white spot detection. Also, the enhanced model attained 80 FPS

real-time performance. The deployment test on the NVIDIA Jetson Orin Nano edge device showed that GDE-YOLO could achieve precise diseases identification with confidence threshold > 0.8 and inference speed maintaining 18 FPS, satisfying edge computing requirements of accuracy and real-time performance. This research provides critical technical foundations for vision-guided precision sprayers in tea plantations, while promoting the practical implementation of machine vision in intelligent agricultural management.

© The Author(s) 2026. Published by Higher Education Press. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0)

1 Introduction

Tea is one of the most widely consumed non-alcoholic beverages globally[1], with over two billion cups consumed daily[2]. Tea is not only delicate and delicious but also effective in reducing cholesterol and blood pressure, improving human immunity, and preventing and treating Alzheimer's disease[3,4]. China is the leading global producer of tea[5]. In 2023, the tea-growing area in China reached 3.43 Mha, accounting for 62.2% of the global total, with a year-on-year growth of 3.3%[6]. According to recent statistics from the UN Food and Agriculture Organization of the United Nations, tea production in China was 3.25 Mt in 2023, representing nearly half of the global production, with a 2.2% annual growth rate[7]. However, tea plants are susceptible to various diseases during the summer-autumn growing season[8], posing a major constraint to both yield and quality. Although boom sprayers are widely used due to their time saving, labor saving, efficient and stable characteristics, their use creates issues such as pesticide waste and environmental pollution, struggling to meet the development demands of modern agriculture for green practices, intelligent solutions, precision operations and sustainability[9]. With the rapid advancement of precision agriculture and artificial intelligence (AI) technologies, precision spraying machines are gradually becoming key way to replace standard boom sprayers. Among these, the rapid and accurate identification of tea leaf diseases in complex tea plantation environments (i.e., scenarios where factors such as light changes, vegetation obstruction, weather variations and background interference that affect the performance of the model in disease detection) serves as a key component in the deployment of precision spraying machines[10]. Therefore, developing a tea leaf disease detection method that combines high accuracy, real-time performance and strong robustness is of significant importance.

Current methods for tea leaf disease detection primarily rely on expert diagnosis through direct visual observations. This approach is time-consuming, incurs high labor costs, is not suitable for large-scale tea plantations and is prone to misjudgment due to subjective factors[11]. Compared to visual diagnostic methods, chemical detection methods can detect diseases at early stages, providing a scientific and accurate diagnosis for diseases prevention and control[12,13]. Common chemical detection technologies primarily include high performance liquid chromatography, gas chromatography, mass spectrometry and gas chromatography-mass spectrometry[14]. However, these methods have certain limitations due to their reliance on specialized instruments for data acquisition, such as microscopes, chromatographs and mass spectrometers, which are characterized by high costs, challenging maintenance requirements and the need for highly skilled technical personnel. Subsequently, researchers have adopted remote sensing imagery for detection, which offers advantages including broad applicability, reduced labor costs and multidimensional data acquisition capabilities[15]. However, this approach has constrained efficacy due to susceptibility to resolution limitation and environmental interference, with applicability restricted to specific disease types. Hyperspectral imaging has emerged as a non-destructive sensing technology demonstrating significant potential for detecting crop diseases and pest infestations[16]. This technique enables the simultaneous acquisition of spatial and spectral information, characterized by non-destructive operation, high efficiency and measurement precision[17,18]. Nevertheless, its practical implementation presents significant challenges, including substantial equipment costs, complex data processing requirements and environmental sensitivity under field conditions[19]. Machine vision showcases the capability to replace manual operations through low-cost, high-precision image processing and automated analysis, significantly enhancing agricultural production efficiency. Consequently,

machine vision technology has gained extensive attention in the research of crop diseases and pests identification in precision agriculture.

In the evolution of machine vision architectures, the initial image processing phase predominantly uses manual feature engineering methodologies combined with mathematical transformation techniques (e.g., scale-invariant feature transformation and histogram of oriented gradients) for elementary object recognition, which exhibited constrained generalization capabilities[20]. With the rapid advancement of computer technologies, machine learning models including support vector machines and Adaboost[21] gained prominence in tea leaf diseases identification due to their operational efficiency and algorithmic flexibility, demonstrating commendable diagnostic performance. Nevertheless, these approaches still required manual feature extraction while showing constrained competence in processing high-dimensional and unstructured data[22,23]. The recent surge in computational hardware capabilities and exponential expansion of agricultural big data have catalyzed the integration of deep learning into crop diseases and pests identification systems. This paradigm shift capitalizes on the capacity of end-to-end learning architectures to model intricate nonlinear relationships[24]. Not only can it automatically extract target features with minimal manual intervention, but it also captures nonlinear relationships within data, making it particularly suited for processing high-dimensional and unstructured data, such as images and speech signals[25,26]. It is worth noting that a growing number of researchers have applied deep learning to tea leaf diseases identification and achieved significant results. Hu et al.[27] improved the convolutional neural network (CNN) by using multiscale feature extraction modules and depthwise separable convolutions. Experimental results presented that the enhanced model achieved higher average recognition accuracy compared to standard machine learning and deep learning methods, while exhibiting significantly fewer parameters and requiring fewer convergence iterations than established deep learning network models such as VGG16 and AlexNet. Liu et al.[28] combined a trainable attention mechanism for explanations module with the backbone network of the Faster R-CNN framework for tea leaf disease detection. Experimental results showed that the proposed algorithm achieved superior mean average precision (mAP) compared to YOLOv5, YOLOv7 and original Faster R-CNN. Although existing tea leaf diseases identification methods have achieved high accuracy rates, the data sets they use are captured under well-lit conditions and do not account for the impacts of complex environmental factors on model performance[29], such as excessive lighting, insufficient lighting, rainy weather, soil coverage, and obstruction by leaves and branches. Therefore, there is a clear need to develop a real-time tea leaf disease detection method that maintains high accuracy and robust stability in authentic complex tea plantation environments.

Building upon this background, this study aimed to: (1) construct a tea leaf disease data set encompassing multiple complex plantation environments and disease types, (2) develop a machine vision model with high robustness and accuracy for real-time detection of tea leaf diseases in complex plantation environments, (3) analyze and validate model performance through ablation study, visualization and comparative evaluations, and (4) deploy the developed model on edge devices to assess its feasibility for real-world implementation in future research.

2 Materials and methods

This study aimed to develop a tea leaf disease detection model applicable for natural tea plantation environments. Through architectural refinements to the YOLOv8n model, we enabled high-precision real-time detection across diverse environmental conditions, with subsequent edge deployment validation on the NVIDIA Jetson Orin Nano platform. The overall flowchart of this study is given in Fig. 1.

2.1 Data acquisition

2.1.1 Image acquisition

The image data set used was obtained from the tea plantation located at the Cangshan Campus of Fujian Agriculture and Forestry University (26°4′54″ to 26°4′55″ N; 119°14′23″ to 119°14′27″ E), Fuzhou City, Fujian Province, China. Figure 2 shows the tea plantation situation and symptomatic leaf samples corresponding to different diseases. Under the guidance of phytopathology experts, RGB images of three prevalent tea leaf diseases, viz., tea algal leaf spot (TALS), tea anthracnose (TA) and tea white spot (TWS), were captured using a Xiaomi 13 smartphone equipped with an aperture of f/1.8 and a focal length of 23 mm. The Xiaomi 13 smartphone integrates a

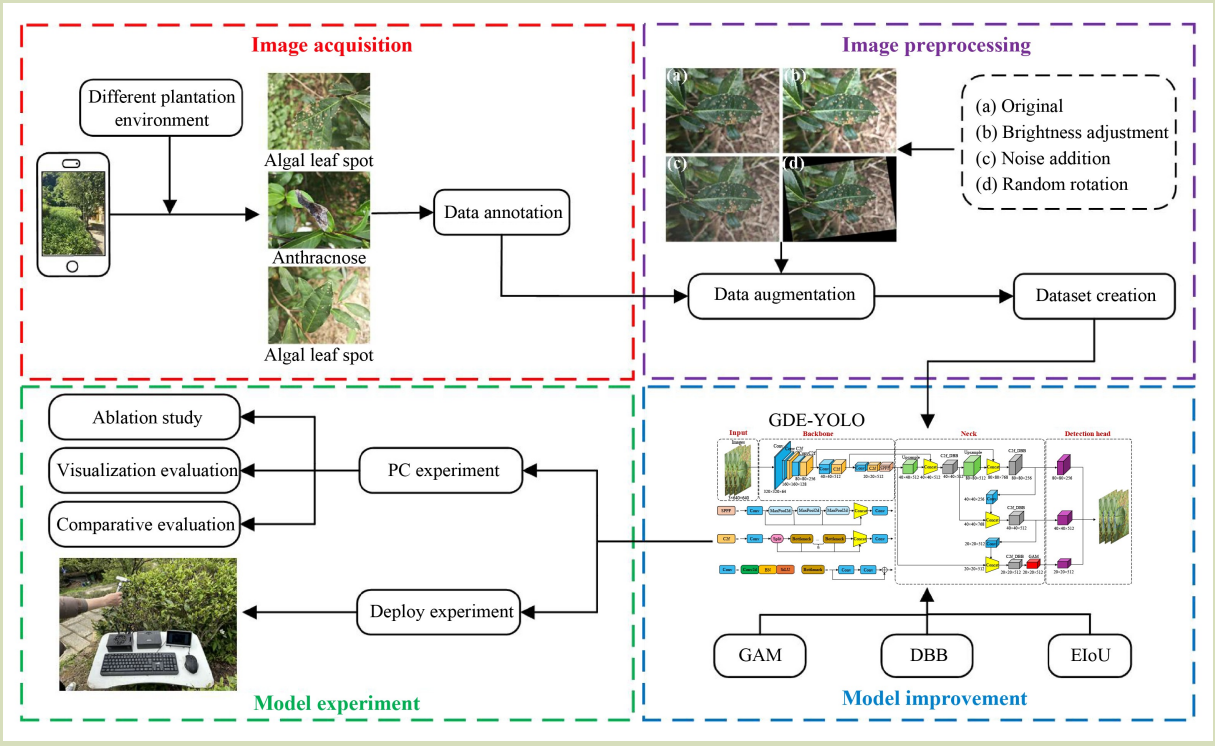


Fig. 1 Overall flowchart of tea leaf disease detection for diverse environmental conditions and disease types.



Fig. 2 Tea plantation context and symptom on leaf samples corresponding to three tea diseases: (a) tea plantation, (b) algal leaf spot, (c) anthracnose, and (d) white spot.

triple-camera system co-engineered with Leica. The primary camera is equipped with a 54-MP Sony IMX800 sensor featuring HyperOIS optical image stabilization, the secondary camera array comprises a 12-MP ultra-wide-angle lens (112° field of view) and a 10-MP telephoto lens (75 mm focal length, 3× optical zoom). During image capture, the native Leica authentic-look profile of the phone was used, with HDR mode and distortion correction enabled, and a photo quality setting of 'High'. Image acquisition was conducted in September to November 2024 at a working distance of 15–25 cm, image resolution of is 3072 × 4096 px and saved in JPG format. Given that varying weather conditions and lighting at different times of day can lead to significant differences in images, data was collected during three time periods (7:00–9:00, 11:00–13:00, 17:00–19:00 local time) under different weather conditions including on sunny, cloudy and rainy days to enhance data set diversity. A total of 1811 images were collected, covering nine distinct environmental conditions: frontlighting, dim light, backlighting, reflection, shadow, staining, after rain, leaf obstruction and branch obstruction (Fig. 3).

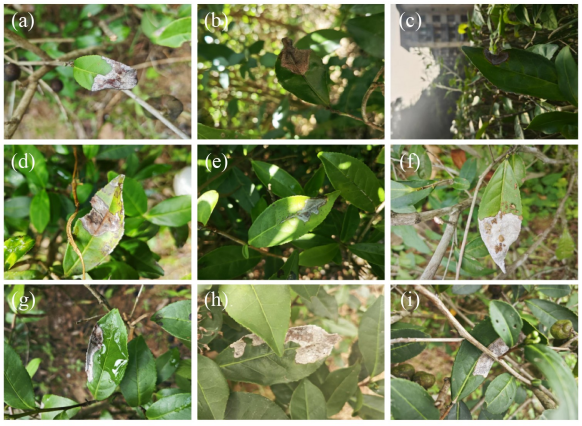


Fig. 3 Tea leaf diseases in different environments: (a) frontlighting, (b) dim light, (c) backlighting, (d) reflection, (e) shadow, (f) staining, (g) after raining, (h) leaf obstruction, and (i) branch obstruction.

2.1.2 Data augmentation

To enhance model recognition performance, data augmentation[30] was used to increase data set diversity. This approach not only improves the generalization capability and robustness of the model but also mitigates challenges posed by sample imbalance in complex tea plantation environments[31]. PyCharm 2023 software (JetBrains, Prague, Czech Republic) and image processing tools were used for data augmentation, including brightness adjustment [using the *addWeighted* function in the OpenCV library, the alpha value range was (1.1, 1.3), and the other parameters value were 0], noise addition (the noise type is Gaussian and the variance was 0.1), and random rotation (the image was rotated by a random angle around the center point, ranging from –180° to 180°), as shown in Fig. 4. The original data set was quadrupled through this data augmentation technique.

2.1.3 Data set creation

The open-source annotation tool LabelImg[32] was implemented to annotate the image data set. Following script execution, all tea leaf diseases images were subsequently annotated with predefined categorical leaf disease labels (viz., TALS, TA and TWS). Subsequently, standard text files compliant with YOLO annotation protocol were generated for each image, encapsulating tea leaf diseases classifications paired with normalized bounding box parameters (center coordinates, width and height) in floating-point representation. These annotation files were subsequently used as training data for the convolutional neural network. Data augmentation operations did not change the actual positions of tea leaf diseases in the images but rescale the coordinates of the bounding boxes through specific transformations to update the diseases coordinates in the text annotation files. In the field of computer vision, the normal ratio for dividing a small-scale data set (no more than 10,000 samples) into a training set, a validation set and a test set is 6:2:2. In this case, the training data are relatively limited, which may lead to underfitting of the model. Therefore, we adopted a 7:2:1 ratio for division, increasing the training data to enhance the model training effect. Although the proportion of the test set has decreased, all its images were strictly selected, covering all types of tea leaf diseases and complex backgrounds, ensuring the diversity and reality of the test data, which is sufficient for reliable evaluation. The augmented data set was ultimately partitioned into a training set (4212 images, 70%), a validation set (301 images, 20%) and a test set (152 images, 10%). The details of the tea leaf disease detection data set are formally documented in Table 1.

2.2 GDE-YOLO model

YOLOv8, one of the most recent iterations in the YOLO series of real-time object detection architectures, features a network structure comprising four principal components: input image



Fig. 4 Tea leaf disease after data augmentation: (a) original, (b) brightness adjustment, (c) noise addition, and (d) random rotation.

Table 1 Details of the tea leaf disease detection data set

Category	Image set (number of images)		
	Training	Validation	Test
TALS	1388	99	50
TA	1472	105	53
TWS	1352	97	49
Total	4212	301	152

Note: TALS, tea algal leaf spot; TA, tea anthracnose; TWS, tea white spot.

preprocessing, backbone for hierarchical feature extraction, neck for multiscale feature fusion and head for bounding box regression and classification predictions[33] YOLOv8 implements multiple architectural enhancements to optimize the speed-accuracy balance, including three core innovations: replacement of C3 module with computationally efficient C2f module in both backbone and neck networks, adoption of decoupled detection heads superseding standard coupled architectures and paradigm shift from anchor-based to anchor-free detection framework[34]. Through these architectural refinements, the YOLOv8 model can deliver significantly enhanced overall performance in object detection tasks. However, when used for tea leaf disease detection in natural plantation environments, the model has several persistent challenges that warrant further investigation.

(1) Under complex environmental conditions encompassing varying illumination, rain and stains, the detection model exhibits heightened susceptibility to both false negatives (missed detections) and false positives (erroneous identifications) during tea leaf diseases diagnosis.

(2) In practical tea plantations characterized by high-density planting and interlacing foliage, diseased leaves are prone to obstruction by adjacent leaves and branches, resulting in compromised lesion localization accuracy in detection models.

(3) Current research on tea leaf disease detection, while achieving improved accuracy, often results in increased computational complexity of models and reduced detection speed, thereby hindering their deployment on edge computing devices.

To address these challenges, we first integrated a global attention mechanism (GAM) module into the neck network of YOLOv8n, then modify the C2f module in the neck network using a diverse branch block (DBB) module, and finally combine an efficient intersection over union (EIoU) loss function, resulting in the proposed GDE-YOLO architecture. These components work synergistically. The neck network takes advantage of the feature that the DBB module can integrate various scales and complex branches while maintaining model lightweightness, and uses its improved C2f-DBB module to perform multiscale fusion on the features extracted from the backbone, providing a more comprehensive multiscale feature basis for the GAM module, enabling it to perform feature selection more accurately. At the end of the neck, the GAM module was used to enhance the significant features of the disease and suppress background noise, thereby reducing the loss of target information and improving the recognition performance of the model for smaller detection boxes that are closer in size to tea leaves. Also, the EIoU loss function is introduced to enable the model to find better solutions more quickly, converting the powerful feature extraction capabilities enhanced by GAM and DBB into more accurate detection results, thereby comprehensively improving model performance. The structure of the GDE-YOLO network is illustrated in Fig. 5.

2.2.1 Global attention mechanism

Existing attention mechanisms predominantly focus on single-dimensional information processing, neglecting the crucial information interaction between channel and spatial dimensions. GAM[35] addresses this limitation by synergistically integrating the strengths of channel attention (CA) and spatial attention (SA) mechanisms, thereby enhancing cross-dimensional feature fusion while mitigating target information loss in complex environmental scenarios. The CA module first uses 3D permutation to preserve three-dimensional information in the input feature map; then uses a

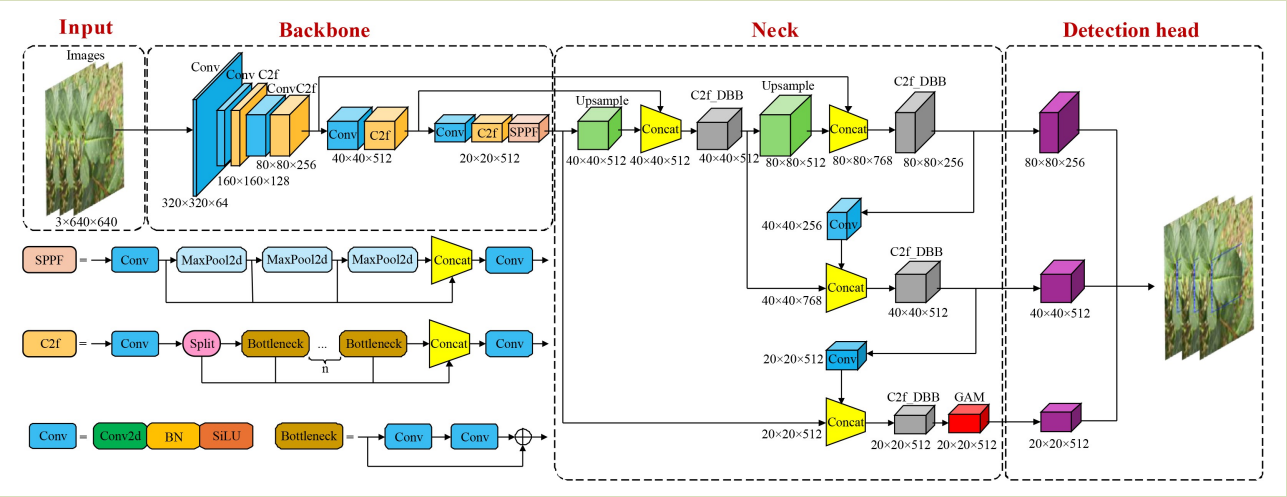


Fig. 5 The structure of GDE-YOLO network, with Conv being Conv2d, BN and SiLU, i.e., Conv2d the convolution layer; BN the batch normalization, and SiLU the sigmoid linear unit activation function; and Maxpool2d the two-dimensional max-pooling operation.

two-layer multilayer perceptron with a reduction ratio r to amplify cross-dimensional channel-spatial dependencies; and finally applies a sigmoid activation function to generate the channel attention feature map. The SA module processes adaptive features by taking as input the element-wise product between the channel attention feature map and residual-derived feature maps from the original input. To emphasize spatial information, SA uses dual 7 × 7 convolutional operations for spatial feature fusion. Also, max-pooling operations are eliminated to enhance feature representation fidelity, as empirical studies demonstrate their detrimental impact on information use efficiency[36]. Figure 6 illustrates the operational mechanism of the GAM, where the input image is sequentially processed by the CA and SA modules through element-wise multiplication. Given an input feature map, the intermediate state and output are computed as:

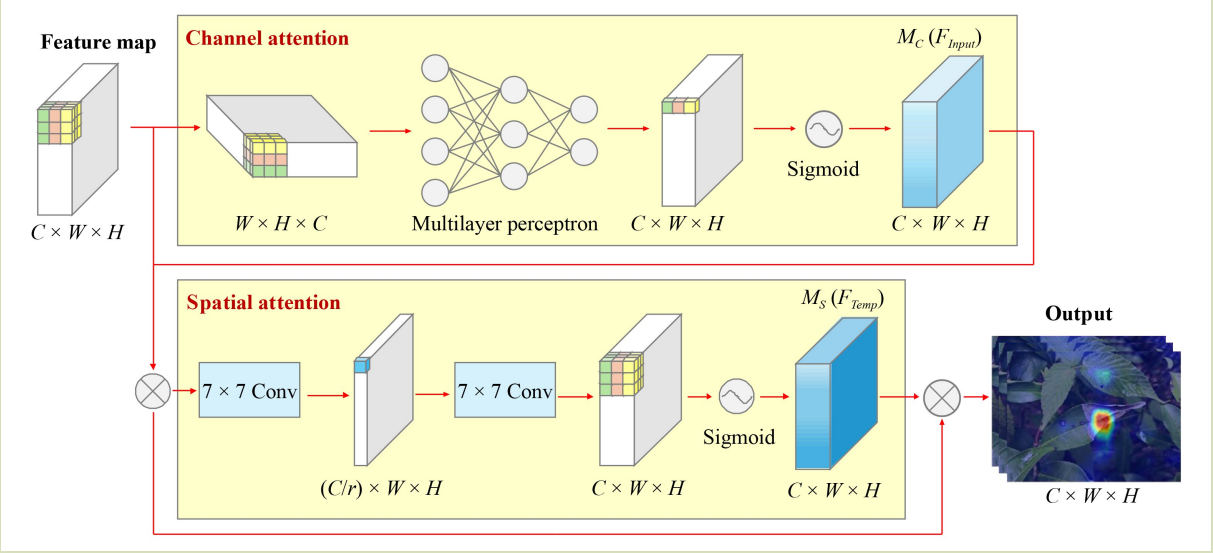


Fig. 6 Network structure of the GAM module. *C*, *H* and *W* are the number of channels, height and width, respectively.

$$F_{Temp} = M_C(F_{Input}) \otimes F_{Input} \quad (1)$$

$$F_{Output} = M_S(F_{Temp}) \otimes F_{Temp} \quad (2)$$

where, F_{Input} is the input feature map, F_{Output} is the output feature map, F_{Temp} is the intermediate transition feature, M_C and M_S are the channel and spatial attention map, respectively, and $\otimes$ is element-wise multiplication.

2.2.2 Diverse branch block

Under normal circumstances, improving model performance often increases computational load and inference time. However, DBB[37] achieves a perfect balance between these two aspects. DBB uses a sophisticated microstructure during training while being computationally equivalent to a single convolution during inference, maintaining model lightweightness while enhancing performance[38]. The DBB structure is illustrated in Fig. 7. Typically, a kernel consists of an output channel D, an input channel C and a kernel size is $K \times K$, which is essentially a fourth-order tensor $F \in R^{D \times C \times K \times K}$ with an optional offset $b \in R^D$. It takes a C channel feature map $I \in R^{C \times H \times W}$ as input and outputs a D channel feature map $O \in R^{D \times H \times W}$, where H and W are determined by K, padding and stride configurations. $*$ is the convolution operator and the offset is $REP(b) \in R^{D \times H \times W}$.

The convolution has the following form:

$$O = I * F + REP(b) \quad (3)$$

The value at (h, w) on the j output channel is given by the following equation:

$$O_{j,h,w} = \sum_{c=1}^{C} \sum_{u=1}^{K} \sum_{v=1}^{K} F_{j,c,u,v} X(c, h, w)_{u,v} + b_j \quad (4)$$

where, $X(c, h, w) \in R^{K \times K}$ is the sliding window on the c channel of I corresponding to position (h, w) on O, and this correspondence is determined by the padding and stride. From Eq. (4), it is easy to deduce the linear properties of convolution, including homogeneity and additivity.

$$I * (pF) = p(I * F), \forall p \in R \quad (5)$$

$$I * F^{(1)} + I * F^{(2)} = I * \left(F^{(1)} + F^{(2)}\right) \quad (6)$$

Note that the additive property holds only when the two convolutions share identical configurations (e.g., number of channels, kernel size, padding and stride).

Based on the two fundamental properties mentioned above, six transformations are summarized below. These use batch normalization (BN), branch addition, depth concatenation, multiscale operations, average pooling and convolution sequences to transform DBB into a single convolution.

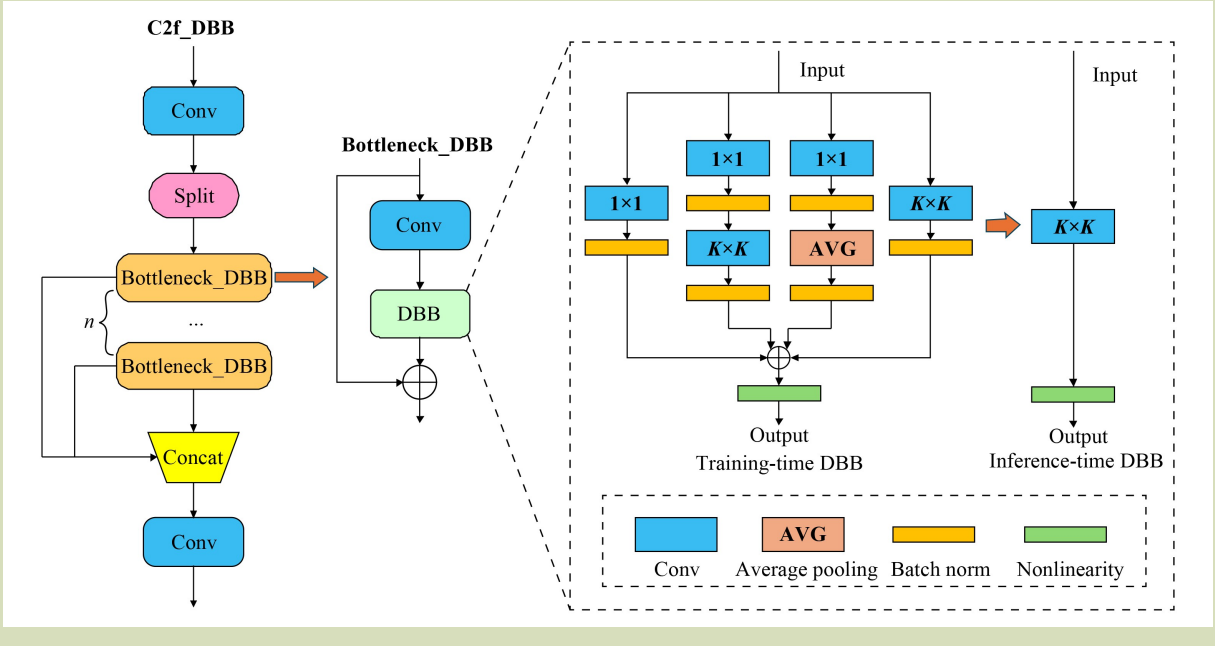


Fig. 7 Improved C2f module of the neck using the diverse branch block (DBB) module. ⊕ is the addition of elements one by one.

Transformation 1 - Conv-BN fusion. Convolutions are usually equipped in deep learning with a BN layer, which performs normalization and linear scaling of the channel. Let j be the channel index, μ_j and σ_j be the accumulated channel mean and standard deviation, γ_j and β_j be the learned scaling factor and bias term, respectively, and the output channel j becomes:

$$O_{j,:,:} = \left((I * F)_{j,:,:} - \mu_j\right)\frac{\gamma_j}{\sigma_j} + \beta_j \tag{7}$$

The homogeneity of convolutions allows BN to be fused into the previous convolutional layers for inference. In practice, simply build a convolution with kernel F' and offset b'. It can be obtained through Eqs. (3) and (7) that each output channel j has:

$$F'_{j,:,:,:} \leftarrow \frac{\gamma_j}{\sigma_j} + F_{j,:,:,:},\ b'_j \leftarrow -\frac{\mu_j\gamma_j}{\sigma_j} + \beta_j \tag{8}$$

Transform 2 - branch addition. Using the additivity of convolutions, two or more convolutional layers with the same configuration can be merged into a single convolution. The two convolutions are merged by the following formula:

$$F' \leftarrow F^{(1)} + F^{(2)},\ b' \leftarrow b^{(1)} + b^{(2)} \tag{9}$$

Transformation 3 - sequence convolution fusion. In convolution, a sequence of 1×1 Conv-BN - $K\times K$ Conv-BN can be merged into a $K\times K$ convolution. Suppose the kernel shapes of the convolutions of 1×1 and $K\times K$ are $D\times C\times1\times1$ and $D\times C\times K\times K$, respectively, where D can be any number. Let $F^{(1)}\in R^{D\times C\times1\times1}$, $b^{(1)}\in R^{D}$, $F^{(2)}\in R^{E\times D\times K\times K}$, $b^{(2)}\in R^{E}$. The output is:

$$O' = \left(I * F^{(1)} + REP\left(b^{(1)}\right)\right) * F^{(2)} + REP\left(b^{(2)}\right) \tag{10}$$

Applying additivity can be obtained:

$$O' = I * F^{(1)} * F^{(2)} + REP\left(b^{(1)}\right) * F^{(2)} + REP\left(b^{(2)}\right) \tag{11}$$

Transformation 4 - splicing and fusion. The Inception unit uses deep connections to combine branches. Given $F^{(1)}\in R^{D_1\times C\times K\times K}$, $b^{(1)}\in R^{D_1}$, $F^{(2)}\in R^{D_2\times C\times K\times K}$, $b^{(2)}\in R^{D_2}$, concatenate them into $F'\in R^{(D_1+D_2)\times C\times K\times K}$ and $b'\in R^{(D_1+D_2)}$, obviously:

$$\begin{aligned} &CONCAT\left(I * F^{(1)} + REP\left(b^{(1)}\right), I * F^{(2)} + REP\left(b^{(2)}\right)\right) \\ &= I + F' + REP(b') \end{aligned} \tag{12}$$

Transformation 5 - average pooling layer transformation. An average pooling with kernel size K and stride s applied to the C channel is equivalent to a convolution with the same K and s. Such a kernel $F'\in R^{C\times C\times K\times K}$ is represented as follows:

$$F'_{d,c,:,:} \begin{cases} \dfrac{1}{K^2} & if\ d = c \\ 0 & elsewise \end{cases} \tag{13}$$

Transformation 6 - multiscale convolution fusion. Considering that an $k_h\times k_w(k_h\leqslant K, k_w\leqslant K)$ core is equivalent to a $K\times K$ core with some zero entries, it is possible to convert an $k_h\times k_w$ core into a $K\times K$ core by zero-padding. Specifically, 1×1, $1\times K$ and $K\times1$ convolutions are particularly practical since they can be implemented efficiently. In addition, the input should also be padded to align the sliding window.

The DBB architecture used 1×1, $1\times1-K\times K$ and $1\times1-\mathrm{AVG}$ to augment the original $K\times K$ convolutional layer, which greatly enriches the feature space by fusing branches of multiple scales and complexity, and improves the performance of the model without additional inference time cost.

2.2.3 Efficient IoU loss function

YOLOv8n adopts a complete intersection over union (CIoU) loss function as the bounding box regression loss, which enhances the standard intersection over union (IoU) by incorporating center distance and aspect ratio to provide a more comprehensive metric. However, the second penalty term of CIoU loss function only constrains aspect ratio similarity while ignoring independent differences in width and height. This leads to a zero penalty when matching aspect ratios, thereby limiting model optimization effectiveness[39]. Based on the aforementioned issues, this study introduces the EIoU loss function[40] to replace the CIoU loss function. EIoU decouples the impact factors of width and height from the aspect ratio in the CIoU penalty term, separately calculating the differences between the predicted box and ground truth box in terms of width and height. The EIoU loss function comprises three components: IoU loss, distance loss and width-height loss. The first two components inherit the methodology from CIoU, while the width-height loss directly minimizes the differences between the predicted box and ground truth box in width and height. This approach accelerates convergence speed, improves regression accuracy and provides more precise fitting for the model. The EIoU loss function can be calculated through the following formula:

$$\begin{aligned} L_{EIoU} &= L_{IoU} + L_{dic} + L_{asp} \\ &= 1 - IoU + \frac{\rho^2(b,b^{gt})}{(c_w)^2+(c_h)^2} + \frac{\rho^2(w,w^{gt})}{(c_w)^2} + \frac{\rho^2(h,h^{gt})}{(c_h)^2} \end{aligned} \tag{14}$$

where, L_{IoU}, L_{dic} and L_{asp} are IoU loss, distance loss and width-height loss, respectively, $\rho^2(b,b^{gt})$ is the Euclidean distance between the center points of the real box and the predicted box, $\rho^2(w,w^{gt})$ and $\rho^2(h,h^{gt})$ are the width and

height differences between the real box and the predicted box respectively, and c_w and c_h are the width and height of the minimum bounding rectangles between the predicted bounding box and the true bounding box.

2.3 Ablation study

This study used the GAM and DBB modules along with the EIoU loss function to enhance YOLOv8n architecture, thereby improving model performance. To systematically evaluate the contributions of individual improvement modules to overall performance and their synergistic interactions, ablation study was conducted to compare and analyze the network performance before and after modification. Ablation study is a scientific research method widely adopted in artificial intelligence that aims to quantify the contributions of model components by iteratively removing or modifying specific elements (e.g., layers and modules). This methodology enables researchers to identify critical components influencing model predictions[41]. The specific models and modules investigated are systematically compared in Table 2.

2.4 Deployment experiment

To validate the real-time performance and practical efficacy of the GDE-YOLO tea leaf disease detection model in edge computing scenarios, and provide core technical support for subsequent development of precision pesticide spraying systems for tea gardens, this study deployed the GDE-YOLO model on the NVIDIA Jetson Orin Nano embedded platform for experiments. The NVIDIA Jetson Orin Nano is a powerful embedded AI device based on the NVIDIA Ampere architecture, delivering up to 40 TOPS of AI computing power. It had a 6-core ARM Cortex-A78AE CPU and a GPU with 512 CUDA cores, supporting up to 8 GB LPDDR5 memory. This configuration enables seamless execution of complex AI models and real-time multitasking[42,43]. Given its compact size and exceptional energy efficiency, this study selected the NVIDIA Jetson Orin Nano as the edge computing platform for deploying the GDE-YOLO model. The hardware and software configurations of the device are detailed in Table 3.

The detection system comprised four key components: an imaging camera, an NVIDIA Jetson Orin Nano core processing module, a DC power supply and a display unit. Specifically, the imaging camera is responsible for capturing and transmitting real-time image data streams, while the NVIDIA Jetson Orin Nano executes algorithm inference and analytical processing. A stable 19 V DC power supply ensures uninterrupted operation, and the display unit served as a human-machine interface for visualizing detection results and system status. We used TensorRT[47] to accelerate the GDE-YOLO model. TensorRT is a high-performance deep learning inference engine developed by NVIDIA, specifically designed for deployment in production environments to enhance model inference speed and computational efficiency[48]. In the aspect of optimizing parameter settings, we adopt the half-precision floating-point quantization method to export the TensorRT engine. This approach can markedly enhance performance with minimal loss of the original accuracy of the model. The batch size is configured to be 1, simulating the real-time single-frame processing scenario. Additionally, the input image size is set to 640 × 640 px, which is consistent with the size used during the training process.

Table 2 Improved models for the ablation study

Improved model	YOLOv8n architecture	GAM	DBB	EIoU
YOLOv8n	√	–	–	–
Case 1	√	√	–	–
Case 2	√	–	√	–
Case 3	√	–	–	√
Case 4	√	√	√	–
Case 5	√	√	–	√
Case 6	√	–	√	√
GDE-YOLO	√	√	√	√

Note: √, this module is included; –, this module is not included.

Table 3 Deployment environment of NVIDIA Jetson Orin Nano

Hardware/Software environment	Version
Development board	NVIDIA Jetson Orin Nano
Operating system: Ubuntu[44]	20.04
Python	3.8.18
PyTorch[45]	1.11.0
Torchvision[46]	0.12.0
Ultralytics	8.2.79
TensorRT	8.5.2.2

2.5 Hardware and software configurations

All models were trained under identical hardware and software configurations. The hardware specifications included an Intel(R) i5-9400 CPU (2.9 GHz), 8 GB DDR4 RAM, and an NVIDIA GeForce GTX 1650 GPU. The software environment comprised CUDA 12.1, Python 3.9 and PyTorch 2.2.1 for deep learning implementation. Training parameters were configured as follows: an initial learning rate of 0.01, a weight decay coefficient of 0.0005, a momentum coefficient of 0.937 and 200 training epochs. Since larger batch sizes have been demonstrated to improve model detection performance[49], a batch size of 8 was selected for this study.

2.6 Performance evaluation

Standard deep learning performance metrics were used to evaluate model efficacy and validate the applicability of GDE-YOLO for tea leaf disease detection. The metrics included *Precision*, *Recall*, *mAP* (mean average precision), *Parameters*, *FLOPs* (floating-point operations per second), *Model Size* and *FPS* (frames per second). Where *Precision* is the ratio of correctly predicted samples in the test set to the total number of predicted positive samples, also indicating the possibility that a sample belongs to a specific class. *Recall* is the ratio of correctly predicted samples in the test set to the total number of actual positive samples, reflecting the probability of the model correctly identifying instances of a specific category. *mAP* is to the average precision across all classes during neural network model training, representing overall model performance across all categories. To ensure accuracy, the IoU threshold for the *mAP* evaluation metric was set to 0.5, enabling a comprehensive assessment of model performance. Additionally, metrics such as *Parameters*, *FLOPs*, *Model Size* and *FPS* reflect model performance in resource consumption, computational efficiency, deployment feasibility, detection speed and other aspects. The corresponding definitions are:

$$Precision = \frac{TP}{TP+FP} \times 100\% \tag{15}$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \tag{16}$$

$$mAP = \frac{\sum_{i=1}^{n} AP_i}{n} \times 100\% \tag{17}$$

$$AP = \int_0^1 Precision(Recall)\,\mathrm{d}(Recall) \tag{18}$$

where, TP is the number of samples correctly predicted as tea leaf diseases, FP is the number of samples incorrectly predicted as tea leaf diseases, FN is the number of samples mistakenly predicted as background, and TN is the number of samples correctly predicted as background. n is the number of categories, which is set to 3. AP_i is the average accuracy of category i.

3 Experimental results and analysis

3.1 Analysis of the ablation study

This study evaluated the performance of multiple enhanced models using a custom tea leaf disease data set, with experimental results summarized in Table 4. The results demonstrate that the original YOLOv8n model provided an overall accuracy of 88.6%, with precision values of 95.0% for TALS and 92.5% for TA, while the precision for TWS was only 78.2%, indicating poor recognition performance. Compared to the original YOLOv8n, Case 1 improved the recall and mAP@0.5 of TA by 0.9% and 0.3%, respectively, and increased the precision and mAP@0.5 of TWS by 3.3% and 0.7%. These results indicate that integrating the GAM module into YOLOv8n enhances its perceptual capability for TA and TWS by reducing global information loss and preserving critical features. In Case 2, the precision and mAP@0.5 of TA improved by 1.7% and 0.5% compared to the original YOLOv8n, while the precision of TWS significantly increased by 5.2%, with no notable rise in model parameters or computational load. This confirms the ability of the DBB module to improve accuracy while retaining model lightweightness. In Case 3, the adoption of the EIoU loss function substantially enhanced TWS precision and mAP@0.5 by 7.9% and 1.5%, proving that adopting the EIoU loss function substantially strengthens the ability of the model to

Table 4 Comparison of detection performance for different improved networks

Model	Category	Precision (%)	Recall (%)	mAP@0.5 (%)	Parameter (M)	FLOPs (G)	Model size (MB)	FPS
YOLOv8n	All	88.6 (0.0%)	82.4	90.3	3.0	8.1	6.2	133 (0)
	TALS	95.0 (0.0%)	76.8	91.5				
	TA	92.5 (0.0%)	89.4	93.9				
	TWS	78.2 (0.0%)	81.0	85.4				
Case 1	All	85.5 (↓3.1%)	81.6	89.3	4.6	9.4	9.5	82 (↓51)
	TALS	85.4 (↓9.6%)	75.0	87.6				
	TA	89.5 (↓3.0%)	90.3	94.2				
	TWS	81.5 (↑3.3%)	79.4	86.1				
Case 2	All	88.7 (↑0.1%)	79.7	88.1	3.3	8.7	7.0	91 (↓42)
	TALS	88.5 (↓6.5%)	76.8	84.5				
	TA	94.2 (↑1.7%)	86.3	94.4				
	TWS	83.4 (↑5.2%)	76.0	85.5				
Case 3	All	88.7 (↑0.1%)	79.1	88.7	3.0	8.1	6.2	102 (↓31)
	TALS	89.1 (↓5.9%)	72.3	87.3				
	TA	90.9 (↓1.6%)	85.0	91.8				
	TWS	86.1 (↑7.9%)	80.0	86.9				
Case 4	All	88.4 (↓0.2%)	78.8	89.1	5.0	10.0	10.3	80 (↓53)
	TALS	93.3 (↓1.7%)	74.4	89.7				
	TA	92.9 (↑0.4%)	85.8	94.0				
	TWS	79.0 (↑0.8%)	76.0	83.8				
Case 5	All	82.0 (↓4.6%)	84.6	88.7	4.6	9.4	9.5	95 (↓38)
	TALS	81.9 (↓13.1%)	80.4	88.8				
	TA	88.5 (↓4.0%)	89.4	93.2				
	TWS	75.5 (↓2.7%)	84.0	84.1				
Case 6	All	84.4 (↓4.2%)	84.9	89.2	3.3	8.7	7.0	86 (↓47)
	TALS	86.2 (↓8.8%)	78.6	88.0				
	TA	85.8 (↓6.7%)	92.0	93.4				
	TWS	81.1 (↑2.9%)	84.0	86.1				
GDE-YOLO	All	91.7 (↑3.1%)	78.8	89.0	5.0	10.0	10.3	80 (↓53)
	TALS	91.2 (↓3.8%)	74.4	87.6				
	TA	93.2 (↑0.7%)	85.2	93.2				
	TWS	90.6 (↑12.4%)	76.8	86.2				

Note: TALS, tea algal leaf spot; TA, tea anthracnose; TWS, tea white spot.

recognize TWS. For Cases 4 to 6, we conducted experiments by pairwise combinations of GAM, DBB and EIoU to evaluate the impact of interactions between the improved modules on model performance. Compared to the original YOLOv8n, all three cases demonstrated consistent improvements in the precision of TA and TWS. Of these models, only Case 6 provided no significant increase in parameters, FLOPs or model size. In contrast, Cases 4 and 5, due to the inclusion of the GAM module, provided modest increases in parameters, FLOPs and model size.

It is noteworthy that GDE-YOLO achieves the highest precision of 91.7% among all models, surpassing the original YOLOv8n by 3.1%. Specifically, TA and TWS precision improved by 0.7% and 12.4%, respectively, highlighting a breakthrough in TWS recognition. According to the research proposed by Li et al.[50], the FPS for real-time field processing needs to exceed 4.8 FPS. With an FPS of 80, GDE-YOLO satisfies real-time detection requirements and facilitates subsequent deployment. The ablation experiment results indicate that GDE-YOLO outperforms seven alternative architectures in accuracy while maintaining real-time detection capabilities, validating the effectiveness of the proposed improvements in enhancing model performance.

3.2 Visualization evaluation of GDE-YOLO by Grad-CAM

To comprehensively evaluate the detection performance of GDE-YOLO for tea leaf diseases in natural tea garden environments, this study uses Gradient-weighted class activation mapping (Grad-CAM)[51] for visual assessment. Figure 8 presents the heat map visualization results generated by Grad-CAM for multiple improved models across different tea leaf diseases categories and environmental conditions. Vertically arranged images (top to bottom) represent: TALS, TA, TWS, frontlighting, dim light, backlighting, reflection, shadow, staining, after rain, leaf obstruction and branch obstruction. Horizontally aligned are eight improved models from the ablation experiments. Warm-colored regions in the heat maps denote areas with higher contribution weights to detection outcomes[52].

In Fig. 8, although the original YOLOv8n can detect disease regions in TWS images, it suffers significant interference from the background, leading to reduced focus on the disease areas. Combined with the data in Table 4, this is the primary reason for low recognition accuracy of YOLOv8n for TWS disease. A similar trend is observed for TALS and under conditions such as frontlighting, dim light, shadow, after rain, leaf obstruction and branch obstruction. Also, YOLOv8n struggled to effectively focus on disease regions in backlighting environment, resulting in false positives and missed detections. Compared to the original YOLOv8n, Case 1 enhanced model robustness, reduces background interference and improves network attention to disease regions in TALS, TWS and under frontlighting, after rain, leaf obstruction and branch obstruction scenarios, with particularly notable improvements for TWS and leaf obstruction. Case 1 also achieved better localization of disease regions in backlighting environments, though its performance slightly declines under dim light. Case 2, compared to Case 1, improved detection accuracy for disease regions in dim light but shows reduced effectiveness in backlighting and after-rain conditions. In contrast, Case 3 outperformed Cases 1 and 2 by accurately detecting disease regions, especially in backlighting and after rain conditions, significantly mitigating complex background interference and achieving superior regression precision. Cases 4–6 provided minor improvements in focusing on TWS, but redundant information remains prevalent in the images. Under backlighting, strong light interference caused these models (Cases 4–6) to struggle in extracting critical disease features, often misclassifying healthy leaves or background regions as diseased.

GDE-YOLO achieved the best recognition performance among all models by effectively integrating the strengths of GAM, DBB and EIoU, enhancing tea leaf disease detection capabilities. Compared to the original YOLOv8n, GDE-YOLO significantly improved accuracy by intensifying network focus on TWS. As indicated in Table 4, GDE-YOLO delivered a precision of 90.6% for TWS, surpassing the original YOLOv8n by 12.4%, while overall model precision increased by 3.1%. Also, GDE-YOLO reduced the interference of strong backlighting, precisely localized diseases in images and enhanced model robustness. For other diseases and environmental conditions, it minimizes redundant information and amplifies the predictive contribution of disease regions. Heat map visualization revealed that GDE-YOLO substantially suppressed background noise, strengthened critical disease feature extraction and fully exploited global contextual information, which ultimately elevated detection performance.

3.3 Comparative evaluation of other object detection models

To demonstrate the overall advantages of GDE-YOLO in detecting tea leaf diseases, comparative evaluation of GDE-YOLO and other established object detection models, including Faster RCNN[53], SSD[54], YOLOv3-tiny[55], YOLOv5n[56] and YOLOv8n[57] was conducted. The detailed comparative results are given in Table 5.

From Table 5, it can be seen that YOLOv5n had the smallest

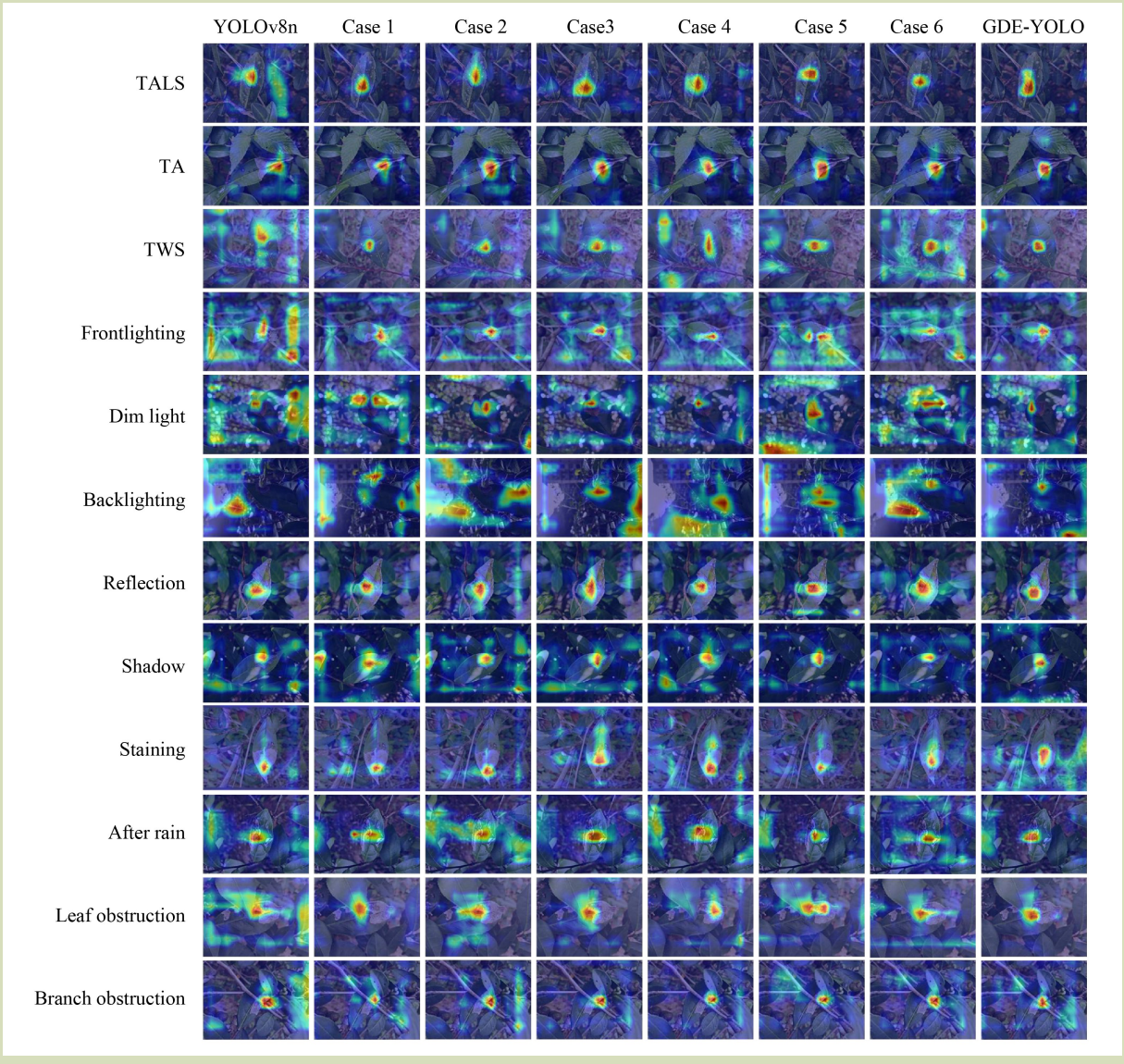


Fig. 8 Visualization of tea leaf disease as heat maps for disease detection. TALS, tea algal leaf spot; TA, tea anthracnose; TWS, tea white spot.

parameters (1.8 M), FLOPs (4.1 G) and model size (3.6 MB), while YOLOv3-tiny achieved the fastest detection speed at 159 FPS. Although the GDE-YOLO model lacked advantages in complexity and detection speed metrics, it surpassed other models in detection precision for tea leaf diseases, with the improved model achieving an overall precision of 91.7%, outperforming Faster RCNN, SSD, YOLOv3-tiny, YOLOv5n and YOLOv8n by 33.5%, 1.9%, 17.2%, 5.0% and 3.1%, respectively. For TALS detection, the YOLOv8n model delivered the highest precision at 95.0%, and compared to the original YOLOv8n, GDE-YOLO had a 3.8% decline in TALS precision but still reached 91.2%, maintaining a high accuracy level. In TA detection, GDE-YOLO achieved the highest precision of 93.2%, surpassing Faster RCNN, SSD, YOLOv3-tiny, YOLOv5n and YOLOv8n by 33.7%, 0.3%, 10.6%, 3.2% and 0.7%, respectively. Notably, for TWS detection, GDE-YOLO effectively addressed the low baseline precision caused by illumination interference, significantly improving TWS precision to 90.6%, which is 40.6%, 7.1%, 27.8%, 14.7% and 12.4% higher than Faster RCNN, SSD, YOLOv3-tiny,

Table 5 Experimental comparison results of other object detection models

Model	Category	Precision (%)	Recall (%)	mAP@0.5 (%)	Parameters (M)	FLOPs (G)	Model size (MB)	FPS
Faster RCNN	All	58.2	73.7	82.2	136.7	401.7	108.0	13
	TALS	65.2	87.0	86.1				
	TA	59.5	91.3	90.6				
	TWS	50.0	72.7	69.9				
SSD	All	89.8	70.8	80.5	23.9	274.0	91.6	15
	TALS	92.9	74.0	84.3				
	TA	92.9	75.6	85.3				
	TWS	83.5	62.7	72.0				
YOLOv3tiny	All	74.5	74.0	77.7	8.7	12.9	17.4	159
	TALS	78.1	72.3	78.3				
	TA	82.6	79.6	84.6				
	TWS	62.8	70.0	70.1				
YOLOv5n	All	86.7	87.7	92.1	1.8	4.1	3.6	121
	TALS	94.3	87.5	91.9				
	TA	90.0	87.7	94.7				
	TWS	75.9	88.0	89.6				
YOLOv8n	All	88.6	82.4	90.3	3.0	8.1	6.2	133
	TALS	95.0	76.8	91.5				
	TA	92.5	89.4	93.9				
	TWS	78.2	81.0	85.4				
GDE-YOLO	All	91.7	78.8	89.0	5.0	10.0	10.3	80
	TALS	91.2	74.4	87.6				
	TA	93.2	85.2	93.2				
	TWS	90.6	76.8	86.2				

Note: TALS, tea algal leaf spot, TA, tea anthracnose; TWS, tea white spot.

YOLOv5n and YOLOv8n. In terms of detection speed, GDE-YOLO achieved 80 FPS, outperforming Faster RCNN and SSD but lagging behind YOLOv3-tiny, YOLOv5n and YOLOv8n, yet still meeting real-time detection requirements. Therefore, the GDE-YOLO tea leaf disease detection model demonstrates superior precision over other established models, effectively enhancing accuracy for TA and TWS, with the decline in TALS precision and speed reduction remaining within acceptable limits to meet practical demands.

Additionally, the recognition performance of GDE-YOLO and other established object detection models were tested on the three tea leaf diseases (TALS, TA and TWS) with results shown in Fig. 9. Only Faster RCNN, YOLOv3-tiny, YOLOv5n and GDE-YOLO accurately identified TALS disease. However, SSD missed detection due to blurred lesion features caused by strong light reflections, while YOLOv8n misclassified healthy leaves as TWS under stain interference. For TA detection, all models achieved accurate identification. In TWS detection, Faster RCNN, SSD, YOLOv3-tiny, YOLOv8n and GDE-YOLO demonstrated robust performance. However, YOLOv5n produced false positives by misclassifying healthy leaves as TWS under complex background lighting and stain interference. Through the above analysis, GDE-YOLO provided superior overall performance compared to other established object detection models.

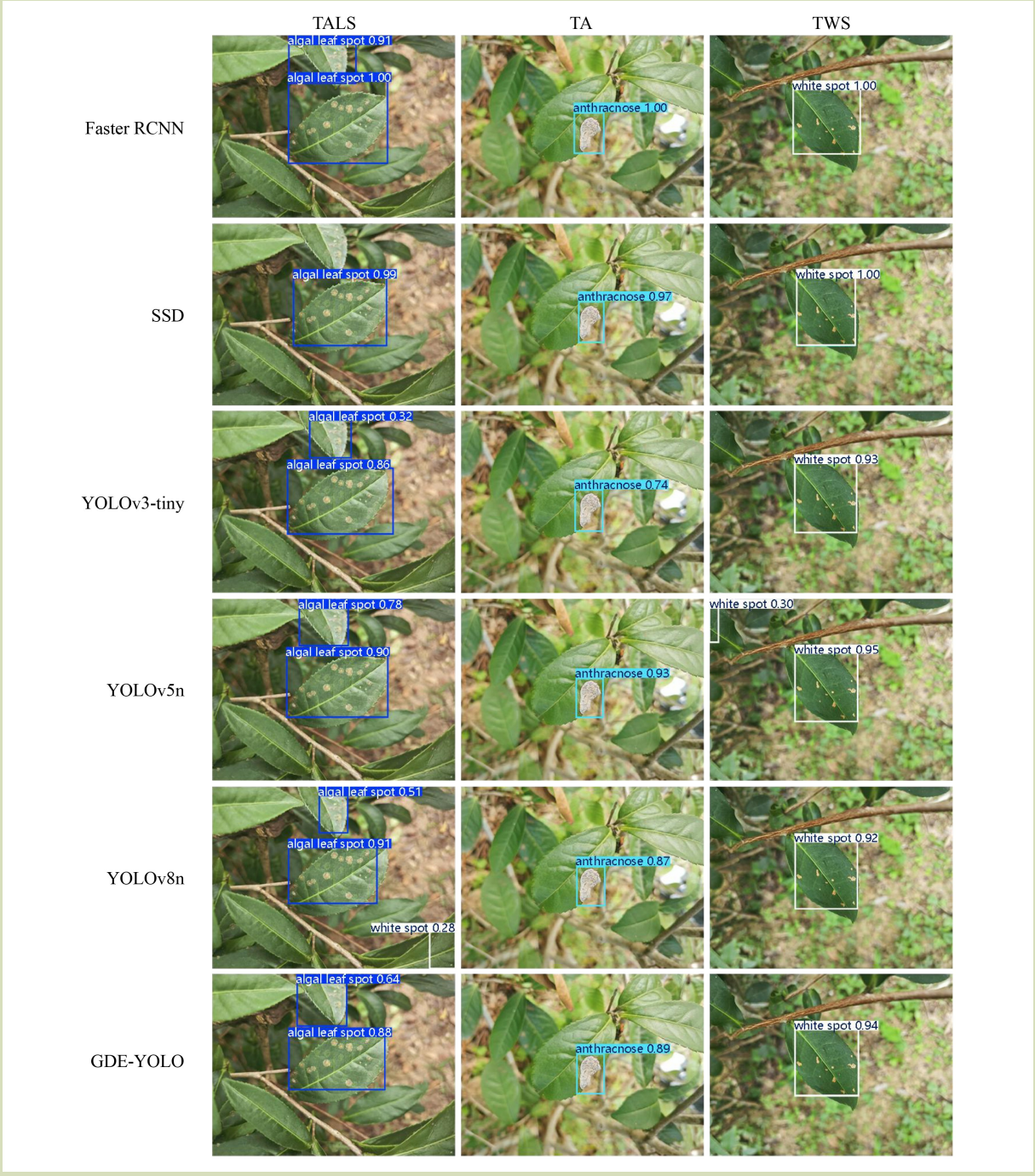


Fig. 9 Comparative evaluation of tea leaf disease detection between GDE-YOLO and other established object detection models. TALS, tea algal leaf spot, TA, tea anthracnose; TWS, tea white spot.

3.4 Deployment experiment results of GDE-YOLO

The deployment experiment results are illustrated in Fig. 10. Field tests under real-world tea garden conditions demonstrated that the edge deployed GDE-YOLO model achieved confidence scores exceeding 0.8 for TALS detection despite strong lighting interference and complex foliage

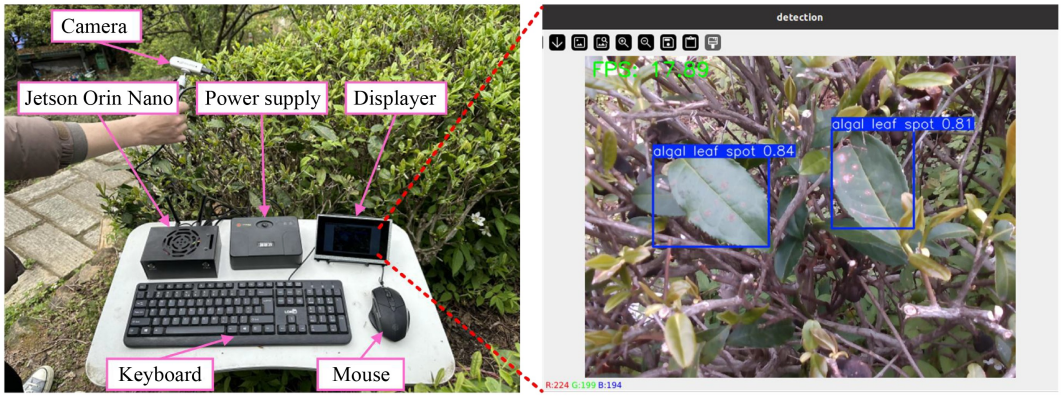


Fig. 10 Deployment of GDE-YOLO on the NVIDIA Jetson Orin Nano.

backgrounds, while maintaining a detection speed of 18 FPS, which surpassed real-time operational requirements. These results demonstrate that GDE-YOLO offers superior accuracy, real-time responsiveness and environmental robustness in edge computing scenarios.

4 Discussion

This study addressed the challenge of tea quality and yield reduction caused by leaf diseases during summer and autumn seasons by proposing GDE-YOLO, using a novel tea leaf disease detection model designed for object detection and real-time identification in complex tea garden environments. The model provided exceptional performance through high detection accuracy while maintaining low missed detection and false positive rates. This advancement has significant implications for enhancing tea production, mitigating agricultural loss, guiding disease control strategy and promoting intelligent agricultural management in tea production systems.

The method proposed in the present study successfully achieved rapid and efficient real-time detection of tea leaf diseases in complex tea garden environments. A comparative evaluation of the original YOLOv8n and GDE-YOLO on the tea leaf disease data set revealed that the GDE-YOLO model can attain a detection accuracy of 91.7%, representing a significant improvement of 3.1% over YOLOv8n. Also, in the detection of multiple disease categories, the enhanced model demonstrates a remarkable 12.4% precision increase, specifically for TWS detection. In the ablation experiments, Case 1 achieved a 3.3% precision improvement for TWS detection, demonstrating that the GAM module enhanced the network feature extraction for TWS by minimizing global information loss and suppressing background interference. Case 2 provided 1.7% and 5.2% precision gains for TA and TWS detection compared to the original YOLOv8n, respectively, while maintaining nearly identical parameter count and computational complexity, proving that the DBB module improved accuracy while preserving model lightweightness. In Case 3, the precision of TWS increased by 7.9% compared to the original YOLOv8n, and the detection speed (102 FPS) was the highest of all tested models except YOLOv8n, indicating that the EIoU loss function enables the model to achieve faster convergence and higher regression precision. Visual evaluation demonstrated that GDE-YOLO effectively reduced redundant information, enhanced resistance to background interference, improved model robustness and amplified the contribution of disease regions during prediction. Of all models in this evaluation, YOLOv8n achieved the highest detection accuracy for TALS but gave significantly lower precision in TWS detection. This could be attributed to the limited perceptual capability of YOLOv8n for TWS and poor interference resistance, which led to frequent misclassification of background regions as TWS, thereby degrading detection accuracy. While YOLOv3-tiny provided the fastest detection (159 FPS), its overall detection performance was suboptimal. The underlying cause likely stems from its simplified network architecture. Although drastically reducing parameter count and computational complexity, this design sacrificed detail capture capability and

processing depth for complex scenes and objects, ultimately compromising detection precision. The GDE-YOLO model proposed in the present study not only provided exceptional detection accuracy but also significant advantages in detection speed, achieving 80 FPS and striking an optimal balance between precision and efficiency. When deployed on the NVIDIA Jetson Orin Nano in the field, the model accurately identified tea leaf diseases in images with confidence scores exceeding 0.8, while maintaining a detection speed of nearly 18 FPS. These results validate that GDE-YOLO fulfills both accuracy requirements and real-time operational demands in edge computing scenarios.

Beyond having significant advantages in overall accuracy and real-time performance, GDE-YOLO also maintained robust recognition stability in complex tea plantation environments. Visual analysis showed that although the original YOLOv8n model can detect disease areas in TWS images, it was easily disturbed by complex backgrounds, leading to a decrease in attention to the target areas and limited recognition accuracy. In various environments including TALS and under conditions of frontlighting, dim light, backlighting, shadow, after rain, leaf obstruction and branch obstruction, the redundant information in the image background increased the difficulty of accurately identifying and locating diseases, with the impact of backlight conditions being particularly significant. In contrast, GDE-YOLO had the ability to focus on the features of TWS disease, significantly increasing the recognition accuracy to 90.6%, a 12.4% improvement over the original model. At the same time, GDE-YOLO can suppress the interference of background redundant information when dealing with other disease types and different environmental scenarios, enhancing the effective contribution of disease areas to the prediction results. Particularly, it can effectively suppress the strong light interference in backlight environments, accurately locate disease areas and improve recognition stability under such adverse conditions. These changes led to a 3.1% overall accuracy improvement for GDE-YOLO, demonstrating superior generalization ability and robustness in complex backgrounds.

In recent years, researchers have developed numerous advanced technologies to improve the accuracy of tea leaf disease recognition. Lin et al.[58] proposed a tea leaf disease detection model called TSBA-YOLO, which enhances the YOLOv5 architecture by integrating modules such as Transformer[59], BiFPN[60], ASFF[61] and SIoU[62]. This method achieves high detection accuracy and speed on the tea leaf disease data set constructed in this research, with a precision of 86.8% and an inference rate of 57 FPS. Ye et al.[63] proposed BRA-YOLOv7, a tea leaf disease detection model based on YOLOv7, which introduces MPDIoU[64], PConv[65] and FasterNet[66] to replace the original loss function and backbone network architecture, thereby improving model efficiency and convergence speed. Additionally, their model incorporates a dual-level routing dynamic sparse attention mechanism to enhance its capability of capturing global contextual information for tea leaf disease detection. The experimental comparison of BRA-YOLOv7 on our data set showed that this approach achieves a precision of 84.3% and a frame rate of 46 FPS. The real-time tea leaf disease detection model GDE-YOLO proposed in the present study achieved an accuracy of 91.7%, representing improvements of 4.9% and 7.4% compared to the TSBA-YOLO and BRA-YOLOv7 models, respectively. This demonstrates the potential of the GDE-YOLO model to enhance the accuracy of tea leaf disease recognition. Additionally, GDE-YOLO exhibits significant advantages in detection speed, achieving an FPS of 80, which is 23 and 34 FPS higher than TSBA-YOLO and BRA-YOLOv7, respectively.

The present study demonstrates that the GDE-YOLO model exhibits high detection accuracy and speed on the tea leaf disease data set. This technology integrates with IoT devices (e.g., field cameras, drones and satellite remote sensing) to establish real-time monitoring networks for early disease warnings, enabling farmers to implement targeted control measures promptly and prevent the spread of diseases. Additionally, it has the potential to be combined with robot technology (e.g., three-dimensional space precise positioning and path planning) to build an automatic spraying device for tea garden for precise pesticide application, which would help reduce pesticide usage, thereby mitigating environmental pollution and minimizing resource waste[67–69]. However, a significant drawback limiting the widespread application of machine vision technology in agricultural contexts remains the substantial time and cost required for data set creation[70]. Under manual labeling conditions, it is impossible to obtain high-quality and large-scale data sets in a short period of time. Although there are currently some automatic annotation tools, such as AutoLabelImg[71], they demonstrate suboptimal annotation velocities and low annotation quality, particularly when handling large-scale annotations or small targets[72]. Therefore, developing a self-supervised adaptive object

detection algorithm-potentially empowered through technologies such as reinforcement learning is crucial. We aim to achieve this in the future, providing critical technical support for intelligent management in tea production and contributing to sustainable agricultural practices.

5 Conclusions

This paper proposes GDE-YOLO, a real-time tea leaf disease detection model for complex tea garden environments, leveraging machine vision technologies and convolutional neural networks. The three conclusions are made from this study.

(1) Integrating GAM, DBB and EIoU into the YOLOv8n network architecture significantly enhanced the detection capabilities of the model. Compared to the original YOLOv8n, the GDE-YOLO model provided a 3.1% improvement in overall detection accuracy on the tea leaf disease data set, with specific accuracy gains of 0.7% for TA disease and 12.4% for TWS disease, indicating substantial optimization of model performance.

(2) Analysis of ablation study and heat map visualization revealed that the three methodological improvements in this study effectively strengthen the feature extraction capabilities of the model, enhanced contextual information utilization, mitigate background interference in predictions, and improved detection efficiency and accuracy. In the comparative evaluation, GDE-YOLO provided a recognition accuracy of 91.7% while maintaining a detection speed of 80 FPS, demonstrating superior overall performance compared to other object detection models.

(3) Deploying the GDE-YOLO model on the NVIDIA Jetson Orin Nano for field experiments enabled accurate identification of tea leaf diseases in images, with confidence scores consistently > 0.8 and a detection speed of 18 FPS, demonstrating that the model meets both accuracy and real-time performance requirements in edge computing scenarios.

The GDE-YOLO model proposed in this study effectively fulfills the requirements for both accuracy and real-time performance. It not only enables tea farmers to promptly suppress diseases spread, thereby reducing loss and improving yield and quality, but also holds potential for integration into automated vision systems within precision spraying devices in tea plantations, providing critical technical support for intelligent agricultural management.

Acknowledgements

This work was supported by Open Foundation of Fujian Key Laboratory of Green Intelligent Drive and Transmission for Mobile Machinery, China (GIDT-202308).

Compliance with ethics guidelines

Lei Guo, Hongbin Chen, Zheming Cui, Zhenhui Lai, Weiliang Huang, Tianliang Lin, Jufei Wang, Shuhe Zheng, and Wuxiong Weng declare that they have no conflicts of interest or financial conflicts to disclose. This article does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

1. Zhai X, Zhang L, Granvogl M, Ho C T, Wan X. Flavor of tea (*Camellia sinensis*): a review on odorants and analytical techniques. *Comprehensive Reviews in Food Science and Food Safety*, 2022, **21**(5): 3867–3909
2. Pan S Y, Nie Q, Tai H C, Song X L, Tong Y F, Zhang L J F, Wu X W, Lin Z H, Zhang Y Y, Ye D Y, Zhang Y, Wang X Y, Zhu P L, Chu Z S, Yu Z L, Liang C. Tea and tea drinking: China's outstanding contributions to the mankind. *Chinese Medicine*, 2022, **17**(1): 27
3. Yang C S, Hong J. Prevention of chronic diseases by tea: possible mechanisms and human relevance. *Annual Review of Nutrition*, 2013, **33**(1): 161–181
4. Drew L. The growth of tea. *Nature*, 2019, **566**(7742): S2–S4
5. Liu Y, Liu H, Xu W, Wang L, Wang Q, Ou G, Wu M, Hong Z.

Advances and challenges of carbon storage estimation in tea plantation. *Ecological Informatics*, 2024, **81**: 102616

6. Hu Z H, Huang T, Zhang N, Chen C, Yang K X, Sun M Z, Yang N, Chen Y, Tao J P, Liu H, Li X H, Chen X, You X, Xiong A S, Zhuang J. Interference of skeleton photoperiod in circadian clock and photosynthetic efficiency of tea plant: in-depth analysis of mathematical model. *Horticulture Research*, 2024, **11**(10): uhae226
7. Statistics Division of the Food and Agriculture Organization of the United Nations (FAOSTAT). Crops and Livestock Products. *FAOSTAT*, 2023. Available at FAO website on December 21, 2024
8. van Bruggen A H, Gamliel A, Finckh M R. Plant disease management in organic farming systems. *Pest Management Science*, 2016, **72**(1): 30–44
9. Chen H B, Lan Y B, Fritz B K, Hoffmann W C, Liu S B. Review of agricultural spraying technologies for plant protection using unmanned aerial vehicle (UAV). *International Journal of Agricultural and Biological Engineering*, 2021, **14**(1): 38–49
10. Ye K, Hu G, Tong Z, Xu Y, Zheng J. Key intelligent pesticide prescription spraying technologies for the control of pests, diseases, and weeds: a review. *Agriculture*, 2025, **15**(1): 81
11. Bao W, Fan T, Hu G, Liang D, Li H. Detection and identification of tea leaf diseases based on AX-RetinaNet. *Scientific Reports*, 2022, **12**(1): 2183
12. Zheng Z, Zhang C. Electronic noses based on metal oxide semiconductor sensors for detecting crop diseases and insect pests. *Computers and Electronics in Agriculture*, 2022, **197**: 106988
13. Sun Y, Wang J, Sun L, Cheng S, Xiao Q. Evaluation of E-nose data analyses for discrimination of tea plants with different damage types. *Journal of Plant Diseases and Protection*, 2019, **126**(1): 29–38
14. Tang T, Luo Q, Yang L, Gao C, Ling C, Wu W. Research review on quality detection of fresh tea leaves based on spectral technology. *Foods*, 2024, **13**(1): 25
15. 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 Transactions on Geoscience and Remote Sensing*, 2024, **62**: 5601218
16. Wang J, Chen G, Ju J, Lin T, Wang R, Wang Z. Characterization and classification of urban weed species in Northeast China using terrestrial hyperspectral images. *Weed Science*, 2023, **71**(4): 353–368
17. 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. *Computers and Electronics in Agriculture*, 2019, **167**: 105039
18. 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. *Computers and Electronics in Agriculture*, 2022, **193**: 106717
19. Lu B, Dao P D, Liu J, He Y, Shang J. Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sensing*, 2020, **12**(16): 2659
20. Zhao Z Q, Zheng P, Xu S T, Wu X. Object detection with deep learning: a review. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, **30**(11): 3212–3232
21. Cao Y, Miao Q G, Liu J C, Gao L. Advance and prospects of AdaBoost algorithm. *Acta Automatica Sinica*, 2013, **39**(6): 745-758 (in Chinese)
22. Zhou L, Pan S, Wang J, Vasilakos A V. Machine learning on big data: opportunities and challenges. *Neurocomputing*, 2017, **237**: 350–361
23. Usama M, Qadir J, Raza A, Arif H, Yau K A, Elkhatib Y, Hussain A, Al-Fuqaha A. Unsupervised machine learning for networking: techniques, applications and research challenges. *IEEE Access: Practical Innovations, Open Solutions*, 2019, **7**: 65579–65615
24. Pouyanfar S, Sadiq S, Yan Y, Tian H, Tao Y, Reyes M P, Shyu M L, Chen S C, Iyengar S S. A survey on deep learning: algorithms, techniques, and applications. *ACM Computing Surveys*, 2018, **51**(5): 1–36
25. Dargan S, Kumar M, Ayyagari M R, Kumar G. A survey of deep learning and its applications: a new paradigm to machine learning. *Archives of Computational Methods in Engineering*, 2020, **27**(4): 1071–1092
26. Sarker I H. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2021, **2**(6): 420
27. Hu G, Yang X, Zhang Y, Wan M. Identification of tea leaf diseases by using an improved deep convolutional neural network. *Sustainable Computing: Informatics and Systems*, 2019, **24**: 100353
28. Liu T, Bai L, Rayhana R, Zou X, Liu Z. TAME-Faster R-CNN Model for Image-based Tea Diseases Detection. In: 2024 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). Kingston, ON, Canada: *IEEE*, 2024, 38–42
29. Xue Z, Xu R, Bai D, Lin H. YOLO-tea: a tea disease detection model improved by YOLOv5. *Forests*, 2023, **14**(2): 415
30. Shorten C, Khoshgoftaar T M. A survey on image data augmentation for deep learning. *Journal of Big Data*, 2019, **6**(1): 60
31. Li Y, Chen T, Xia F, Feng H, Ruan Y, Weng X, Weng X. TTPRNet: a real-time and precise tea tree pest recognition model in complex tea garden environments. *Agriculture*, 2024, **14**(10): 1710
32. GitHub. HumanSignal/LabelImg. *GitHub*. Available at GitHub website on September 2, 2024
33. Zou Z, Chen K, Shi Z, Guo Y, Ye J. Object detection in 20 years: a survey. *Proceedings of the IEEE*, 2023, **111**(3): 257–276
34. Gu Z, He D, Huang J, Chen J, Wu X, Huang B, Dong T, Yang

Q, Li H. Simultaneous detection of fruits and fruiting stems in mango using improved YOLOv8 model deployed by edge device. *Computers and Electronics in Agriculture*, 2024, **227**: 109512

35. Liu Y, Shao Z, Hoffmann N. Global attention mechanism: retain information to enhance channel-spatial interactions. arXiv, 2021: 2112.05561
36. Akgül İ. A pooling method developed for use in convolutional neural networks. *Computer Modeling in Engineering & Sciences*, 2024, **141**(1): 751–770
37. Ding X, Zhang X, Han J, Ding G. Diverse Branch Block: Building a Convolution as an Inception-like Unit. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: *IEEE*, 2021, 10881–10890
38. Wang C, Wu Y, Wang Y, Chen Y. Scene recognition using deep softpool capsule network based on residual diverse branch block. *Sensors*, 2021, **21**(16): 5575
39. Zhao J, Hao S, Dai C, Zhang H, Zhao L, Ji Z, Ganchev I. Improved vision-based vehicle detection and classification by optimized YOLOv4. *IEEE Access: Practical Innovations, Open Solutions*, 2022, **10**: 8590–8603
40. Zhang Y F, Ren W, Zhang Z, Jia Z, Wang L, Tan T. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing*, 2022, **506**: 146–157
41. Lin Y. Design of urban road fault detection system based on artificial neural network and deep learning. *Frontiers in Neuroscience*, 2024, **18**: 1369832
42. Majeed Y, Ojo M O, Zahid A. Standalone edge AI-based solution for Tomato diseases detection. *Smart Agricultural Technology*, 2024, **9**: 100547
43. Di Pierro B, Hank S. CPU and GPU parallel efficiency of ARM based single board computing cluster for CFD applications. *Computers & Fluids*, 2024, **272**: 106187
44. Tabassum M, Mathew K. Software Evolution Analysis of Linux (Ubuntu) OS. In: 2014 International Conference on Computational Science and Technology (ICCST). Kota Kinabalu, Malaysia: *IEEE*, 2015, 1–7
45. Imambi S, Prakash K B, Kanagachidambaresan G R. PyTorch. In: Prakash K B, Kanagachidambaresan G R, eds. Programming with TensorFlow. Cham: *Springer International Publishing*, 2021, 87–104
46. Marcel S, Rodriguez Y. Torchvision the Machine-vision Package of Torch. In: Proceedings of the 18th ACM International Conference on Multimedia. Firenze, Italy: *ACM*, 2010, 1485–1488
47. Nvidia Developer. NVIDIA TensorRT. *Nvidia Developer*. Available at Nvidia Developer website on February 15, 2025
48. Jeong E, Kim J, Tan S, Lee J, Ha S. Deep learning inference parallelization on heterogeneous processors with TensorRT. *IEEE Embedded Systems Letters*, 2022, **14**(1): 15–18
49. Hoffer E, Ben-Nun T, Hubara I, Giladi N, Hoefler T, Soudry D. Augment your batch: better training with larger batches. arXiv, 2019: 1901.09335. https://arxiv.org/abs/1901.09335
50. Li S, Zhang Z, Du F, He Y. A new automatic real-time crop row recognition based on SoC-FPGA. *IEEE Access: Practical Innovations, Open Solutions*, 2020, **8**: 37440–37452
51. Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks *via* gradient-based localization. *International Journal of Computer Vision*, 2020, **128**(2): 336–359
52. Lee D, Lee S H, Jung J H. The effects of topological features on convolutional neural networks—An explanatory analysis *via* Grad-CAM. *Machine Learning: Science and Technology*, 2023, **4**(3): 035019
53. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(6): 1137–1149
54. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y, Berg A C. SSD: Single Shot MultiBox Detector. In: Computer Vision—ECCV 2016. Cham: *Springer International Publishing*, 2016, 21–37
55. Adarsh P, Rathi P, Kumar M. YOLO v3-Tiny: Object Detection and Recognition Using One Stage Improved Model. In: 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). Coimbatore, India: *IEEE*, 2020, 687–694
56. Zhang Y, Guo Z, Wu J, Tian Y, Tang H, Guo X. Real-time vehicle detection based on improved YOLO v5. *Sustainability*, 2022, **14**(19): 12274
57. Sohan M, Sai Ram T, Rami Reddy C V. A Review on YOLOv8 and Its Advancements. In: Data Intelligence and Cognitive Informatics. Singapore: *Springer Nature Singapore*, 2024, 529–545
58. Lin J, Bai D, Xu R, Lin H. TSBA-YOLO: an improved tea diseases detection model based on attention mechanisms and feature fusion. *Forests*, 2023, **14**(3): 619
59. Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, Tang Y, Xiao A, Xu C, Xu Y, Yang Z, Zhang Y, Tao D. A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, **45**(1): 87–110
60. Chen J, Mai H, Luo L, Chen X, Wu K. Effective Feature Fusion Network in BIFPN for Small Object Detection. In: 2021 IEEE International Conference on Image Processing (ICIP). Anchorage, AK, USA: *IEEE*, 2021, 699–703
61. Qiu M, Huang L, Tang B H. ASFF-YOLOv5: multielement detection method for road traffic in UAV images based on multiscale feature fusion. *Remote Sensing*, 2022, **14**(14): 3498
62. Gevorgyan Z. SIoU loss: more powerful learning for bounding box regression. arXiv, 2022: 2205.12740
63. Ye R, Gao Q, Li T. BRA-YOLOv7: improvements on large leaf

disease object detection using FasterNet and dual-level routing attention in YOLOv7. *Frontiers in Plant Science*, 2024, **15**: 1373104

64. Ou J, Shen Y. Underwater target detection based on improved YOLOv7 algorithm with BiFusion neck structure and MPDIoU loss function. *IEEE Access: Practical Innovations, Open Solutions*, 2024, **12**: 105165–105177
65. Lyon M, Armitage P, Álvarez M A. Spatio-angular convolutions for super-resolution in diffusion MRI. arXiv, 2023: 2306.00854
66. Yang Q, Meng H, Gao Y, Gao D. A real-time object detection method for underwater complex environments based on FasterNet-YOLOv7. *Journal of Real-Time Image Processing*, 2024, **21**(1): 8
67. Wang H, Zhang G, Cao H, Hu K, Wang Q, Deng Y, Gao J, Tang Y. Geometry-aware 3D point cloud learning for precise cutting-point detection in unstructured field environments. *Journal of Field Robotics*, 2025, **42**(7): 3063–3076
68. Tang Y C, Qi S J, Zhu L X, Zhuo X R, Zhang Y Q, Meng F. Obstacle avoidance motion in mobile robotics. *Journal of System Simulation*, 2024, **36**(1): 1–26 (in Chinese)
69. Tang Y, Chen C, Leite A C, Xiong Y. Editorial: precision control technology and application in agricultural pest and disease control. *Frontiers in Plant Science*, 2023, **14**: 1163839
70. Ju J, Chen G, Lv Z, Zhao M, Sun L, Wang Z, Wang J. Design and experiment of an adaptive cruise weeding robot for paddy fields based on improved YOLOv5. *Computers and Electronics in Agriculture*, 2024, **219**: 108824
71. GitHub. Wufan-tb/AutoLabelImg. *GitHub*. Available at GitHub website on October 25, 2024
72. Kaur J, Singh W. Tools, techniques, datasets and application areas for object detection in an image: a review. *Multimedia Tools and Applications*, 2022, **81**(27): 38297–38351