Mathematical Biosciences and Engineering

Research article

Lightweight tea bud recognition network integrating GhostNet and YOLOv5

Miaolong Cao[1],*, Hao Fu[1], Jiayi Zhu[1] and Chenggang Cai[2]

[1] School of Mechanical and Energy Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China

[2] School of Biological and Chemical Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China

* **Correspondence:** Email: cml_q@163.com.

Abstract: Aiming at the problems of low detection accuracy and slow speed caused by the complex background of tea sprouts and the small target size, this paper proposes a tea bud detection algorithm integrating GhostNet and YOLOv5. To reduce parameters, the GhostNet module is specially introduced to shorten the detection speed. A coordinated attention mechanism is then added to the backbone layer to enhance the feature extraction ability of the model. A bi-directional feature pyramid network (BiFPN) is used in the neck layer of feature fusion to increase the fusion between shallow and deep networks to improve the detection accuracy of small objects. Efficient intersection over union (EIOU) is used as a localization loss to improve the detection accuracy in the end. The experimental results show that the precision of GhostNet-YOLOv5 is 76.31%, which is 1.31, 4.83, and 3.59% higher than that of Faster RCNN, YOLOv5 and YOLOv5-Lite respectively. By comparing the actual detection effects of GhostNet-YOLOv5 and YOLOv5 algorithm on buds in different quantities, different shooting angles, and different illumination angles, and taking F1 score as the evaluation value, the results show that GhostNet-YOLOv5 is 7.84, 2.88, and 3.81% higher than YOLOv5 algorithm in these three different environments.

Keywords: YOLOv5; tea bud detection; deep learning; GhostNet; coordinate attention

1. Introduction

With the continuous improvement of people's quality of life, tea has become an indispensable drink in people's leisure life. Therefore, to further improve the influence of tea culture, it is necessary to improve the control of tea quality [1]. The processing technology of tea is divided into picking, killing, crushing, initial drying, shaping, and drying. Among them, the quality of tea picking will directly affect the economic benefits. At present, tea picking is mainly divided into manual and mechanical picking [2]. Manual picking is mainly based on workers' experience and according to the differences in the color and shape of tea buds, which is too time-consuming, labor-intensive, and costly [3]. Mechanical picking solves the problem of slow tea picking, but it can not accurately identify the buds. Therefore, it is urgent to study the detection algorithm of tea buds.

The detection algorithm based on machine vision has always been a research hotspot [4]. In the field of tea detection, different detection algorithms have different usages [5]. There are detection methods for tea leaf diseases, tea bud leaves, and tea sprout's picking points. Mukhopadhyay et al. [6] presented a novel approach for automatically detecting tea leaves diseases based on image processing technology. Yang et al. [7] developed an effective, simple, apt computer vision algorithm to detect tea disease area using infrared thermal image processing techniques and to estimate tea disease. Karunasena et al. [8] proposed a new method for tea bud detection using a cascade classifier, which carried out the detection of tea buds by combining histogram of oriented gradient features and support vector machine classification. Zhang et al. [9] proposed a method to obtain information on the picking point on the basis of the Shi-Tomasi algorithm. The test identified 1042 effective shoots for tender buds, and 887 picking points were marked, with a success rate of 85.12%. These methods all provide a theoretical basis for the automatic picking of tea leaves.

Subsequently, object detection algorithms based on deep learning rose rapidly and then evolved into two-stage network models such as fast region-based convolutional network method (R-CNN) [10], faster R-CNN [11], and one-stage network models such as YOLO [12], single shot multibox detector [13]. These algorithms have achieved certain applications in the agricultural field [14]. Lawal et al. [15] proposed a modified YOLOv3 model called YOLO-Tomato model to detect tomatoes under complex environmental conditions, and its performance is superior to other state-of-the-art methods. Roy et al. [16] presented a high-performance real-time fine-grained object detection framework for plant disease detection. At a detection rate of 70.19 FPS, the proposed model obtained a precision value of 90.33%, F1 score of 93.64%, and a mean average precision value of 96.29%. It can be seen that the detection effect of deep learning on general crops is excellent. However, using deep learning to detect tea buds in complex background is difficult, so there are few related studies. Yang et al. [17] proposed a complete solution, including the mechanical structure, the visual recognition system and the motion control system of the high-quality tea automatic plucking robot. Tao et al. [18] proposed a method of tea picking point location based on mask RCNN. By training the tea picking point location model and identifying the segmentation of tea buds, the coordinates of tea picking points can be located. Li et al. [19] proposed a real-time tea bud detection method using the channel and layer pruned YOLOv3-SPP deep learning algorithm. The test results show that the number of parameters, model size, and inference time of the tea bud detection model after compression reduced by 96.82, 96.81, and 59.62% respectively, while the mean average precision of the model is only 0.40% lower than that of the original model. The above research provides a basis for the study of tea bud detection under complex background.

Xu et al. [20] proposed a detection and classification approach of a two-level fusion network with a variable universe to solve the detection and classification problems of different grades of tea. This has established the foundation for realizing the automation of famous and excellent tea picking, but it lacks the consideration of actual environmental factors. Considering that tea bud picking needs certain rapidity and accuracy, this paper takes the YOLOv5 model as the basic framework and improves the backbone layer based on GhostNet, to reduce the amount of calculation and shorten the detection speed. The coordinate attention mechanism is introduced to improve the recognition accuracy of small targets such as buds. EIOU is used as the loss function of the positioning box. Through the self-made data set, the algorithm designed in this paper is compared with the original algorithm. The paper is organized as follows: Section 2 introduces the YOLO framework; Section 3 describes the proposed GhostNet-YOLOv5 model; Section 4 discusses the experimental results and comparative analysis of the proposed object detection model. Finally, the conclusions and prospects of the current work have been discussed in Section 5.

2. Related works

YOLO, fully known as "you only look once", is an object detection algorithm based on a deep neural network. Its biggest feature is that it runs very fast and can be used in real-time systems. Different from the two-stage of R-CNN, YOLO uses the one-stage platform, that is, end-to-end prediction [21]. It can directly predict the categories and locations of different targets using only one CNN network.

At present, the YOLO algorithm has been upgraded in five versions: YOLOv1, YOLOv2, YOLOv3, YOLOv4, and YOLOv5. The improvement ideas of predecessors are mainly divided into the following four parts:

1) Input: in the model training stage, some improvement measures are proposed, such as mosaic data enhancement, adaptive image scaling, adaptive anchor box calculation;

2) Feature extraction layer: optimize the main detection algorithms, such as focus structure and CSP structure;

3) Feature fusion layer: FPN structure is used in the early stage, and FPN+PANet [22] structure is used in the later stage;

4) Output layer: the main improvement is the loss function GIOU-Loss [23] and non-maximum suppression (NMS) [24].

The overall framework of the current YOLOv5 algorithm is shown in Figure 1. Generally, it is divided into four modules, including the input layer, backbone layer, neck layer, and output layer. Today, there are four versions of YOLOv5, including YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. Considering the model size, this paper gives priority to the fastest YOLOv5s. On this basis, the network is improved to improve the accuracy of object detection.

YOLOv5 is indeed a detection algorithm with fast detection speed and high accuracy. It performs well on some open source data sets, but it still needs to be improved on the detection task of tea buds. Aiming at the problems of complex background, fuzzy color, and small size of tea buds, this paper improves and optimizes based on YOLOv5, and proposes a tea bud detection algorithm GhostNet-YOLOv5. Experimental and test results show the effectiveness of the algorithm.
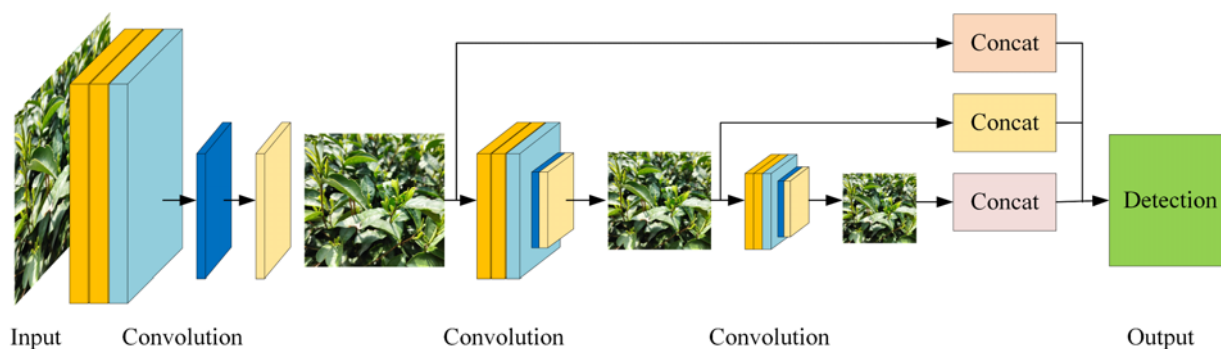
Figure 1. Network structure of YOLOv5.

The YOLOv5 network evolved from the YOLO model. Compared with the faster R-CNN network, the YOLO network model turns the detection problem into a regression problem [25]. It does not need the suggested region, and directly generates the bounding box coordinates and probabilities of each class through the regression [26]. Compared with R-CNN, the detection speed has been greatly improved. The detection principle of YOLO is shown in Figure 2. First, divide the input picture into s * s cells. If the center of the detected target falls into the specified cell, the cell is responsible for detecting the target. Assuming that each cell produces B anchor boxes, the target location and category prediction can be expressed by the tensor of $s \times s \times b \times (4 + 1 + C)$, where 4 represents the coordinates (x, y), width and height (w, h), 1 represents the confidence score, and C represents the category number. The position, confidence, and category information of the final prediction target can be obtained by continuous regression of the real bounding box through training.
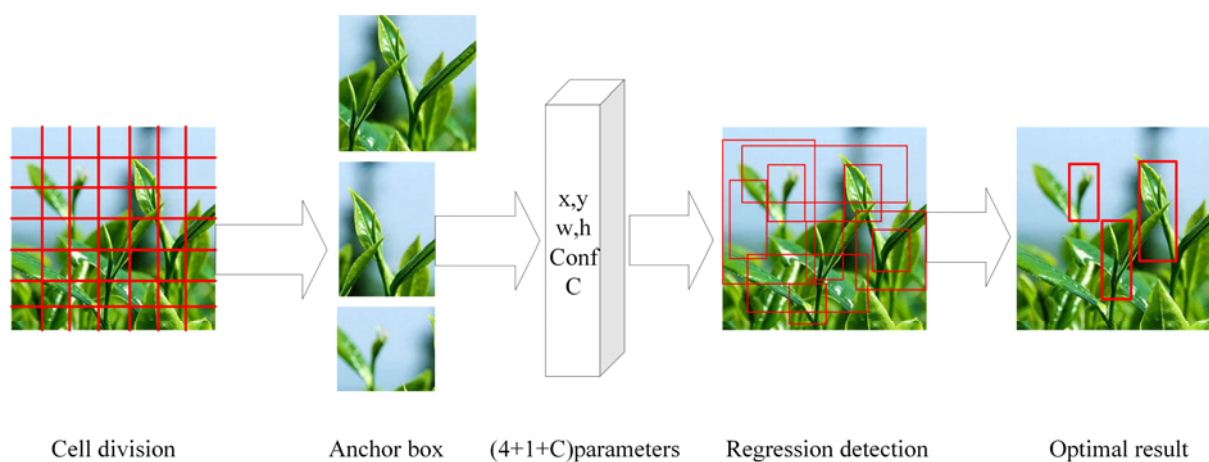

Figure 2. Detection principle of YOLO.

3. Methodology

3.1. GhostNet network

A Ghost module is designed in GhostNet to replace general convolution [27]. The Ghost

module integrates the general convolution into two parts. First, 1×1 convolution is used to obtain the necessary feature concentration of the input features. Secondly, the deep separable convolution is performed. This deep separable convolution is layer by layer convolution, and a similar feature map is generated by using the feature concentration obtained in the previous step. The Ghost module is shown in Figure 3.
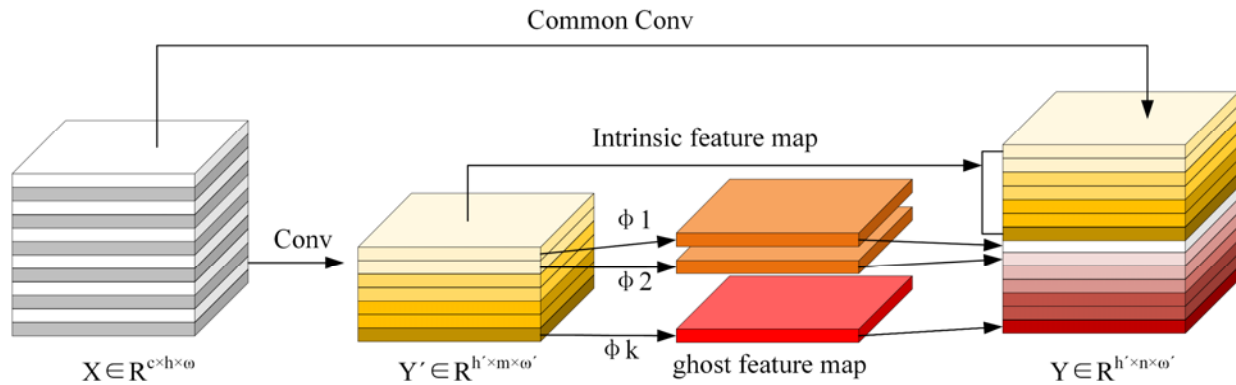

Figure 3. Schematic diagram of Ghost module.

The algorithm and calculation amount of general convolution can be expressed as:

$$Y = X \times f + b \tag{1}$$

$$FLOP_{s1} = n \times h' \times \omega' \times c \times k \times k \tag{2}$$

where, X represents a feature map with channel number c, height h and width w, Y represents a feature map with channel number n, height h and width w, f represents n convolution kernels with size k * k, and b represents a bias term.

The algorithm and calculation amount of ghost module can be expressed as:

$$\begin{cases} Y' = X \times f' \\ Y_{ij} = \varphi_{i,j}(Y'_i), i \in (1, m), j \in [1, s] \end{cases} \tag{3}$$

$$FLOP_{s2} = m \times h' \times \omega' \times c \times k \times k + (n - m) \times h' \times \omega' \times c \times d \times d \tag{4}$$

where, Y' is the m intrinsic feature maps of input x for general convolution output, f' represents m convolution kernels of size k × k, Y'_i represents the i-th intrinsic feature map in Y', and Y_{ij} represents the j-th ghost feature map generated by convolution kernel linear transformation of d × d, and m ≤ n, k = d.

The Ghost module first generates a part of the original feature map by ordinary convolution, and then convolutes these feature maps one by one to obtain another part of the corresponding feature map, and splices this part with the original feature map. It can be seen from the principle of the algorithm that the convolution layers after splicing are the same as the original ones, but compared with flops, it can be seen that the calculation amount of the network is reduced, and the improved convolution structure is shown in Figure 4.
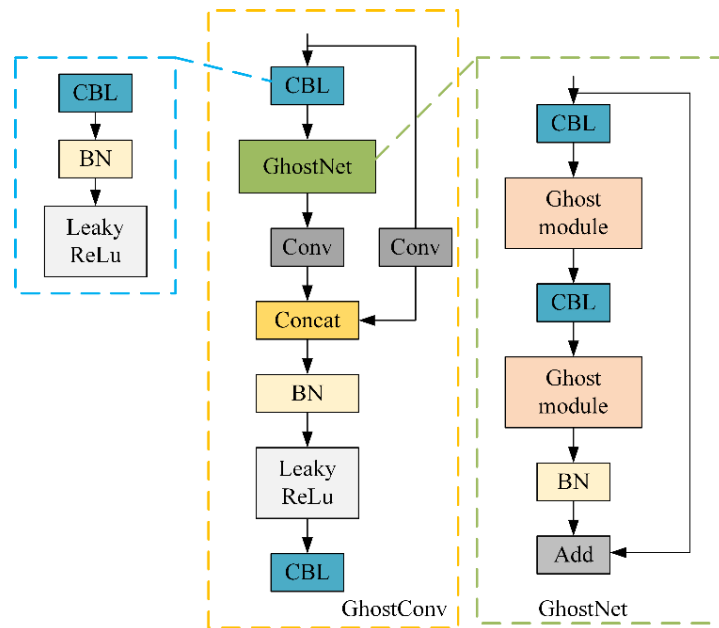
Figure 4. Schematic diagram of improved convolution.

3.2. Coordinate attention

Attention mechanism comes from the study of human vision. In cognitive science, due to the bottleneck of information processing, human beings will selectively pay attention to a part of all information and ignore other visible information [28]. These mechanisms are often referred to as attention mechanisms. To make rational use of the limited visual information processing resources, human beings need to select a specific part of the visual region and then focus on it. At present, the commonly used attention mechanisms are sequence and exception (SE) and convolution block attention module (CBAM). However, SE only considers remeasuring the importance of each channel by modeling the channel relationship and ignores the location information, which is very important for generating a spatial selective attention map. On the contrary, CBAM can not capture the long-range dependence on the feature map, although it introduces local location information into the channel.

As a new and efficient attention module, coordinate attention (CA) [29] is mainly divided into two steps: coordinate information embedding and coordinate attention generation. Firstly, to mitigate the loss of position information caused by 2D global pooling, the channel attention is decomposed into two parallel (X and Y direction) one-dimensional feature coding processes, and the spatial coordinate information is effectively integrated into the generated attention map. Then, the two feature maps embedded with specific direction information are coded into two attention maps respectively, and each attention map captures the long-range dependence of the input feature map along one spatial direction. Specifically, for input x, the pooled cores with dimensions (H, 1) and (1, W) are used to encode each channel along the horizontal and vertical coordinate directions. The output formula of the c-th channel with height h is as follows:

$$z_c^h(h) = \frac{1}{W} \sum_{i=0}^{W-1} x_c(h,i) \tag{5}$$

Accordingly, the output of the c-th channel with width w is as follows:

$$z_c^w(w) = \frac{1}{H}\sum_{j=0}^{H-1} x_c(j,w) \tag{6}$$

Secondly, after the transformation in the information embedding, the fusion operation is carried out, and then 1×1 convolution transform function F1 performs transform operation on it:

$$f = \delta\left(F_1([z^h, z^w])\right) \tag{7}$$

After normalization and nonlinear processing, f is divided into two independent tensors f^h and f^w along the spatial dimension, and then f^h and f^w are transformed into the same number of channels as X by two 1×1 convolutions, as shown in the following formula.

$$g^h = \sigma\left(F_h(f^h)\right) \tag{8}$$

$$g^w = \sigma\left(F_w(f^w)\right) \tag{9}$$

where, σ represents sigmoid activation function.

Finally, the output y of coordinate attention can be written as:

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{10}$$

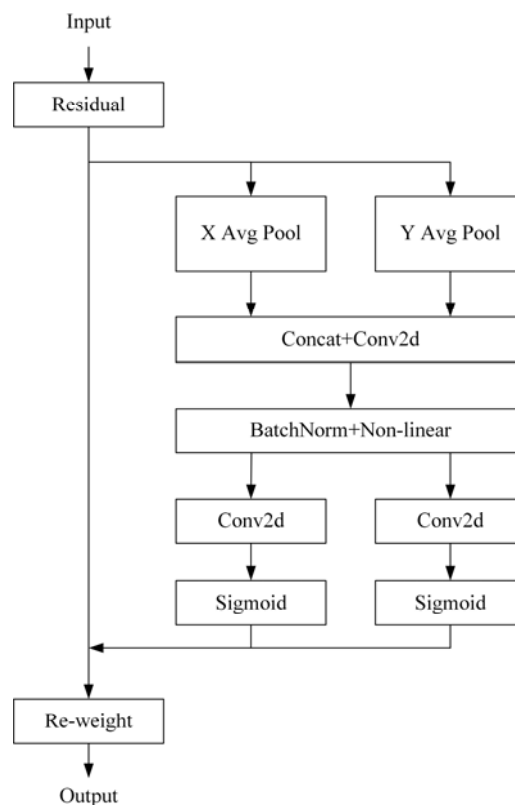The coordinate attention module is shown in Figure 5.

Figure 5. Coordinate attention model.

3.3. BiFPN

The original YOLOv5 backbone network adopts FPN + PAN structure. This paper retains the original network architecture, and also extracts the features of the upper, middle and bottom feature layers, and then transmits them to BiFPN to strengthen the feature extraction network. Since the BiFPN network has five input layers, the BiFPN network is simplified to three input layers to integrate with YOLOv5 and reduce the amount of calculation. The improved BiFPN network structure is shown in Figure 6.
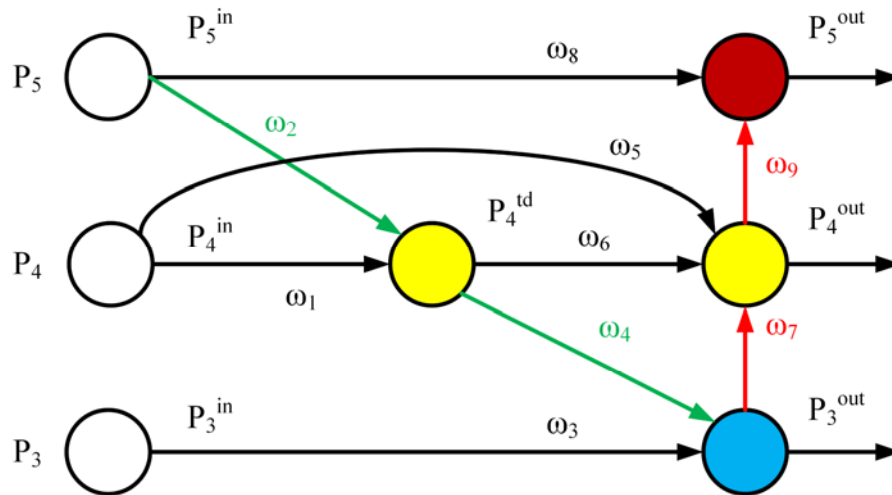

Figure 6. Improved BiFPN structure.

The calculation formula of the output of each fusion node after the improved network is as follows.

$$P_4^{td} = conv\left(\frac{\omega_1 \cdot P_4^{in} + \omega_2 \cdot resize\left(P_5^{in}\right)}{\omega_1 + \omega_2 + \varepsilon}\right) \quad (11)$$

$$P_3^{out} = conv\left(\frac{\omega_3 \cdot P_3^{in} + \omega_4 \cdot resize\left(P_4^{td}\right)}{\omega_3 + \omega_4 + \varepsilon}\right) \quad (12)$$

$$P_4^{out} = conv\left(\frac{\omega_5 \cdot P_4^{in} + \omega_6 \cdot P_4^{td} + \omega_7 \cdot resize(P_3^{out})}{\omega_5 + \omega_6 + \omega_7 + \varepsilon}\right) \quad (13)$$

$$P_5^{out} = conv\left(\frac{\omega_8 \cdot P_5^{in} + \omega_9 \cdot resize(P_4^{out})}{\omega_8 + \omega_9 + \varepsilon}\right) \quad (14)$$

where, P_i^{in}, P_i^{out} represent input and output feature layer respectively. *conv* represents convolution, *resize* represents upsampling or downsampling of the input, $\omega_i \geq 0$ is a learnable weight, and ε is a small quantity to ensure numerical stability.

Compared with previous networks, the YOLOv5 network has many improvements in recognition accuracy and model size. However, the identification of small targets such as tea buds is still somewhat difficult. Therefore, in this paper, GhostConv is used to replace the ordinary convolution module, a CA module is used in the backbone layer, and the BiFPN structure is used in the neck layer to replace the original FPN + PAN structure. The improved YOLOv5 structure is shown in Figure 7. The improved parts are indicated by red boxes.
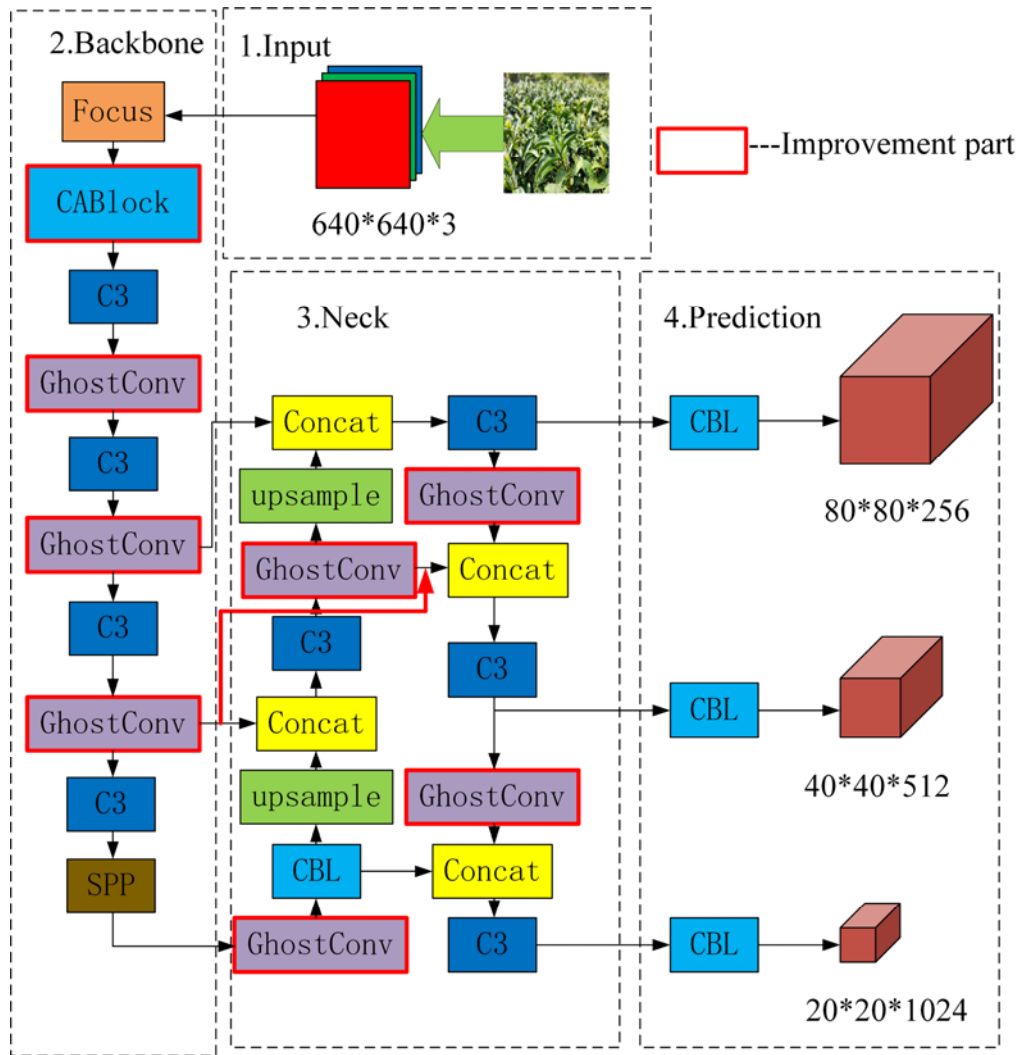
Figure 7. The framework of GhostNet-YOLOv5.

3.4. Loss function

In the original YOLOv5 algorithm, the loss function is mainly composed of three parts: classification loss, confidence loss, and bounding box loss [30]. The cross entropy loss function is used for classification loss and target loss, while generalized intersection over union (GIOU) loss is the positioning loss of the bounding box, loss keeps the main properties of IOU and avoids the shortcomings of IOU. It not only pays attention to overlapping areas but also to other non-overlapping areas, which can well reflect the coincidence degree between the two.

GIOU first calculates the minimum closure area C of the predicted frame and the real frame and then calculates the proportion of the closure area that does not belong to the two frames in the total area. Finally, GIOU is obtained by subtracting this proportion from IOU, as shown in the following formula.

$$IOU = \left| \frac{B \cap B^{gt}}{B \cup B^{gt}} \right| \tag{15}$$

where, B and B^{gt} represent prediction box and real box respectively.

$$GIOU = IOU - \frac{|C \backslash (A \cup D)|}{|C|} \tag{16}$$

where, A represents the prediction box and D represents the real box, both of which are tensors. C represents the minimum area surrounding A and D.

CIOU loss considers the overlapping area, center point distance and aspect ratio of bounding box regression. However, the difference of aspect ratio reflected by V in its formula, rather than the real difference between width and height and their confidence, sometimes hinders the effective optimization similarity of the model. According to the characteristics of tea bud with high aspect ratio and small target, EIOU is introduced to integrate the overlapping part, center distance and aspect ratio of the two frames into the calculation of loss function.

$$L_{EIOU} = L_{IOU} + L_{dis} + L_{asp} = 1 - IOU + \frac{\rho^2(b,b^{gt})}{c^2} + \frac{\rho^2(\omega,\omega^{gt})}{C_\omega^2} + \frac{\rho^2(h,h^{gt})}{C_h^2} \tag{17}$$

where, c_h and c_w are the height and width of the smallest circumscribed rectangle covering the real box of the prediction box, respectively.

4. Experiments results

4.1. Dataset

The shooting place is WuChao mountain tea garden in Hangzhou City, Zhejiang Province. The shooting time is during the Qingming Festival, which is the best time for picking tea. The shooting equipment selects the mobile camera, selects 1000 pieces of collected images as the data set for tea bud detection, and labels the pictures through labelimg. Then, the training is conducted according to the ratio of 8:2 to verify.

4.2. Experimental platform

This experiment is carried out under the pytorch framework. The configuration parameters of software and hardware are shown in Table 1.

Table 1. Software and hardware platform configuration parameters.

Configuration	Parameter
(CPU)	AMD R5-4600H
(GPU)	NVIDIA GTX1650
Operation system	Windows10
CUDA	10.1
Opencv	3.4.6

4.3. Evaluation indicators

There are many evaluation criteria for deep learning, such as accuracy, confusion matrix, precision, recall, average precision, mean average precision (map), intersection union ratio (IOU),

non maximum mechanism (NMS), etc. This paper mainly introduces the accuracy rate, recall rate and map as the evaluation criteria for identifying tea buds. Next, it will be introduced separately.

Precision (P) is the ratio of true position in the recognized image.

$$Precision = \frac{t_p}{t_p+f_p} = \frac{t_p}{n} \tag{18}$$

Recall (R) is the proportion of the left and right positive samples in the test set that are correctly identified as positive samples.

$$Recall = \frac{t_p}{t_p+f_n} \tag{19}$$

where, t_p refers to the result of correctly identifying the category of tea buds, f_p refers to the result of incorrectly identifying the category of tea buds, and f_n refers to the result of failing to detect the category of tea buds.

Since the accuracy and recall may be contradictory in some cases, this test defines F1 score as the comparative evaluation index, and the formula is as follows:

$$F1 = \frac{2*P*R}{P+R} \tag{20}$$

4.4. Data comparison

In order to verify the detection performance of the algorithm, the improved method is compared with the original YOLOv5 and the mainstream object detection algorithms such as deep learning YOLOv5-Lite and Faster RCNN in the loss value and recall rate. The results are shown in Figure 8.
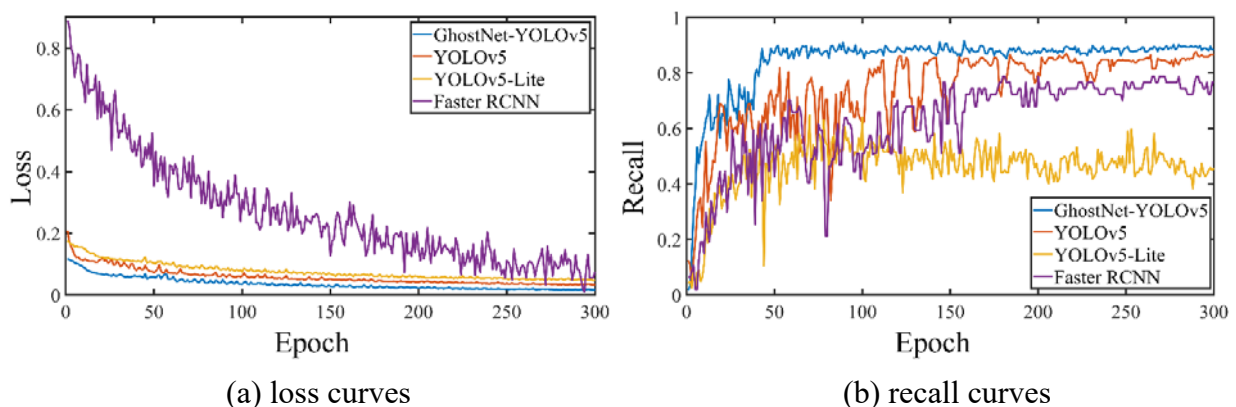
(a) loss curves   (b) recall curves

Figure 8. Loss curves and recall curves.

The training loss curves obtained from Faster RCNN, YOLOv5, YOLOv5-Lite and GhostNet-YOLOv5 have been compared in Figure 8(a). At the initial stage of the loss curve, the Faster RCNN loss begins to reduce significantly after approximately 250 epochs, whereas, for YOLOv5-Lite and YOLOv5, loss reduction happens after approximately 50 epochs. However, the loss in the GhostNet-YOLOv5 rapidly decreases within approximately 30 epochs indicating the best convergence characteristics. After 200 epochs, the loss curve becomes stable and the model gradually tends to

converge. It can be seen from Figure 8(b) that after 50 epochs, the recall rate of GhostNet-YOLOv5 is basically stable at 85%, while the YOLOv5-Lite has been floating between 40% and 60%. For YOLOv5 and Faster RCNN, the recall rate were stable at 80 and 75% after 250 epochs. The specific data comparison is shown in Table 2.

Table 2. Comparative experimental test results of different networks.

Algorithm	P (%)	R (%)	Parameters	Weight size (MB)
Faster RCNN	75	76.6	573,571,072	22
YOLOv5	71.48	86.61	7,063,542	15
YOLOv5-Lite	72.72	45.9	1,542,966	6
GhostNet-YOLOv5	76.31	88.42	5,936,737	10

In order to verify the effectiveness of the improved YOLOv5 network model in detecting tea buds, various modules are compared and tested, and the results are shown in Table 3.

Table 3. Comparison results before and after adding different modules.

Algorithm	GhostNet	CA	BiFPN	P (%)	Time (s)
YOLOv5				71.48	0.078
	√			71.97	0.047
	√	√		73.65	0.047
	√	√	√	76.31	0.063

It can be seen that the detection time of the model with GhostNet alone is significantly shortened by 0.031 s, while the detection time of the model with coordinate mechanism is increased by 1.68% on map. Finally, although the detection time of GhostNet-YOLOv5 is slightly improved after the improvement of BiFPN, it is still reduced by 0.015 s compared with the original YOLOv5, and the overall accuracy is improved by 4.83%.

To further verify the effectiveness of the model, it is necessary to detect the efficiency of the algorithm in various actual environments. This paper will take the number of buds, shooting angle, and illumination angle as control variables, compare the detection effect with the original algorithm and evaluate the performance with F1 score. Thirty images will be randomly selected for each category from the test set as F1 score evaluation.

In the actual tea image, there are often different numbers of bud objects, which will have different effects on the detection. For example, in a single bud image, the buds occupy a large area, the outline is complete and clear, and the recognition difficulty is low. In a multi-target image, due to the reduction of size and the increase of number, the recognition object will be blocked, so the recognition difficulty will increase. Therefore, set the comparison test of bud detection under different numbers, including 1–4 buds, 5–10 buds, and more than 10 buds. Compare the detection performance of the two algorithms under different numbers, as shown in Figure 9.

Figure 9. Detection results of two algorithms under different numbers.

Two different algorithms are used to detect the number of positive samples, the number of total samples, the number of undetected samples and the number of false samples. The accuracy and recall are calculated to obtain F1 scores. Finally, the three results are averaged to obtain the comprehensive effect, and the final results are shown in Table 4.

Table 4. Comparison results of different numbers of buds.

numbers	algorithm	F1 (%)
1~4	YOLOv5	73.68
	GhostNet-YOLOv5	84.21
5~10	YOLOv5	72.72
	GhostNet-YOLOv5	81.48
>10	YOLOv5	79.71
	GhostNet-YOLOv5	83.91
Comprehensive results	YOLOv5	75.37
	GhostNet-YOLOv5	83.21

The Figure 10 shows the detection results of the two algorithms at different shooting angles.
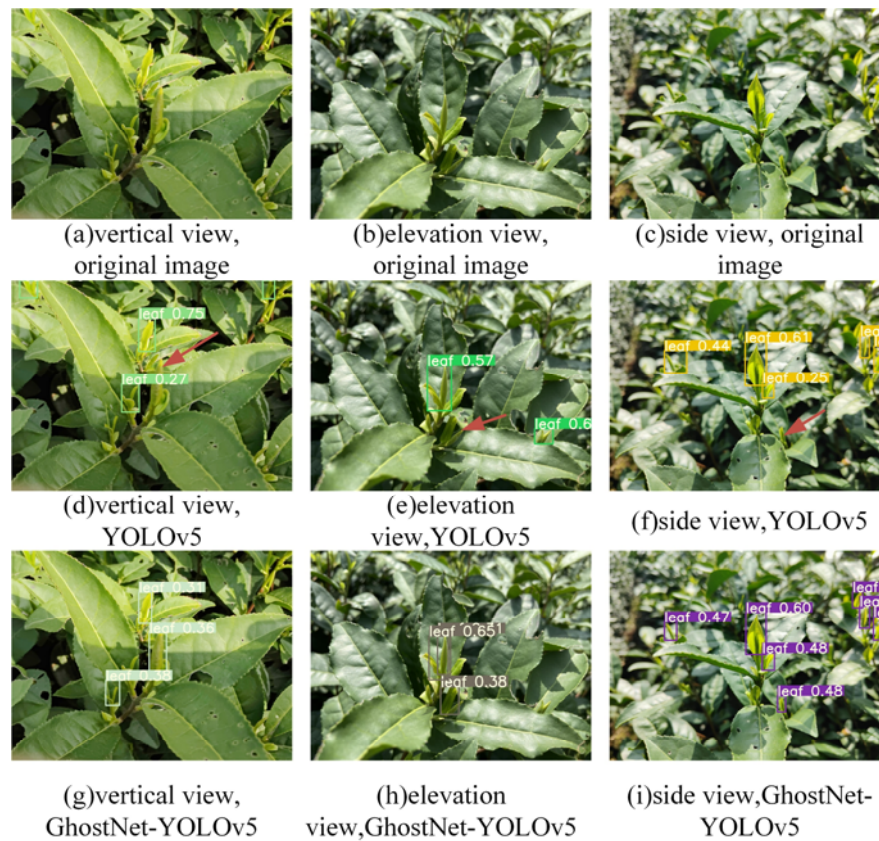
Figure 10. Detection results of two algorithms under different shooting angles.

As can be seen from Figure 10, the shape characteristics of tea buds are obvious when facing up. Therefore, both algorithms can be well detected. However, when looking down and sideways, the original YOLOv5 was seriously missed. The test adopts the same method as the previous section, and repeats the random sampling data set for different shooting angles. The statistical results are shown in Table 5.

Table 5. Comparison results of different shooting angles.

Shooting angle	algorithm	F1 (%)
Elevation view	YOLOv5	82.19
	GhostNet-YOLOv5	86.11
Side view	YOLOv5	78.87
	GhostNet-YOLOv5	80.55
Vertical view	YOLOv5	73.84
	GhostNet-YOLOv5	76.92
Comprehensive results	YOLOv5	78.31
	GhostNet-YOLOv5	81.19

As an external influence factor, the illumination angle is equally important. In this test, forward light, reverse light and side light are selected for comparative analysis. Figure 11 shows the detection results of the two algorithms under different illumination angles.

Figure 11. Detection results of two algorithms under different illumination angles.

It can be seen from Figure 11 that the brightness of the buds is enhanced in the forward light, and the outline of the buds is obvious in the reverse light, although there are some shadows, while the buds and branches and leaves are easy to distinguish under the side light. According to the sampling method in the previous section, select three gradient data sets for comparison, and the statistical results are shown in Table 6.

Table 6. Comparison results of different illumination angles.

Illumination angles	algorithm	F1 (%)
Side light	YOLOv5	80.64
	GhostNet-YOLOv5	83.87
Direct light	YOLOv5	74.57
	GhostNet-YOLOv5	77.96
Backlight	YOLOv5	70.17
	GhostNet-YOLOv5	75.01
Comprehensive results	YOLOv5	75.13
	GhostNet-YOLOv5	78.94

To sum up, the three groups of comparative tests have fully proved the effectiveness and real-time of tea bud detection based on GhostNet-YOLOv5 algorithm.

5. Conclusions

In this paper, a tea bud detection model integrating YOLOv5 and GhostNet is proposed. To improve the detection efficiency, the GhostNet module is used to replace the CSP module in the backbone layer first, and a coordinate attention mechanism is added to enhance the ability of feature extraction. To enhance multi-scale fusion, BiFPN is introduced. In addition, EIOU is used as the optimized loss function. The experimental results show that the precision of this model is improved by 4.83, 3.59, and 1.31% respectively compared with YOLOv5, YOLOv5-Lite, and Faster RCNN, and the recall rate is increased by 1.81, 42.52, and 11.82%, respectively. However, under the influence of different numbers, shooting angles, and illumination angles in the actual environment, the F1 score of GhostNet-YOLOv5 is 7.84, 2.88, and 3.81% higher than that of YOLOv5. In the future, I will further optimize the network model.

Acknowledgments

Conflict of interest

The authors declare that they have no conflict of interest in this work.

References

1. X. L. Yu, D. W. Sun, Y. He, Emerging techniques for determining the quality and safety of tea products: A review, *Compr. Rev. Food Sci. Food Saf.*, **19** (2020), 2613–2638. https://doi.org/10.1111/1541-4337.12611

2. C. Chen, J. Lu, M. Zhou, J. Yi, M. Liao, Z. Gao, A YOLOv3-based computer vision system for identification of tea buds and the picking point, *Comput. Electron. Agric.*, **198** (2022), 107116. https://doi.org/10.1016/j.compag.2022.107116

3. N. Gan, M. F. Sun, C. Y. Lu, M. H. Li, Y. J. Wang, Y. Song, et al., High-speed identification system for fresh tea leaves based on phenotypic characteristics utilizing an improved genetic algorithm, *J. Sci. Food Agric.*, **2022** (2022). https://doi.org/10.1002/jsfa.12047

4. Z. Huang, Y. Li, T. Zhao, P. Ying, Y. Fan, J. Li, Infusion port level detection for intravenous infusion based on Yolo v3 neural network, *Math. Biosci. Eng.*, **18** (2021), 3491–3501. https://doi.org/10.3934/mbe.2021175

5. M. Cao, J. Zhu, J. Zhang, S. Cao, M. Pang, Orthogonal optimization for effective classification of different tea leaves by a novel pressure stabilized inclined chamber classifier, *J. Food Process Eng.*, **2022** (2022), e14141. https://doi.org/10.1111/jfpe.14141

6. S. Mukhopadhyay, M. Paul, R. Pal, D. De, Tea leaf disease detection using multi-objective image segmentation, *Multimedia Tools Appl.*, **80** (2021), 753–771. https://doi.org/10.1007/s11042-020-09567-1

7. N. Yang, M. F. Yuan, P. Wang, R. B. Zhang, J. Sun, H. P. Mao, Tea diseases detection based on fast infrared thermal image processing technology, *J. Sci. Food Agric.*, **99** (2019), 3459–3466. https://doi.org/10.1002/jsfa.9564

8. G. M. K. B. Karunasena, H. Priyankara, Tea bud leaf identification by using machine learning and image processing techniques, *Int. J. Sci. Eng. Res.*, **10** (2020). https://doi.org/10.14299/ijser.2020.08.02

9. L. Zhang, L. Zou, C. Y. Wu, J. N. Chen, H. P. Chen, Locating famous tea's picking point based on Shi-Tomasi algorithm, *CMC-Comput. Mater. Continua*, **69** (2021), 1109–1122. https://doi.org/10.32604/cmc.2021.016495

10. R. Girshick, Fast R-CNN, in *2015 IEEE International Conference on Computer Vision (ICCV)*, (2015), 1440–1448. https://doi.org/10.1109/ICCV.2015.169

11. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (2017), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

12. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 779–788. https://doi.org/10.1109/CVPR.2016.91

13. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, et al., SSD: Single shot MultiBox detector, in *Computer Vision—ECCV 2016*, (2016), 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

14. A. M. Roy, J. Bhaduri, Real-time growth stage detection model for high degree of occultation using DenseNet-fused YOLOv4, *Comput. Electron. Agric.*, **193** (2022), 106694. https://doi.org/10.1016/j.compag.2022.106694

15. M. O. Lawal, Tomato detection based on modified YOLOv3 framework, *Sci. Rep.*, **11** (2021). https://doi.org/10.1038/s41598-021-81216-5

16. A. M. Roy, R. Bose, J. Bhaduri, A fast accurate fine-grain object detection model based on YOLOv4 deep neural network, *Neural Comput. Appl.*, **34** (2022), 3895–3921. https://doi.org/10.1007/s00521-021-06651-x

17. H. L. Yang, L. Chen, Z. B. Ma, M. T. Chen, Y. Zhong, F. Deng, et al., Computer vision-based high-quality tea automatic plucking robot using Delta parallel manipulator, *Comput. Electron. Agric.*, **181** (2021), 105946. https://doi.org/10.1016/j.compag.2020.105946

18. O. M. Lawal, Development of tomato detection model for robotic platform using deep learning, *Multimedia Tools Appl.*, **80** (2021), 26751–26772. https://doi.org/10.1007/s11042-021-10933-w

19. Y. T. Li, L. Y. He, J. M. Jia, J. N. Chen, J. Lyu, C. A. Y. Wu, High-efficiency tea shoot detection method via a compressed deep learning model, *Int. J. Agric. Biol. Eng.*, **15** (2022), 159–166. https://doi.org/10.25165/j.ijabe.20221503.6896

20. W. Xu, L. Zhao, J. Li, S. Shang, X. Ding, T. Wang, Detection and classification of tea buds based on deep learning, *Comput. Electron. Agric.*, **192** (2022), 106547. https://doi.org/10.1016/j.compag.2021.106547

21. J. Redmon, A. Farhadi, YOLO9000: Better, faster, stronger, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), 6517–6525. https://doi.org/10.1109/CVPR.2017.690

22. M. Tan, R. Pang, Q. V. Le, EfficientDet: Scalable and efficient object detection, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2020), 10778–10787. https://doi.org/10.1109/CVPR42600.2020.01079

23. X. Dong, S. Yan, C. Duan, A lightweight vehicles detection network model based on YOLOv5, *Eng. Appl. Artif. Intell.*, **113** (2022), 104914. https://doi.org/10.1016/j.engappai.2022.104914

24. A. Neubeck, L. V. Gool, Efficient non-maximum suppression, in *18th International Conference on Pattern Recognition (ICPR'06)*, (2006), 850–855. https://doi.org/10.1109/ICPR.2006.479

25. Z. Wang, L. Jin, S. Wang, H. Xu, Apple stem/calyx real-time recognition using YOLO-v5 algorithm for fruit automatic loading system, *Postharvest Biol. Technol.*, **185** (2022), 111808. https://doi.org/10.1016/j.postharvbio.2021.111808

26. M. P. Mathew, T. Y. Mahesh, Leaf-based disease detection in bell pepper plant using YOLO v5, *Signal Image Video Process.*, **16** (2022), 841–847. https://doi.org/10.1007/s11760-021-02024-y

27. K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, C. Xu, GhostNet: More features from cheap operations, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2020), 1577–1586. https://doi.org/10.1109/CVPR42600.2020.00165

28. A. Pandey, K. Jain, A robust deep attention dense convolutional neural network for plant leaf disease identification and classification from smart phone captured real world images, *Ecol. Inf.*, **70** (2022), 101725. https://doi.org/10.1016/j.ecoinf.2022.101725

29. S. Yi, J. Li, X. Liu, X. Yuan, CCAFFMNet: Dual-spectral semantic segmentation network with channel-coordinate attention feature fusion module, *Neurocomputing*, **482** (2022), 236–251. https://doi.org/10.1016/j.neucom.2021.11.056

30. D. Yuan, X. Shu, N. N. Fan, X. J. Chang, Q. Liu, Z. Y. He, Accurate bounding-box regression with distance-IoU loss for visual tracking, *J. Visual Commun. Image Represent.*, **83** (2022), 103428. https://doi.org/10.1016/j.jvcir.2021.103428