YOSBG: UAV image data-driven high-throughput field tobacco leaf counting method

Hengshu Xiao
Yunnan University
Junying Li (ljy1250@163.com)
Yunnan Academy of Tobacco Agriculture Science
Hong Liang
Yunnan University
Liuchen Zhang
Yunnan Academy of Tobacco Agriculture Science
Hong Zhang
Yunnan Academy of Tobacco Agriculture Science
Erdeng Ma
Yunnan Academy of Tobacco Agriculture Science
Xiaopeng Deng
Yunnan Academy of Tobacco Agriculture Science

Research Article

Keywords: Plant phenotype, Plant leaf detection, tobacco leaf count, UAV, YOLOv8, BiFPN, High-throughput

Posted Date: December 4th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3686207/v1

Additional Declarations: No competing interests reported.

YOSBG: UAV image data-driven high-throughput field tobacco leaf counting method

Hengshu Xiao [1]; Junying Li [2] *; Hong Liang [1]; Liuchen Zhang [2]; Hong Zhang [2]; Erdeng Ma [2]; Xiaopeng Deng [2]

[1] School of Information, Yunnan University, Kunming 650504, China.

[2] Yunnan Academy of Tobacco Agriculture Science, Kunming 650021, China.

*corresponding author

Corresponding author:

Correspondence to Junying Li. Email: ljy1250@163.com

Abstract

Background: Estimating tobacco leaf yield is a crucial task. The number of leaves is directly related to yield. Therefore, it is important to achieve intelligent and rapid high-throughput statistical counting of field tobacco leaves. Unfortunately, the current method of counting the number of tobacco leaves is expensive, imprecise, and inefficient. It heavily relies on manual labor and also faces challenges of mutual shading among the field tobacco plants during their growth and maturity stage, as well as complex environmental background information. This study proposes an efficient method for counting the number of tobacco leaves in a large field based on unmanned aerial vehicle (UAV) image data. First, a UAV is used to obtain high-throughput vertical orthoimages of field tobacco plants to count the leaves of the tobacco plants. The tobacco plant recognition model is then used for plant detection and segmentation to create a dataset of images of individual tobacco plants. Finally, the improved algorithm YOLOv8 with Squeeze-and-Excitation (SE) and bidirectional feature pyramid network (BiFPN) and GhostNet (YOSBG) algorithm is used to detect and count tobacco leaves on individual tobacco plants.

Results: Experimental results show YOSBG achieved an average precision (AP) value of 93.6% for the individual tobacco plant dataset with a model parameter (Param) size of only 2.5 million (M). Compared to the YOLOv8n algorithm, the F1 (F1-score) of the improved algorithm increased by 1.7% and the AP value increased by 2%, while the model Param size was reduced by 16.7%. In practical application discovery, the occurrence of false detections and missed detections is almost minimal. In addition, the effectiveness and superiority of this method compared to other popular object detection algorithms have been confirmed.

Conclusions: This article presents a novel method for high-throughput counting of tobacco leaves based on UAV image data for the first time, which has a significant reference value. It solves the problem of missing data in individual tobacco datasets, significantly reduces labor costs, and has a great impact on the advancement of modern smart tobacco agriculture.

Keywords: Plant phenotype; Plant leaf detection; tobacco leaf count; UAV; YOLOv8; BiFPN; High-throughput

Introduction

Tobacco plants are considered important due to their low production costs and high economic value. Research on tobacco phenotype, especially leaf yield, quality, resistance, and genetic breeding, is attracting great attention among scientists. The measurement and analysis of Param of plant phenotype [1] can effectively regulate the breeding process and ultimately influence the growth, development, and crop yield of plants. Characteristics of tobacco plant phenotype mainly include plant height, number of leaves, and leaf area. Accurately counting leaf numbers is crucial for estimating crop yield and provides scientific researchers with important data support for effective tobacco plant breeding work. In addition, the production of tobacco leaves requires strict management, with the annual quantity and yield strictly controlled, making the statistical counting of the leaves of the tobacco plant a crucial task.

There are currently two main categories of statistical methods for counting leaves: automated and manual. Manual methods are traditional on-site observations and recordings, which are suitable for small-scale investigations or on-site research but have disadvantages such as poor timeliness, heavy workload, and multiple statistical error factors. As technology continues to advance, automated statistical techniques have gradually emerged. For example, near-infrared spectroscopy technology [2], high-performance liquid chromatography [3], remote sensing technology [4-5], image processing [6] and computer vision [7]. In recent years, advances in unmanned aerial vehicle (UAV) technology have provided a means to quickly and efficiently acquire high-throughput plant image data, enabling intelligent monitoring of massive plant growth processes [8]. Researchers use UAV-carrying sensors to collect data on plant height [9–10], leaf number [11–12], leaf area [13–14], etc., enabling real-time monitoring of plant growth and development as well as predicting yield. [15] showed an accuracy rate of 94.4% when using a camera-carrying UAV to count the number of pineapple crowns. Precision agriculture has benefited greatly from the widespread application of convolutional neural networks in object detection and semantic segmentation. One-stage and Two-stage object detection algorithms are the two main types of object detection algorithms available today. The One-stage algorithm simultaneously performs object localization and classification in an individual inference stage, with good real-time performance and faster detection speed. The representative algorithms include CenterNet [16-17], You Only Look Once (YOLO) [18], Single Shot MultiBox Detector (SSD) [19], etc. Two-stage algorithms generally have high accuracy but have slower detection speed. Representative algorithms include Region Convolutional Neural Networks (RCNN) [20], Faster-RCNN [21], etc. Wang M et al. [22] proposed an improved, Faster R-CNN-based dense sweet potato leaf detection scheme that integrates visual attention mechanisms and DIoU-NMS and achieved a detection accuracy of 95.7%. The method performed well in detecting dense or occluded leaves, but the generalization performance of the model was poor. Lu H et al. [23] introduced an advanced plant counting model, TasselNetV2+, which was more efficient in processing high-resolution images than traditional faster R-CNN models. It showed superiority in wheat ear count, corn tassel count, and validation of sorghum ear count, but had a slower detection speed. Xu X et al. [24] proposed a semi-supervised learning framework based on Noisy Student that uses YOLOv5x to identify and count maize leaves after first segmenting entire maize seedlings and creating foreground images with background eliminated using the SOLOv2 model. This method requires only a modest amount of labeled data, reducing the amount of work required for early data annotation. However, the counting performance of the YOLOv5x model is relatively low for newly formed leaves. Ubbens J et al. [25] proposed to train convolutional neural networks with synthetic 3D rendering images and real images together to improve leaf counting performance, thus solving the undertraining of neural networks due to the sparse data sets. However, the process of generating and rendering images is more complex, and synthetic images cannot replace certain specific plant characteristics. Xie X et al. [26] proposed a novel approach leveraging deep neural networks using Inception-Resnet-V2. It extracts leaf skeletons and applies enhancement techniques, eliminating the need for explicit labeling of leaf structures, and enables leaf counting in monocots. It works well when there are strong leaf occlusions in the images, although the process of generating high-quality leaf skeletons and applying enhancement techniques is complex and the model's robustness to noise effects is average.

In summary, it is not difficult to find that there is extensive research on the use of deep learning and computer vision technology to achieve plant leaf counting, which has a certain reference value. However, in practical applications, there are still problems such as slow detection speed, insufficient detection accuracy, poor model generalization performance, and complex detection methods. In addition, there is also the problem of scarce plant data sets to support model training. As for the task of leaf counting of tobacco plants, due to the size and size of tobacco plants, it is difficult to collect data on individual tobacco plant images, and there are serious problems of mutual occlusion of mature tobacco leaves and complex environmental background information. Currently, there is no efficient method for counting tobacco leaves. Therefore, we first obtained high-throughput image data of mature tobacco plants by capturing images in situ using UAV. After image processing and data annotation, we used the constructed tobacco plant recognition model for detection and segmentation and obtained a dataset of 8,000 images of individual tobacco plants. Secondly, compared with the current mainstream object detection algorithms, we found that the YOLO series of algorithms has highly efficient real-time performance and excellent detection capabilities. The latest version, YOLOv8, continues this advantage with even faster and higher detection performance and introduces additional features such as multi-scale detection and optimization of small object detection. Therefore, we optimized the YOLOv8-based model to improve the recognition performance of the model, and developed the YOSBG, a lightweight tobacco leaf detection algorithm based on YOLOv8, which achieved an average AP of 93.6% with model Params reduced to 2.5 M, enables deployment of the model on mobile devices with limited resources. It outperforms most classical detection algorithms in detection experiments and rarely shows false positives or false negatives in practical application detection, making it extremely valuable for high-throughput plant leaf counting. The main contributions of this work are as follows:

- A high-throughput image dataset of field tobacco plants was obtained through UAV photography, and a tobacco plant detection model was used for detection and segmentation to create an image dataset of 8,000 individual tobacco plants.
- The improved algorithm embedded the SE [27] module in the bottleneck of the C2f module of the YOLOv8 network to reduce shallow noise information and strengthen the network's ability to detect small and densely obscured targets.
- By replacing the Path Aggregation Network (PANet) [28] with the more efficient and simpler BiFPN [29-30] module, both the recognition speed of the model and the feature fusion ability are improved.
- The GhostNet [31-32] and C2f modules are merged to form the C2fGhost module [33]. Integrating with the neck network segment of YOLOv8 not only reduces the number of model Params by 16.7% but also improves the overall detection accuracy of the model.
- In the detection experiment of the individual tobacco plant dataset, YOSBG achieved an AP of 93.6%, and the Params of the model were reduced to 2.5M. The comparison experiment confirmed that its detection performance outperforms most of the current classical detection algorithms, including Faster R-CNN, CenterNet, YOLOv3, YOLOv5, etc.

Materials and Methods

Experimental location

The experimental site is located in the World Tobacco Variety Garden in Yuxi City, Yunnan Province, People's Republic of China, at 102.87902° E, 24.652563° N. The climate in this region is pleasant with abundant rainfall, which is very conducive to the growth of tobacco plants favored. In addition, the area has wide and flat terrain, which makes unmanned aerial vehicles suitable for conducting large-scale data collection and analysis in the area. Figure 1 shows the geographical location of the experimental area.

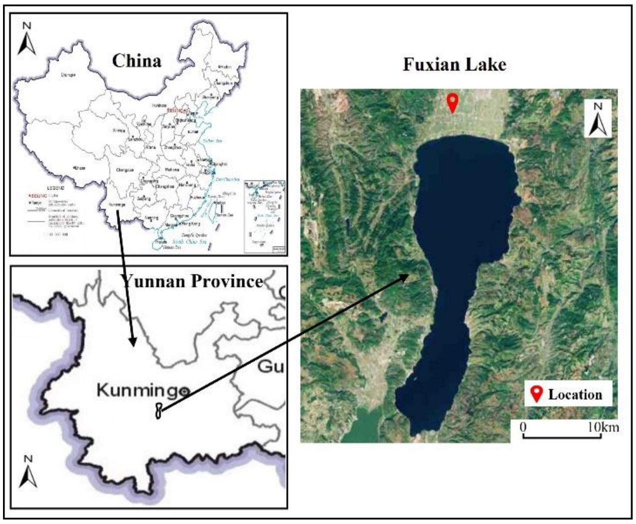


Figure 1. Experimental location

Construct datasets

The data required for the experiment is the vertical orthographic image of the tobacco plants. To ensure the rapid, high-throughput acquisition of high-quality image data, a DJI PHANTOM 4 PRO UAV equipped with a camera with an effective pixel count of 20 million, a focal length of 9 mm, and a maximum resolution of 5472 * 3956 was used as the experimental equipment. The flights were scheduled for clear weather conditions with low wind speeds and appropriate temperature and humidity to avoid strong sunlight and shadows that could affect the quality of the data, thus ensuring the normal operation of the equipment and the accuracy of the data. Before each flight operation, it was necessary to check the status of the device batteries and the strength of the GPS signal and adjust Params such as the UAV flight altitude, image overlap, and sampling distance to obtain high-quality RGB image data. The flights were scheduled for clear weather conditions with low wind speeds and appropriate temperature and humidity to avoid strong sunlight and shadows that could affect the quality of the data, thus ensuring the normal operation of the equipment and the accuracy of the data. Before each flight operation, it was necessary to check the status of the device batteries and the strength of the GPS signal and adjust Params such as the UAV flight altitude, image overlap, and sampling distance to obtain high-quality RGB image data.

During the flight, the image overlap of the main route was set to 80%, the image overlap between the lines was set to 85%, and the UAV with the RGB camera system was set to take photos at altitudes of 5 m, 6 m and 7 m captured ground above the aircraft at a flight speed of 0.4 m/s, the RGB camera was set to take photos at 2-second intervals, and the RGB camera was set to automatic exposure mode. The device adjusted shutter speed, aperture size, and ISO sensitivity in real-time according to changes in ambient light to ensure that the contrast and brightness of the image were in the correct range. Figure 2 shows the equipment and archived tobacco plant photos taken at the site.

Figure 2. Equipment and filed tobacco plant photos

After capturing the orthoimage of tobacco plants by UAV, it is extremely difficult to directly count the number of tobacco leaves in the image data obtained by UAV photography. The individual recognition tasks are not only extensive, the background information is also complex and the recognition effect is unsatisfactory. In order to achieve an accurate determination of the number of tobacco leaves, it is therefore necessary to create an image data set of individual tobacco plants. To this end, this study first annotates the tobacco plants in the orthoimage obtained by the UAV and then trains them in the target detection network to build a tobacco plant detection model. The segmented images of individual tobacco plants obtained through the recognition are stored, resulting in approximately 8000 images in the dataset. Figure 3 shows an image of an individual tobacco plant after segmentation.

Figure 3. Individual tobacco plant

Data annotation and augmentation

In this study, the dataset used for neural network training follows the format of Pascal VOC [34] and the labeling of tobacco leaves is done using the LabelImg tool. LabelImg is an open-source image annotation tool that provides users with a visual interface to edit and modify existing annotations and categories. The label was set to "Leaf" and after selecting the position of the maximum bounding box, the file was saved in XML format containing the pixel coordinates and label of the target tobacco leaves in each plant image. Subsequently, the data set was divided into training, validation, and test sets, with the training set accounting for 80% and the validation and testing set accounting for 10% each. Data augmentation is an effective strategy in machine learning and deep learning tasks because it can improve the performance, generalization ability, and robustness of the model while reducing the impact of sampling imbalance. This is especially important when data is limited. As shown in Figure 4, the data augmentation methods used in this study included concatenating, mirroring, scaling, and adjusting brightness and contrast to enrich the information about the detected objects.

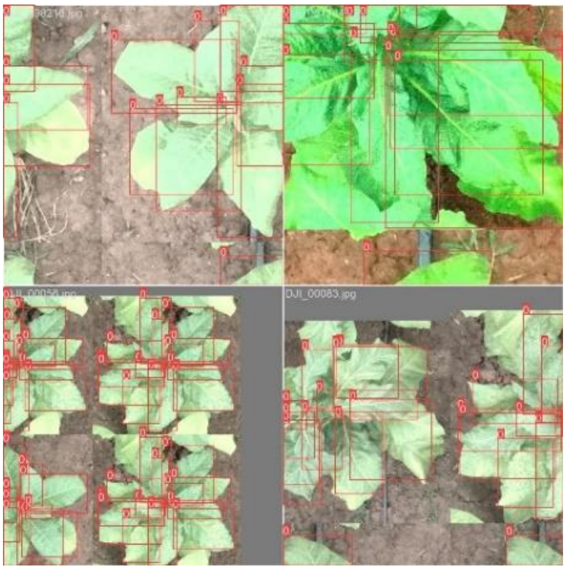


Figure 4. Data augmentation results

Tobacco leaf number counting method

The flowchart shown in Figure 5 below illustrates the process of intelligently counting tobacco plants in the field. First, UAVs are used to capture high-throughput vertical orthoimages of tobacco plants in the field. Next, a tobacco plant detection model is used to create a dataset of images of individual tobacco plants. The third step involves performing data augmentation and using a labeling tool called LabelImg to annotate the tobacco leaves. In the fourth step, the data is input into the constructed neural network model for training, whereby the Params are continuously adjusted to obtain the optimal training model for tobacco leaf recognition. Finally, the improved YOSBG algorithm is used to perform tobacco leaf detection for individual tobacco plants and complete the leaf counting task.

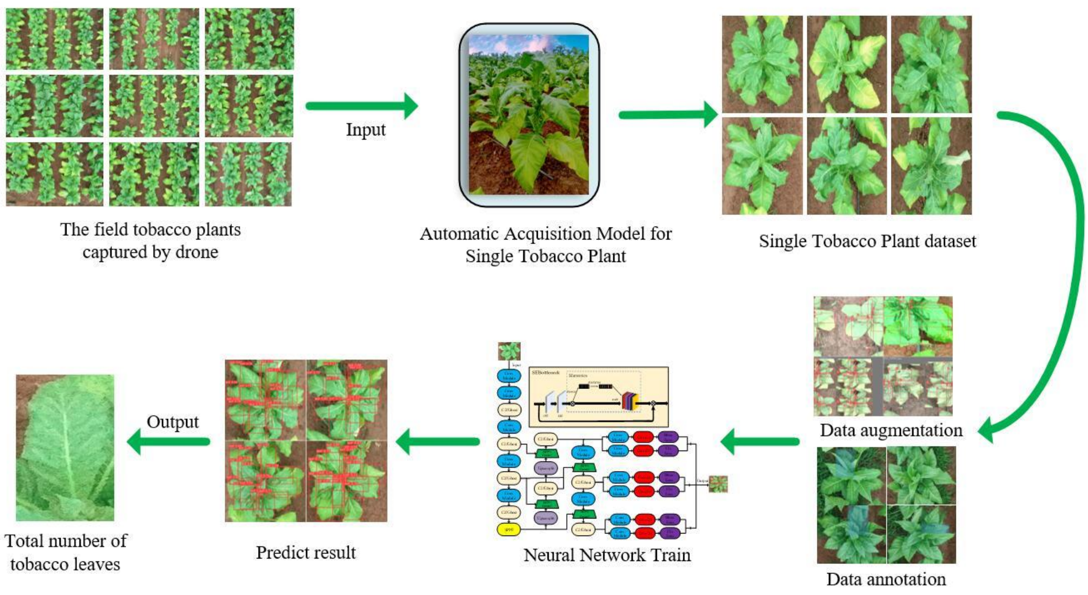


Figure 5. Flow chart of tobacco leaf number counting method

Related Algorithm

YOLOv8 algorithm introduction

From the YOLO series of algorithms, YOLOv8 is the latest version. Its core idea is to transform object detection assignment into a regression problem. The algorithm adopts the idea of Anchor-Free and Feature Pyramid Network (FPN), which achieves efficient object detection by fusing multi-scale features of the image. The feature extraction network of YOLOv8 uses a ResNet-like residual structure and extracts features of the input image through multi-scale convolution and skip connections, thereby achieving better feature representation ability.

FPN repeatedly upsamples the low-level feature maps and combines them with the high-level feature maps to obtain a set of feature maps that contain information about the different scales of the image, thus increasing the accuracy of detection of the object. YOLOv8 has moved from Anchor-based to Anchor-free, making image classification and regression more efficient and accurate without the need to define anchor points in advance for testing. For the calculation of the loss function, YOLOv8 uses the binary cross-entropy loss as the classification loss and the CIOU+DFL loss as the regression loss and introduces the positive sampling strategy TaskAlignedAssigner. YOLOv8 has different versions based on network depth and width, such as YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. Considering the need for rapid large-scale tobacco leaf detection and the need for subsequent deployment on mobile devices, this paper proposes an improvement method YOSBG based on YOLOv8n, with a smaller network depth and width for faster inference speed. YOLOv8 has various versions based on network depth and width, such as YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. Considering the need for large-scale rapid detection of tobacco leaves and the requirement for later deployment on mobile devices, this paper proposes an improvement method YOSBG based on YOLOv8n, with a smaller network depth and width for faster inference speed. The YOLOv8n network structure includes three primary components: the backbone network, the neck network, and the detection head, as shown in Figure 6. The backbone and neck may have adopted the design concept of YOLOv7 ELAN and replaced the C3 structure of YOLOv5 with another enriched gradient flow C2f structure. The modules of the backbone network are C2f, SPPF, and CBS modules. The SPPF module is a spatial feature pyramid pooling layer that improves the expressiveness of feature maps and accelerates operations by merging feature maps with various receptive fields. The head region has been changed from coupled to decoupled, separating the classification head and the detection head.

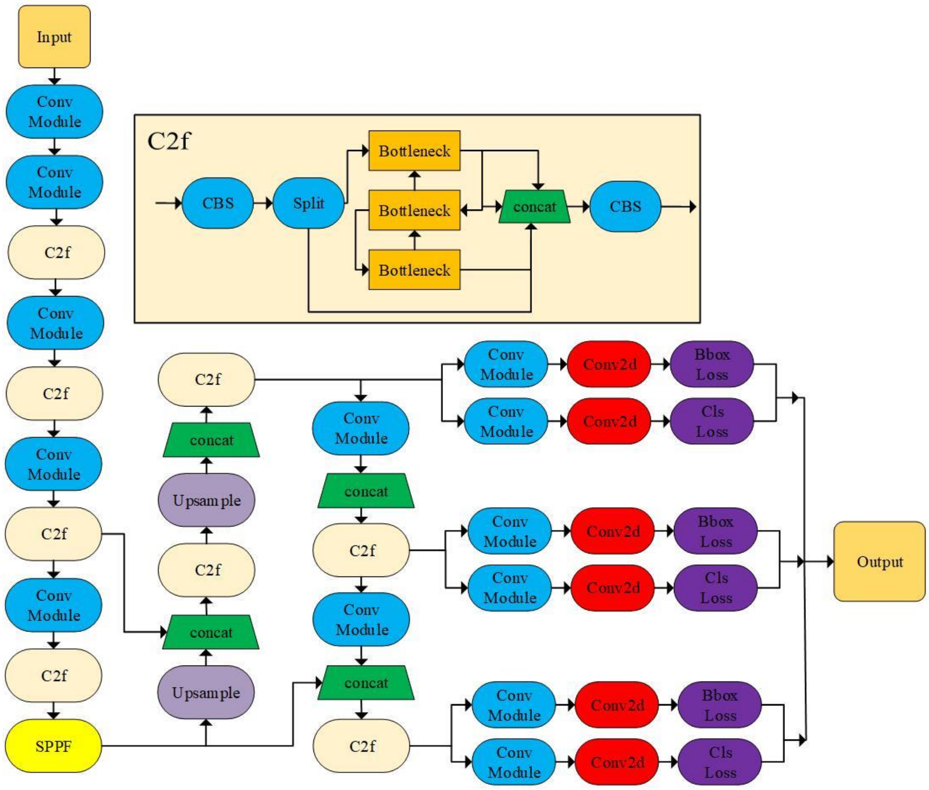


Figure 6. YOLOv8 network structure [35]

Problems of algorithm

After implementing numerous common convolutions and C2f modules in YOLOv8, the precision of the algorithm has increased at the expense of increasing the model size and reducing the speed at which it runs, requiring higher computing resources to achieve higher detection accuracy and better performance of the algorithm. However, when using mobile devices to count tobacco leaves, the limited performance of the device must be taken into account. Therefore, the minimum weight model YOLOv8n was selected in this study. In addition, in practical detection tasks, the YOLOv8n model has lower detection accuracy due to the small size and overlap of tobacco leaves, which leads to problems such as false detections and missed detections.

Improvement of algorithm

In order to improve the accuracy of tobacco leaf detection and solve the problems of low detection accuracy for small targets and overlap, the SE attention mechanism is introduced into the feature fusion network of the YOLOv8 algorithm to enhance the network feature extraction ability for critical information and improve detection capability for small targets and densely obscured targets as well as improved detection accuracy. Using a more efficient and simpler BiFPN module makes the feature fusion method simpler and more efficient by improving detection speed and reducing model complexity. The C2fGhost module, which merges the GhostNet and C2f modules, will be integrated into the Neck network section of YOLOv8n to replace the convolution utilized in the original network. This reduces the Params and computational load of the model, making it more suitable for model detection and final deployment. Figure 7 below shows the improved structure of YOLOv8.

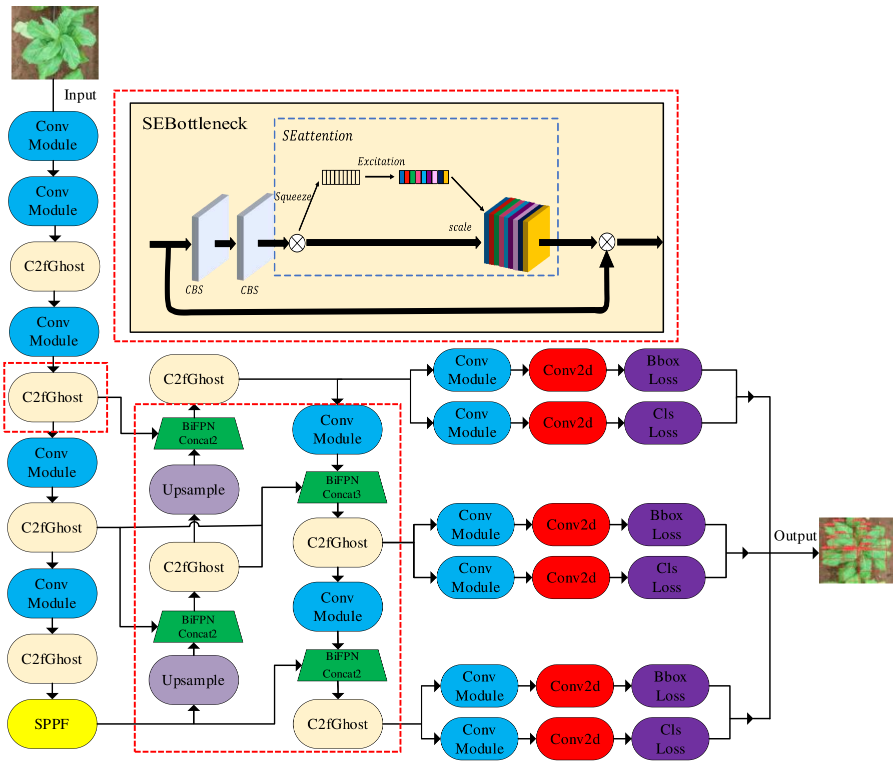


Figure 7. The improved network structure of YOLOv8

Add SE attention module

The feature extraction network of YOLOv8 faces difficulty in detecting small targets and extracting features of occluded targets. Hence, this work introduces the SE attention module to increase the performance of the network without adding additional Params. The primary aim of the SE attention module is to enhance network performance by explicitly modeling the interdependence between convolutional feature channels. The SE attention module assigns different weights to feature maps of channels with different importance, ensuring that the model can focus on information from small and obstructed targets. Figure 8 below shows the structure of the SE attention module.

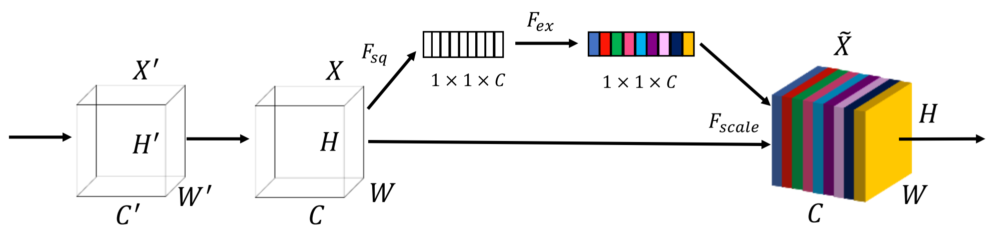


Figure 8. SE attention module structure

X' represents the original input data, H' represents spatial height, W' represents the spatial width, and C' represents the number of channels. X represents the feature map after convolution, H represents the spatial height after convolution, W represents the spatial width after convolution, and C represents the number of channels of the convolutional features. The Squeeze operation compresses the feature maps, the Excitation operation excites the feature maps, $\tilde{X}$ represents the feature recalibration result.

The Squeeze step primarily compresses the input feature map of $H \times W \times C$ into a $1 \times 1 \times C$ vector through global average pooling, thereby reducing the spatial dimensions. Each element of this vector represents a channel of the features of the input feature map and compresses the meaning of each channel of the input feature map while preserving global information. As shown in Eq. 1.

$$z_c = \mathbf{F}_{sq}(\mathbf{u}_c) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}u_c(i,j) \tag{1}$$

Here, z_c represents the global feature, $\mathbf{F}_{sq}$ depicts the squeezing operation, $\mathbf{u}_c$ represents the input feature map, H is the height of the feature map, W is the width of the feature map, $u_c(i,j)$ represents the feature vector of the i-th row and j-th column pixels.

A weighted feature map is created by multiplying the weight values by the original input feature map after normalizing the weight values in the excitation step using a sigmoid function. This approach allows the model to improve performance by highlighting valuable functional channels while suppressing those that are less useful for current tasks. The output of this weighted feature map is the result of the SE attention mechanism, as shown in Eq. 2.

$$s = \mathrm{F}_{ex}(\mathrm{z},\mathrm{W}) = \sigma\big(\mathrm{W}_2\delta(\mathrm{W}_1\mathrm{z})\big) \tag{2}$$

Here, s represents the excitation score vector, F_{ex} represents the excitation operation, $\mathrm{W}_1 \in R^{\frac{C}{r}\times C}$ is the weight matrix of size $\frac{C}{r}$ rows C columns, $\mathrm{W}_2 \in R^{C\times\frac{C}{r}}$ is the weight matrix of size C rows $\frac{C}{r}$ columns, r is the scaling factor, σ is the sigmoid function, and δ is the ReLU activation function.

This study compares the effect of adding the SE attention module at the last layer of the backbone network and adding it to the bottleneck of the C2f module. It was found that placing the SE module at the bottleneck in the C2f module resulted in better performance. The Bottleneck originally consisted of two CBS modules, and the SE module was embedded after the second CBS module. The structure of the bottleneck after embedding the SE module is shown in Figure 9.

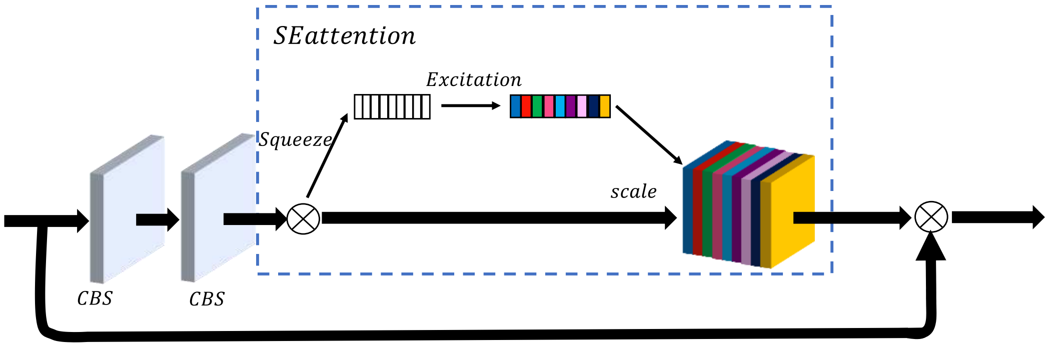


Figure 9. Bottleneck with SE attention mechanism embedded

Add BiFPN multi-scale feature fusion module

A top-down method for fusing multi-scale features to fuse shallow position information and deep semantic information is proposed by the FPN. Through bidirectional cross-scale connections, lower-level target position information is merged with upper-level semantic feature information, changing the original information flow to reduce information loss during calculation, as shown in Figure 10(a).

The PANet used in the YOLOv8 algorithm is based on FPN, as shown in Figure 10(b). In addition to FPN, it is complemented by another path aggregation network, which enables the transmission of bottom-level image information to the prediction feature layer, so that the prediction layer contains top-level semantics and bottom-level feature information at the same time. However, PANet increases the complexity of the network model and reduces the efficiency of information transmission. When detecting tobacco leaves, using the original YOLOv8n model may cause mutual interference between shallow and deep information due to the complex background information, which affects the detection accuracy. In order to achieve efficient and fast multi-scale feature fusion and improve detection accuracy, the neck network segment was upgraded by integrating the BiFPN module.

The BiFPN structure builds on top of the PANet by using cross-scale connections to remove nodes in the PANet that contribute minimally to feature fusion. In addition, to merge more functional information while conserving resources, new channels are added between the initial input and output nodes. The structure is shown in Figure 10 (c). The BiFPN structure performs fast normalization fusion through the ratio of weight to the sum of weights, enhancing the perception of targets in various situations. It can merge information from feature maps at different levels during prediction, effectively combating interference caused by noisy images and other factors. By extending YOLOv8's neck network with BiFPN, multiscale feature fusion is faster and easier. In addition, by adding learnable weights, BiFPN can regularly apply both top-down and bottom-up multiscale feature fusion and learn the relative importance of different input features. Compared to the PANet in the neck network of YOLOv8, BiFPN consumes fewer Params and computing resources. Therefore, tobacco leaves can be better detected and tobacco leaf counting tasks can be performed in real time.

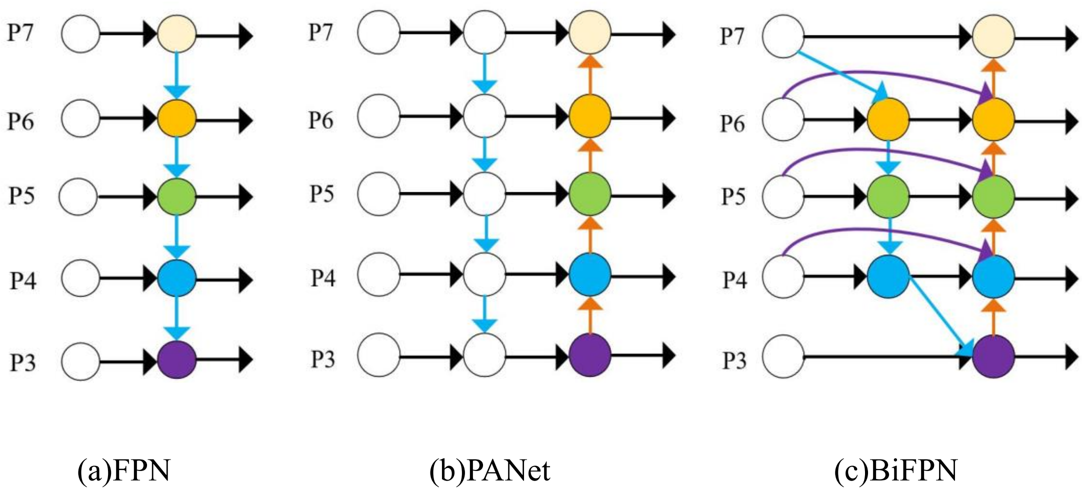


Figure 10. Different feature pyramid structures

C2fGhost module

The core of GhostNet utilizes the Ghost module, which is different from the standard convolutional feature extraction technique. In feature extraction, the Ghost module first uses standard convolution to extract part of the feature maps. Simple linear operations are then used to capture additional feature maps based on these extracted maps. Ultimately, the feature mappings derived from the first two phases are merged to obtain the final output. Figure 11 shows the convolution process between ghost modules.

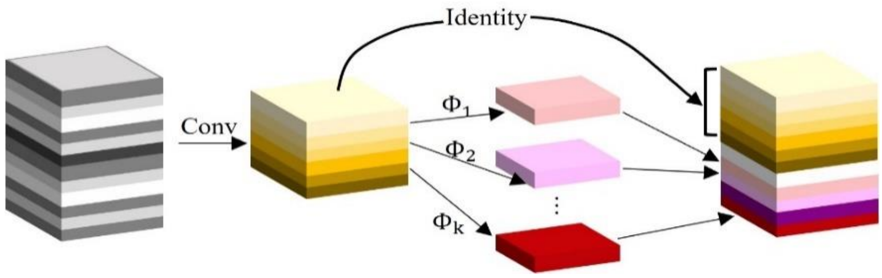


Figure 11. Ghost convolution module

The Ghost module exploits the convolution process by taking advantage of its lightweight advantage. The Ghost module has a much lower computational cost than the traditional convolution method when the two feature maps are of the same scale. Assuming that there are redundant features for each basic feature s, the computational cost of the ghost module is about 1/s of the standard convolution. This can reduce the computational cost without changing the size of the output feature map and maintain model performance so that more feature maps can be obtained. As shown in Figure 12, two consecutive ghost modules form the ghost bottleneck. To increase the number of channels, the primary Ghost module is used as an extension layer and to adjust the link path, the second Ghost module reduces the number of channels. The inputs and outputs of these two Ghost modules are connected by the link.

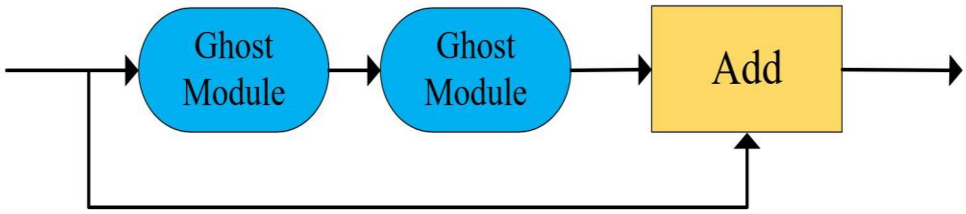


Figure 12. Ghost Bottleneck

In this article, the bottleneck of the C2f module is replaced with a ghost bottleneck. Figure 13 shows the final C2fGhost module. The C2fGhost module's introduction not only reduces the model size without increasing network Params, reduces the inference time of redundant feature maps, realizes network lightweight, and improves the accuracy of network detection.

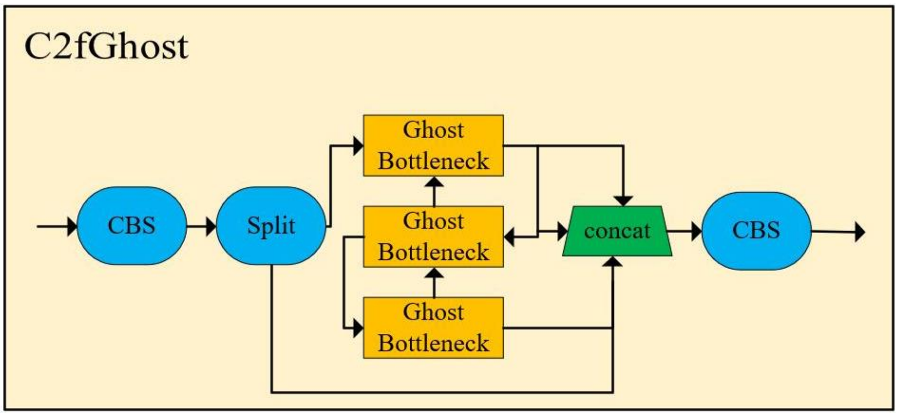


Figure 13. C2fGhost module

Results

Experimental environment

The experiment used the DJI PHANTOM 4 PRO UAV for data collection. The model was developed primarily in Python, using the open-source deep learning framework Pytorch version 2.0.1, Python version 3.9.18, and CUDA version 11.8. The hardware used to train the object detection model included a 13th Generation Intel(R) Core(TM) i5-13600KF CPU and an NVIDIA GeForce RTX 3060Ti GPU with 8GB of memory. The training Param settings for the experiment are listed in Table 1 below.

Table 1 training hyperparameters

Param name	Value
Batch-Size	4
input-images	640×640×3
momentum	0.9
Weight decay	0.0005
initial learning rate	0.002
Epoch	500
optimizer	AdamW

Model evaluation index

The evaluation metrics for this experiment included Recall, Precision, AP, F1, and Params. The size of the model is determined by Params. In general, smaller Params in model inference result in lower performance requirements on hardware devices and easier model deployment on low-end devices. The precision rate refers to the probability of a correct prediction when the result of the model prediction is a positive sample. The recall rate is the proportion of correct results detected by the algorithm among all targets. The F1 balances precision and recall and better reflects the performance of the model. The AP is determined by Recall and Precision, which can intuitively reflect the recognition performance results of each category. Perform an integral operation on the precision-recall curve to find the area enclosed by the curve that represents the AP value. The calculation formula is shown in the Eqs. 3, 4, 5, 6.

$$Precision = \frac{TP}{TP+FP} \times 100\% \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \tag{4}$$

$$F1 = 2\,\frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

$$AP = \int_{0}^{1} p(r)\mathrm{d}r \tag{6}$$

TP: Represents the amount of true positive samples whose detection results were correctly determined to be positive.

FN: Indicates the number of truly positive samples whose detection results were incorrectly classified as negative.

FP: Indicates the number of samples whose detection results were incorrectly classified as positive among the negative samples.

Result analysis

(1) Comparison experiment of different attention mechanisms

We included comparisons with other attention mechanism modules to confirm the effectiveness of including the SE module in this experiment, including the Convolutional Block Attention Module (CBAM) [36], the Expectation-Maximization Attention (EMA) [37], and the Global Attention Mechanism (GAM) [38]. At the same time, this study also investigated the addition of various attention mechanisms to the last layer of the backbone network and the bottleneck of the C2f module. The experimental results are shown in Table 2 below. Where position represents the location where the attention mechanism is added: a represents the addition to the last layer of the backbone, and b represents the addition to the bottleneck of the C2f module. Table 2 shows that at position b added SE attention mechanism, the detection accuracy of the model is significantly increased with only a slight increase in Params, and the effect is better compared to other research attention modules.

Table 2 Comparison experiment of different attention mechanisms

Methods	position	Precision/%	Recall/%	AP/%	Params/M
YOLOv8	——	90.2	84.7	91.8	3.01
CBAM-	a	87.6	86	92.1	3.01
YOLOv8	b	87.8	88.8	92.5	3.05
EMA-	a	87.9	88.4	92.2	3.01
YOLOv8	b	89.9	86.2	92.5	3.05
GAM-	a	89.6	87.2	92.6	4.64
YOLOv8	b	89.1	87.2	92.1	4.36
SE-	a	89.2	86.2	92.5	3.01
YOLOv8(Ours)	b	**90.2**	86.6	**92.8**	3.05

(2) The impact of the C2fGhost module on model performance

In this study, experiments were created to investigate the impact of introducing the C2fGhost module in different network locations on network performance. As shown in Table 3, in the Ghost-all method, the C2f modules in both the backbone network and the neck network were completely replaced with C2fGhost modules. When using the Ghost-backbone method, the C2fGhost module only replaces the C2f module in the backbone network, and the neck network remains unchanged. Ghost-neck means that the C2f module in the neck network is replaced by the C2fGhost module, and the C2f module in the backbone network remains unchanged. As can be seen from Table 3, the introduction of Ghost modules at different locations of the YOLOv8n network reduces the number of Params of the model. The AP value of the Ghost-all algorithm increased by 0.76% and the Params decreased by 28.9%. The Ghost-backbone AP value decreased by 0.33% and the Params value decreased by 14.3%. Ghost-neck AP increased by 0.87% and params decreased by 15.0%. When comparing the experimental results, the Ghost module was introduced to shrink the model without increasing the network Params. Therefore, it is preferable to replace the C2f module in the entire YOLOv8 network.

Table 3 The impact of C2fGhost modules in different positions on network performance

Methods	Precision/%	Recall/%	AP/%	Params/M
YOLOv8n	90.2	84.7	91.8	3.01
Ghost-backbone	88.4	86.5	91.4	2.58
Ghost-neck	89.1	85.1	**92.6**	2.56
Ghost-all（ours）	89.1	86.8	92.5	**2.14**

(3) Ablation experiment

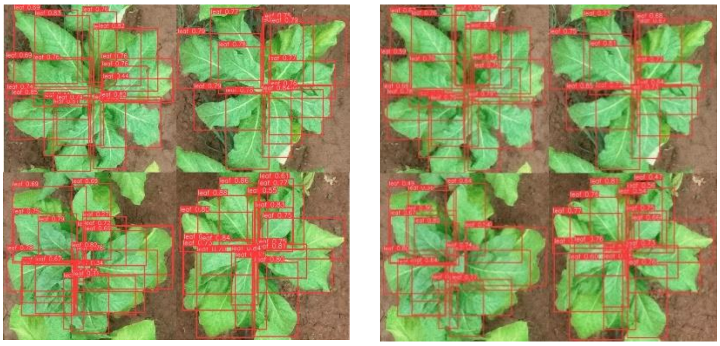
To confirm that the various modules proposed in the YOLOv8n network model are effective, ablation experiments were designed. The experimental results are shown in Table 4. Here, YOLOv8-SE represents the algorithm that introduces the SE module into the bottleneck of the C2f module, YOLOv8-B represents the algorithm that adds the BiFPN module in the neck network, and YOLOv8-G represents the algorithm that Adds the C2fGhost module in the backbone network and the neck network. YOLOv8-SEB, YOLOv8-SEG, YOLOv8-BG, and YOSBG represent the algorithms according to the combination of modules. As shown in Table 4, after the introduction of the SE module and the BiFPN module, the F1 score of the YOSBG model improved by 0.9% and the AP score of the model increased by 1.5% compared to the base model YOLOv8n. In addition, with the introduction of the C2fGhost module, the F1 score of the final model YOSBG used in this article increased to 1.7%, the AP value of the model increased to 2%, and the Params of the model also decreased by 16 .7%. This not only reduces the difficulty of deploying the model to mobile devices but also reduces hardware costs and significantly improves accuracy while meeting real-time requirements.

Table 4 Ablation experiment result

Methods	SE	BiFPN	C2fGhost	F1	AP/%	AP_{50-95}/%	Params/M
YOLOv8n				87.4	91.8	0.581	3.0
YOLOv8-SE	√			88.4	92.8	0.581	3.1
YOLOv8-B		√		87.5	92.2	0.58	3.0
YOLOv8-G			√	87.9	92.5	0.569	2.1
YOLOv8-SEB	√	√		88.2	93.2	0.579	3.1
YOLOv8-SEG	√		√	88.2	93	0.581	2.5
YOLOv8-BG		√	√	88.3	92.8	0.577	2.1
YOSBG(Ours)	√	√	√	**88.9**	**93.6**	**0.583**	2.5

(4) Practical application detection results

We conducted an experiment using the YOLOv8n model and the improved YOSBG model to detect the same individual tobacco leaves and confirm the performance of the improved model in real-world applications. Figure 14(a) shows the practical application effect of the YOLOv8n model with the detection of 17, 13, 19, and 17 leaves in the four images. Figure 14(b) shows the practical application effect of the YOSBG model by detecting 17, 11, 18, and 16 leaves in the four images, where the actual number of leaves in the four images is 17, 11, 18, and 15. The YOLOv8n algorithm has false detections in three images and detects more leaves than the actual values due to the dense and overlapping nature of the leaves, which leads to interference in the detection. The detection results of the improved algorithm are consistent with the actual values, which can better complete the task of completing tobacco leaf detection and statistics.



(a)YOLOv8n (b)YOSBG

Figure 14. Practical application detection results

(5) Comparison of results of different detection algorithms

In this study, the individual tobacco dataset was used to compare YOSBG with other popular object detection algorithms and further validate the performance of the algorithm. The experimental results are listed in Table 5. Compared to Faster-RCNN (ResNet50) and CenterNet (ResNet50), YOSBG improved in AP metric by 8.7% and 12.0%, respectively, with a model Param size of only 1.8% and 7.6%. Compared to other YOLO series algorithms such as YOLOv3, YOLOv5s, and YOLOv6, YOSBG improved in AP metric by 2.5%, 2.7%, and 2.2%, with a model Param size of only 2.4%, 34 .7% and 53.2% respectively. From these results, it can be seen that the improved algorithm not only achieved significant improvements in model accuracy but also significantly reduced the model size. The AP curve of the YOLO series of algorithms is shown in Figure 15 below.

Table 5 Results of different detection algorithms

Methods	F1	Precision/%	Recall/%	AP/%	Params/M
Faster R-CNN	83.0	83.6	82.5	85.5	137.7
CenterNet	87.0	89.2	84.7	82.4	32.7
YOLOv3	87.1	87.1	87.0	91.3	103.6
YOLOv5s	86.6	88.5	84.8	91.1	7.2
YOLOv6	87.5	88.0	86.9	91.6	4.7
YOLOv8n	87.4	90.2	84.7	91.8	3.0
YOSBG(Ours)	**88.9**	**90.7**	**87.1**	**93.6**	**2.5**

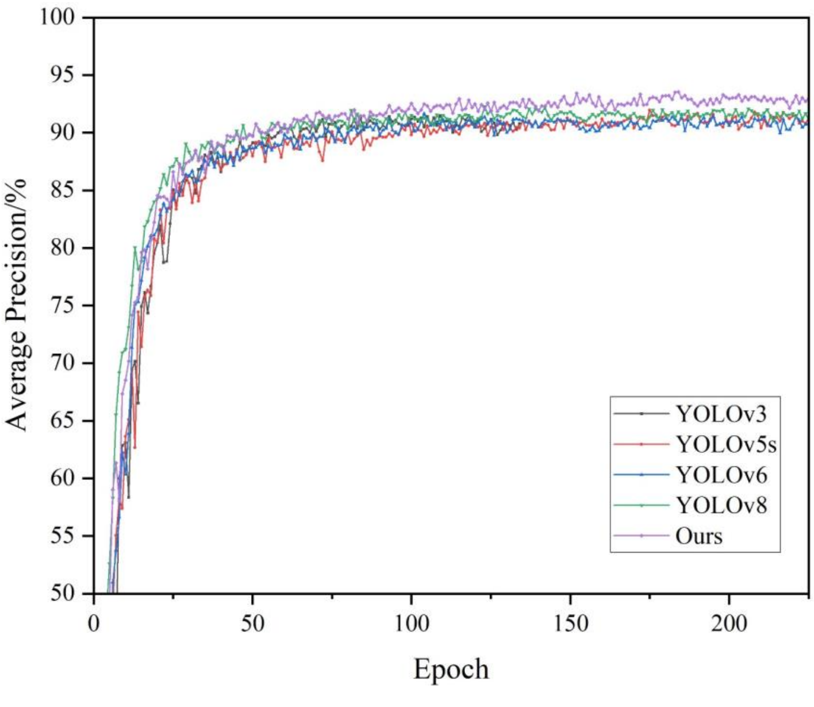


Figure 15. Average precision curve of the YOLO series of algorithms

Discussion

In modern agricultural applications, artificial intelligence methods are widely used to monitor and manage plant growth processes, predict and control crop yields, and select superior crop varieties. Typically, these methods rely on RGB images and computer vision object recognition technologies to collect plant information. Therefore, creating an image dataset is the first step of such research. Collecting data is a very labor-intensive task. In some advanced agricultural regions, professional greenhouse laboratories are installing cameras to collect long-term data. In most cases, however, researchers have to take photographs on-site. The use of UAVs for aerial photography in smart agriculture greatly improves the efficiency of crop data collection.

The current application of deep learning for object recognition is widely used in crop and plant leaf recognition. Crop images captured by high-resolution UAVs often contain high-throughput information that can be used to predict crop yield for the entire production area. Consequently, it is more effective to use UAVs to predict crop yield. However, due to a large amount of information, direct plant detection using cameras mounted on UAVs can only monitor whole plants or very different plants such as corn, eggplant, lettuce, etc. When detecting small and densely distributed targets such as plant leaves, there are significant errors in using data collected directly from UAVs, making it difficult to obtain statistical leaf counts for high-throughput crops. This article is the first to propose a method for counting tobacco leaves in the field, which only requires image data of field tobacco plants captured by UAV to complete the statistical analysis of the total number of tobacco leaves in the field. This has great reference value, not only to solve the problem of missing data sets for individual plants but also to significantly reduce the cost of data collection. As shown in Table 5. The improved YOLOv8 algorithm used in this article provides better tobacco leaf detection performance, better real-time performance, and 93.6% accuracy. It also works well in real applications and our improvements have reduced the model's Param volume by 16.7%, making it easier to deploy the model on mobile devices.

Although the algorithm has achieved good results in counting tobacco leaves, it has some limitations. First, the data captured by the UAV must be vertical orthographic images captured at the top of the plants. When plant recognition models are used to segment plant images at high throughput, skewed segmentation images may occur, which could affect the recognition effectiveness of the algorithm. In addition, some segmented individual plant images may contain some leaf information from neighboring plants, which may lead to duplicate detections in leaf counting. Secondly, when taking plant images, various other factors such as environment, light, soil, plant growth stage, plant shape, plant density, etc. may also have a certain influence on the final recognition. Therefore, further research is needed in the future to evaluate the reliability of this method for high-throughput plant leaf counting.

Conclusions

Fast and accurate high-throughput statistical counting of tobacco plant leaf area can enable prediction and control of tobacco yield. In this study, our proposed method is based on RGB image capture by UAV, followed by detection and segmentation using object recognition technology to obtain a dataset of individual images of tobacco plants, and then counting the number of tobacco leaves. The feasibility has been verified through experiments, and the proposed algorithm YOSBG has high detection accuracy and superior performance in practical applications, which has significant implications for tobacco yield estimation, overall resource optimization, and the advancement of modern smart tobacco agriculture. The proposed YOSBG tobacco leaf detection algorithm based on the improved YOLOv8 has several important innovations. First, to address the low detection accuracy for small targets and overlapping occlusions in the YOLOv8 algorithm, the SE module is introduced in the feature fusion network to improve the feature extraction ability for important information, thereby improving the detection accuracy. Secondly, to address the problem of complex background information and interference between shallow and deep information, the BiFPN module is introduced to replace the PANet structure, achieve simple and fast multi-scale feature fusion, and enhance the feature fusion capability of the model to improve and thus improve the detection speed. Finally, the C2fGhost module is introduced into the YOLOv8 network to reduce the Params of the model while maintaining the speed and accuracy of detection. Experimental results show that YOSBG has advantages such as fewer Params, higher recognition accuracy, and lower calculations, and meets real-time requirements. Compared to existing models, this method has higher detection accuracy while reducing the need for platform computation and storage capacity, making it easier to deploy on resource-constrained devices. However, issues with missed and false detections still exist and will be addressed in future improvements. Furthermore, the improved model is deployed on resource-constrained embedded detection devices to improve the proposed algorithm for practical applications. In addition, due to the limitations of the high-throughput tobacco leaf counting method proposed in this study, further research and evaluation will be carried out to apply this method to the high-throughput leaf counting of other plants.

List of abbreviations

UAV: unmanned aerial vehicle

SE: Squeeze-and-Excitation

BiFPN: Bidirectional feature pyramid network

YOSBG: YOLOv8 with SE and BiFPN and GhostNet

AP: average precision

Param: parameter

M: million

F1: F1-Score

YOLO: You Only Look Once

SSD: Single Shot MultiBox Detector

RCNN: Region Convolutional Neural Networks

PANet: Path Aggregation Network

FPN: Feature Pyramid Network

CBAM: Convolutional Block Attention Module

EMA: Expectation-Maximization Attention

GAM: Global Attention Mechanism

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable

Availability of data and materials

The datasets used or analyzed during the current study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

The research project was funded by the Major Science and Technology Project of Yunnan Province of China Tobacco Corporation (Fund number: 2021530000241025, 2022530000241030), and the Graduate Research Innovation Fund of Yunnan University (Serial number: KC-23235266).

Authors' contributions

Conceptualization and methodology, H.X., J.L., H.L.; formal analysis and investigation, H.X., J.L., H.L., L.Z., H.Z.; resources, H.X., H.L., J.L., H.Z.; writing—original draft preparation, H.X., J.L., E.M.; writing—review and editing, X.D., H.X., J.L., H.L.; project administration and funding acquisition, H.X., J.L., H.L., L.Z. All authors reviewed the manuscript.

Acknowledgments

The authors gratefully acknowledge the financial support of the Major Science and Technology Project of Yunnan Province of China Tobacco Corporation and the Graduate Research Innovation Fund of Yunnan University. We also appreciate everyone who offered us valuable guidance during the experiment. We are also grateful to the reviewers who were able to provide critical comments.

References

1. Kolhar S, Jagtap J. Plant trait estimation and classification studies in plant phenotyping using machine vision–A review[J]. Information Processing in Agriculture, 2023, 10(1): 114-135.
2. Li L, Pan X, Yang H, et al. Supervised dictionary learning with regularization for near-infrared spectroscopy classification[J]. IEEE Access, 2019, 7: 100923-100932.
3. Liu S J, Guo S, Hou Y F, et al. Chemical fingerprinting and multivariate analysis of Paeonia ostii leaves based on HPLC-DAD and UPLC-ESI-Q/TOF-MS/MS[J]. Microchemical Journal, 2023, 184: 108169.
4. Akhtar M N, Ansari E, Alhady S S N, et al. Leveraging on Advanced Remote Sensing-and Artificial Intelligence-Based Technologies to Manage Palm Oil Plantation for Current Global Scenario: A Review[J]. Agriculture, 2023, 13(2): 504.
5. Terentev A, Badenko V, Shaydayuk E, et al. Hyperspectral Remote Sensing for Early Detection of Wheat Leaf Rust Caused by Puccinia triticina[J]. Agriculture, 2023, 13(6): 1186.
6. Huang F M, Ma Q L, Ren J X, et al. Identification of Smoking-Associated Transcriptome Aberration in Blood with Machine Learning Methods[J]. BioMed research international, 2023, 2023.

7. Chen Y, Bin J, Jiang Y. Application of Machine Vision and Convolutional Neural Networks in Distinguishing of Tobacco Leaf Ma-turity by Mobile Devices[J]. Appl. Sci.
8. Amarasingam, N, Ashan Salgadoe, A.S, Powell, K, Gonzalez, L.F, Natarajan, S. A review of UAV platforms, sensors, and applications for monitoring of sugarcane crops. Remote Sens. Appl. Soc. Environ. 2022, 26, 100712.
9. Ji, Y, Chen, Z, Cheng, Q, Liu, R, Li, M, Yan, X, Li, G, Wang, D, Fu, L, Ma, Y, et al. Estimation of plant height and yield based on UAV imagery in faba bean (Vicia faba L.). Plant Methods 2022, 18, 26.
10. Gao, M, Yang, F, Wei, H, Liu, X. Individual Maize Location and Height Estimation in Field from UAV-Borne LiDAR and RGB Images. Remote Sens. 2022, 14, 2292.
11. Li Y, Zhan X, Liu S, Lu H, Jiang R, Guo W, et al. Self-Supervised Plant Phenotyping by Combining Domain Adaptation with 3D Plant Model Simulations: Application to Wheat Leaf Counting at Seedling Stage. Plant Phenomics. 2023;5:0041.
12. Syazwani, R.W.N, Asraf, H.M, Amin, M.M.S, Dalila, K.N. Automated image identification, detection and fruit counting of top-view pineapple crown using machine learning. Alex. Eng. J. 2022, 61, 1265–1276.
13. Wu, S, Deng, L, Guo, L, Wu, Y. Wheat leaf area index prediction using data fusion based on high-resolution unmanned aerial vehicle imagery. Plant Methods 2022, 18, 68.
14. Du, L, Yang, H, Song, X, Wei, N, Yu, C, Wang, W, Zhao, Y. Estimating leaf area index of maize using UAV-based digital imagery and machine learning methods. Sci. Rep. 2022, 12, 15937.

15. Syazwani, R.W.N, Asraf, H.M, Amin, M.M.S, Dalila, K.N. Automated image identification, detection and fruit counting of top-view pineapple crown using machine learning. Alex. Eng. J. 2022, 61, 1265–1276.
16. Li, Xiang, Qian, Yurong, Guo, Rui, et al. I-CenterNet: Road infrared target detection based on improved CenterNet[J].IET IMAGE PROCESSING,2023,17(1):57-66.
17. Zhou X, Wang D, Krähenbühl P. Objects as points[J]. arXiv preprint arXiv:1904.07850, 2019.
18. Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.
19. Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016,
20. Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 9199(10.5555):2969239-2969250.
21. Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
22. Wang M, Fu B, Fan J, et al. Sweet potato leaf detection in a natural scene based on faster R-CNN with a visual attention mechanism and DIoU-NMS[J]. Ecological Informatics, 2023, 73: 101931.
23. Lu H, Cao Z. TasselNetV2+: A fast implementation for high-throughput plant counting from high-resolution RGB imagery[J]. Frontiers in plant science, 2020, 11: 541960.

24. Xu X, Wang L, Liang X, et al. Maize Seedling Leave Counting Based on Semi-Supervised Learning and UAV RGB Images[J]. Sustainability, 2023, 15(12): 9583.
25. Ubbens J, Cieslak M, Prusinkiewicz P, et al. The use of plant models in deep learning: an application to leaf counting in rosette plants. Plant Methods. 2018;14(1):6.
26. Xie X, Ge Y, Walia H, et al. Leaf-counting in monocot plants using deep regression models[J]. Sensors, 2023, 23(4): 1890.
27. Jian Deng, Yanyun Ma, Deng-ao Li, et al. Classification of breast density categories based on SE-Attention neural networks[J]. Computer Methods and Programs in Biomedicine,2020.
28. Wu Y, Yao Q, Fan X, et al. Panet: A point-attention based multi-scale feature fusion network for point cloud registration[J]. IEEE Transactions on Instrumentation and Measurement, 2023.
29. Yu C, Shin Y. SAR ship detection based on improved YOLOv5 and BiFPN[J]. ICT Express, 2023.
30. Chen J, Mai H S, Luo L, et al. Effective feature fusion network in BIFPN for small object detection[C]//2021 IEEE international conference on image processing (ICIP). IEEE, 2021: 699-703.
31. HAN K, WANG Y, TIAN Q, et al. GhostNet:More features from cheap operations [C] // IEEE. Proceedings of 2020 IEEE/CVFConference on Computer Vision and Pattern Recognition. Seattle New York:IEEE, 2020.
32. XIONG Enjie, ZHANG Rongfen, LIU Yuhong, et al. Ghost-YOLOv8 Detection Algorithm for Traffic Signs[J]. Computer Engineering and Applications, 2023, 59(20): 200-207.

33. Ling S, Wang N, Li J, et al. Accurate recognition of jujube tree trunks based on CLAHE image enhancement and improved YOLOv8[J]. 2023.
34. Everingham M, Eslami S M A, Van Gool L, et al. The pascal visual object classes challenge: A retrospective[J]. International journal of computer vision, 2015, 111: 98-136.
35. https://github.com/ultralytics/ultralytics/issues/189. Accessed 25 November 2023.
36. Guo Y, Aggrey S E, Yang X, et al. Detecting broiler chickens on litter floor with the YOLOv5-CBAM deep learning model[J]. Artificial Intelligence in Agriculture, 2023, 9: 36-45.
37. Li X, Zhong Z, Wu J, et al. Expectation-maximization attention networks for semantic segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 9167-9176.
38. Liu Y, Shao Z, Hoffmann N. Global attention mechanism: Retain information to enhance channel-spatial interactions[J]. arXiv preprint arXiv:2112.05561, 2021.