



OPEN

An improved deep learning network for image detection and its application in *Dendrobii caulis* decoction piece

Yonghu Chang¹, Dejin Zhou², Yongchuan Tang³, Shuiping Ou⁴ & Sen Wang²✉

In recent years, with the increasing demand for high-quality *Dendrobii caulis* decoction piece, the identification of *D. caulis* decoction piece species has become an urgent issue. However, the current methods are primarily designed for professional quality control and supervision. Therefore, ordinary consumers should not rely on these methods to assess the quality of products when making purchases. This research proposes a deep learning network called improvedYOLOv5 for detecting different types of *D. caulis* decoction piece from images. In the main architecture of improved YOLOv5, we have designed the C2S module to replace the C3 module in YOLOv5, thereby enhancing the network's feature extraction capability for dense and small targets. Additionally, we have introduced the Reparameterized Generalized Feature Pyramid Network (RepGFPN) module and Optimal Transport Assignment (OTA) operator to more effectively integrate the high-dimensional and low-dimensional features of the network. Furthermore, a new large-scale dataset of Dendrobium images has been established. Compared to other models with similar computational complexity, improvedYOLOv5 achieves the highest detection accuracy, with an average mAP@.05 of 96.5%. It is computationally equivalent to YOLOv5 but surpasses YOLOv5 by 2 percentage points in terms of accuracy.

Keywords *Dendrobii caulis* decoction piece, YOLOv5, Deep learning, Image processing

Dendrobium is a kind of commonly used traditional Chinese herbal medicine in the genus Dendrobium Sw. of the family Orchidaceae. Dendrobium is known as “longevity herb”, located in the first of the nine Chinese herbs. Dendrobium has a long history of application in China, and was first recorded in the Shennong Bencao Jing, which is the top grade of traditional Chinese medicine. Traditional Chinese medicine theory that Dendrobium taste sweet, slightly cold, with the benefit of the stomach, nourish the yin and clear the effect of heat, often used in the treatment of fever and fluid injury, dry mouth, thirst, lack of stomach yin, less food, dry vomiting, after the disease of false heat does not go away, yin deficiency and fire, bone vapour and labour heat, dark and unclear, sinew and bone impotence and softness; modern research has shown that Dendrobium is rich in polysaccharides, alkaloids, flavonoids, phenanthrenes, benzyl, volatile oil and amino acids and other ingredients, has the following properties Antioxidant, anti-inflammatory, lowering uric acid, lowering blood sugar, anti-tumour and immunomodulation and other pharmacological effects, in the treatment of cancer, hyperuricemia, diabetes and hepatic fibrosis liver function injury, enhance immunity, alleviate physical fatigue, skin anti-aging, anti-inflammatory repair, anti-radiation and so on, has a good application prospect¹⁻⁷. There are many medicinal Dendrobium species, according to statistics, there are more than 1,500 species in the world, “Flora of China” recorded that there are about 78 species of Dendrobium spp. in China, which are mainly distributed in Guangxi, Guangdong, Guizhou, Yunnan, Fujian, Zhejiang, Jiangxi, Hunan, Anhui, etc. Among them, there are about 14 species are endemic to China, including Dendrobium hendianense, Dendrobium fanjingshanense, Dendrobium hainanense, etc.⁸. There are about 60 species of Dendrobium spp. in Yunnan Province, which is the region with the largest number of species of Dendrobium spp. in China. The Pharmacopoeia of the People's Republic of China (2020 edition) includes several species such as Dendrobium hoshanense, Dendrobium chrysotoxum,

¹School of Medical Information Engineering, Zunyi Medical University, Zunyi 563000, China. ²School of Pharmacy, Zunyi Medical University, Zunyi 563000, China. ³School of Microelectronics, Northwestern Polytechnical University, Xi'an 710072, China. ⁴Department of Pharmacy, Affiliated Hospital of Zunyi Medical University, Zunyi 563000, China. ✉email: wangsen@zmu.edu.cn

Dendrobium fimbriatum, *Dendrobium officinale*, and *Dendrobium nobile* Lindl. Among them, *D. huoshanense* is mainly produced in Huoshan County, Anhui Province in Dabie Mountain area, *D. chrysotoxum* is mainly produced in southern to western Yunnan, *D. fimbriatum* is mainly produced in southwestern to northwestern China, *D. fimbriatum* is mainly produced in Yunnan and Guizhou, and *D. nobile* Lindl. is mainly produced in Guizhou and Sichuan. At present, *Dendrobium* is used about 80,000 tons per year in China, mainly in the form of roughly processed slices of origin or slices produced in GMP-compliant workshops, which are widely used in products such as Chinese patent medicines, health care products, cosmetics, as well as in Chinese medicine for clinical use, folk health care and dietary therapy. Among them, *D. huoshanense*, *D. chrysotoxum*, *D. fimbriatum*, *D. officinale*, and *D. nobile* Lindl. are circulated in the market in the form of primary processed products of agricultural products, because of their high price, the existence of shoddy and false situations in the market, which poses a great challenge to the authenticity of ordinary consumers.

Different species of *Dendrobium* are different in plant morphology, but their traits are similar after being processed into slices. The Pharmacopoeia of the People's Republic of China (2020 edition) defines the traits of sliced *Dendrobium* species, and relevant national standards, industry standards, group standards and local standards also describe the traits of their slices. The description of the traits of this kind of medicinal herbs is a scientific and rigorous description of the terminology, which must have the expertise of botany, plant taxonomy, identification of herbs and other professionals can master and use, not applicable to non-professionals. Non-professionals have limited knowledge of *Dendrobium*, and when purchasing *Dendrobium* products, they are often faced with situations such as species being difficult to identify, substituting good for bad, and counterfeiting, which brings certain risks and economic losses to consumers.

At present, the main methods for identifying medicinal *Dendrobium* include morphological identification, microscopic identification, chromatographic identification, spectroscopic identification, molecular biology method identification, and recent rapid identification methods, such as near-infrared spectroscopic analysis^{9–14}. These methods are mainly used for product quality testing and official supervision and arbitration testing, with strict requirements for methodological procedures, testing equipment and professional knowledge, and are not suitable for ordinary consumers to identify and judge the products when purchasing.

The motivation for the identification of *D. caulis* decoction piece detection and identification based on the improved YOLOv5 method designed in this paper is as follows.

- Ordinary consumers can only identify according to their own experience when purchasing, such as from the appearance of traits, color, texture, and smell or taste, etc. However, this kind of empirical judgments requires high knowledge of the consumers, the accuracy is not very high, and there are often cases of deviation.
- The image-based intelligent detection method of *D. caulis* decoction piece is an accurate, simple and fast identification method suitable for common consumers. Consumers only need to use their cell phones to take a photo of *Dendrobium* to be able to get the detection and classification results of *Dendrobium* in a visualized way. This can not only reduce the threshold of identification, but also save consumers' time and energy.
- The image-based intelligent detection method of *D. caulis* decoction piece can minimize the possibility of substandard *D. caulis* decoction piece, standardize the market order of traditional Chinese medicine, and promote the healthy development of traditional Chinese medicine industry.

The contributions of this work are as follows. On the one hand, a YOLO-based recognition method for the detection of *D. caulis* decoction piece is proposed. On the other hand, a dataset of *Dendrobium* drinking slices detection and identification is established. The rest of the paper is organized as follows. Section 2 is a review of related research. In Section 3, a YOLO-based *D. caulis* decoction piece detection and recognition algorithm is proposed. The model training process and the result of ablation analysis are presented in Section 4. Section 5 is the conclusion of this work.

Related work

A review of related researches in this work mainly focuses on two aspects. First, methods used to recognize the *D. caulis* decoction piece. Second, deep learning algorithms and their applications in object detection.

Although there is no method for target detection of *D. caulis* decoction piece, the researchers have utilized data augmentation and machine learning techniques to detect and classify plants. Traditional detection methods primarily rely on the extraction of shape and color features, making logical judgments based on the information extracted. Traditional target detection methods include Scale-Invariant Feature Transform (SIFT)¹⁵, Histogram of Oriented Gradients (HOG)¹⁶, Support Vector Machine (SVM)¹⁷, and Selective Search for object recognition¹⁸. Raphael et al. proposed a method for detecting fruits using hue information and color variation curvature, achieving a detection success rate of 78.8%¹⁹. Chunmei et al. extracted the Otsu feature from the image, then used the Otsu threshold algorithm for automatic threshold segmentation and extracted pixels representing the fruit, with an accuracy rate of over 95%²⁰. Zhouzhou et al. improved the YOLOX model using techniques such as CSP Attention Block, SPPCSPC-F, and ASFF, resulting in a model named YOLOX-Nano, which achieved an mAP value of 84.08% for positioning²¹. Yuxiang et al. proposed a universal attention module (AGHRNet) capable of separating the background from the detected subject, which realized higher segmentation accuracy and smaller model parameters²². Mukhiddinov et al.²³ presents a deep learning system for multiclass fruit and vegetable categorization based on an improved YOLOv4 model, but there is a certain loss of accuracy. Muhammad et al.²⁴ proposing a novel DL-based methodology for the detection and classification of eight classes of weeds, but the detection rate is not that fast. Chowdhury et al.²⁵ proposed a deep learning model based on EfficientNet and they used 18,161 tomato leaf images to classify tomato diseases. However, due to the emergence of the gradient vanishing problem, which makes the network difficult to train and difficult to converge. Liu et al.²⁶ proposes a novel

framework that combines hyperspectral imaging (HSI) and deep learning techniques for plant image classification. But it does not perform well in small target scenarios. Teng et al.²⁷ proposes a robust pest detection network based on RCNN. But it's slow to detect. Wagle et al.²⁸ proposed a CNN model with transfer learning from AlexNet to detect nine species of plants from the PlantVillage dataset. But it's more computationally intensive.

Many computer vision algorithms based on CNN and deep learning have been proposed and have been proven to be successful in the recognition and classification of real-world objects^{29,30}. Some significant advancements in the field of computer vision, specifically in object detection, have been primarily focused on the RCNN³¹ series, YOLO³² series, and SSD³³ series algorithms. R-CNN series includes R-CNN, Fast R-CNN³⁴, Faster R-CNN³⁵, and Mask R-CNN³⁶. These methods achieve object detection through a process of region proposal extraction and region classification. YOLO series includes YOLO³⁷, YOLOv2³⁸, YOLOv3³⁹, YOLOv4⁴⁰, and so on. YOLO algorithms transform the object detection task into a regression problem and perform dense predictions directly on the image, enabling real-time object detection. SSD³³ algorithm employs multiple scales of convolutional filters applied to feature maps at different levels to achieve object detection at various scales. It should be noted that the agricultural industry has turned to DL-based models to address these challenges. Deep learning approaches have achieved state-of-the-art results in tasks such as plant identification, fruit harvesting, and pest and disease control.

Methods

General research idea

Aiming at the application requirements of *D. caulis* decoction piece detection and recognition concerning real time and accuracy in actual Buying and selling scenarios, this paper proposes an improved YOLOv5 *D. caulis* decoction piece detection and recognition method for processing *D. caulis* decoction piece images collected by Phone camera. The overall framework is shown in Figure 1.

This study encompasses several key components, namely data collection, data annotation, data augmentation, and the development of an improved YOLOv5-based system for recognizing and classifying *D. caulis* decoction piece. Initially, photographs of *D. caulis* decoction piece were acquired using the smartphone's camera, followed by the annotation process for these images. The collected images were then utilized to construct a comprehensive dataset, which underwent data augmentation techniques to enhance its diversity. Data augmentation methods were performed using an online approach and include random perspective, and HSV adjustments. Subsequently, the improved YOLOv5 model was deployed to accurately recognize and classify *D. caulis* decoction piece. Finally, the identified *D. caulis* decoction piece types were visually presented on the smartphone screen as images.

YOLO-based meter detection and recognition algorithm

The recognition of *D. caulis* decoction piece imposes specific real-time performance requirements, necessitating the selection of mature object detection methods. Among these methods, the YOLO series, as a one-stage approach, exhibits faster detection speed compared to the two-stage RCNN series. Within the YOLO series, the YOLOv5 algorithm has emerged as a superior object detection algorithm due to its optimal trade-off between accuracy and speed. Compared to classic algorithms such as YOLOv3³⁹ and YOLOv4⁴⁰, YOLOv5⁴¹ boasts a more advanced network architecture that offers improved performance characteristics. In contrast to more recent and sophisticated models like YOLOv7⁴² and YOLOv8⁴³, it employs a more lightweight architectural design, enabling it to achieve the desired performance on our dataset with a significantly reduced computational footprint. Therefore, this paper proposes an improved YOLOv5 algorithm for the identification of *D. caulis* decoction piece.

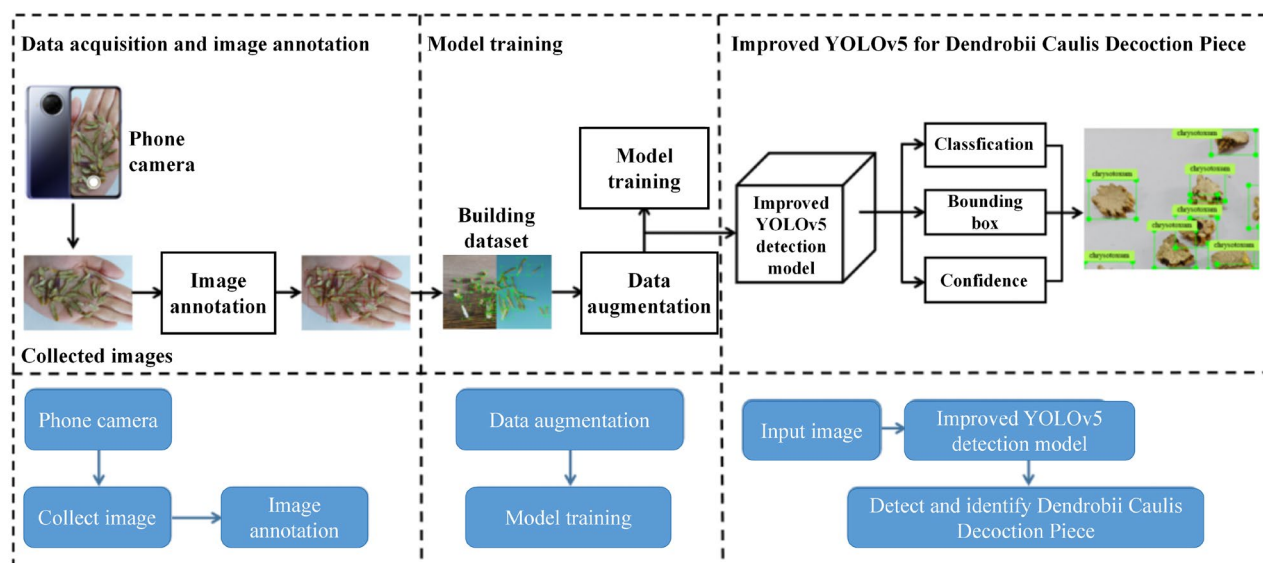


Figure 1. Framework of *D. caulis* decoction piece detection method based on deep learning.

YOLOv5

YOLOv5 employs CSPDarknet53 as its backbone network for extracting image features. It utilizes an FPN network to fuse three distinct output layer features from the backbone network. It uses Mish activation function and the Focal Loss loss function to enhance the model's performance. The overall structure of the YOLOv5 network is shown in Figure 2.

Improved YOLOv5

Given that YOLOv5 is a lightweight network within the deep learning domain, efforts have been made to enhance network accuracy without significantly compromising network speed. The following improvements have been implemented to achieve this objective. The improved YOLOv5 network structure is shown in Figure 3.

- this paper designed a C2S module and add it to YOLOv5 Backbone. The C2S module interacts the feature maps of the current layer with deeper layers, leveraging semantic information from deeper layers to capture the position and detail information of small *D. caulis* targets. This enables the network to adapt to *D. caulis* targets with varying scales.
- This paper introduces the RepGFPN (Repeating Grouped Feature Pyramid Network) module to better utilize feature maps at different scales. The RepGFPN module divides the feature maps into multiple groups, performs feature fusion within each group, and then cascades the fusion results from different groups to achieve more effective feature fusion.
- The loss function incorporates optimal transport assignment, a dynamic label assignment method, into YOLOv5 Loss function to better handle class imbalance and varying target sizes.

Due to the small proportion of pixels occupied by small objects in images, the number of pixels in the feature maps obtained during the feature extraction process of convolutional neural networks gradually decreases after multiple downsampling operations. For instance, when using a stride of 16, a target region of size 32×32 pixels reduces to only 2×2 pixels in the feature map. This results in the loss of effective spatial information for detecting small objects, making it challenging to accurately detect them. Furthermore, as the network depth increases, the feature and positional information of small objects gradually diminish, further reducing the detection capability and localization accuracy of convolutional neural networks for small objects^{44,45}. For tiny objects smaller than 10×10 pixels, their target features become extremely weak or may even disappear after eight downsampling operations. Inspired by the YOLOv8, we made modifications to the C3 module by removing one convolutional block, adding multiple bottleneck layers, and introducing shortcut connections that combine shallow and deep features. This enhancement aims to strengthen the feature representation capability for small objects and improve their detection accuracy.

The feature pyramid network (FPN) is designed to aggregate features of different resolutions extracted from the backbone network, which has been proven to be a crucial and effective component in object detection^{46–48}. FPN⁴⁶ fuses feature maps from different levels through top-down feature propagation and lateral connections. However, it still suffers from the loss of features related to small objects. PAFPN⁴⁷ introduces an additional bottom-up pathway aggregation network at the cost of increased computational complexity. BiFPN⁴⁸ combines

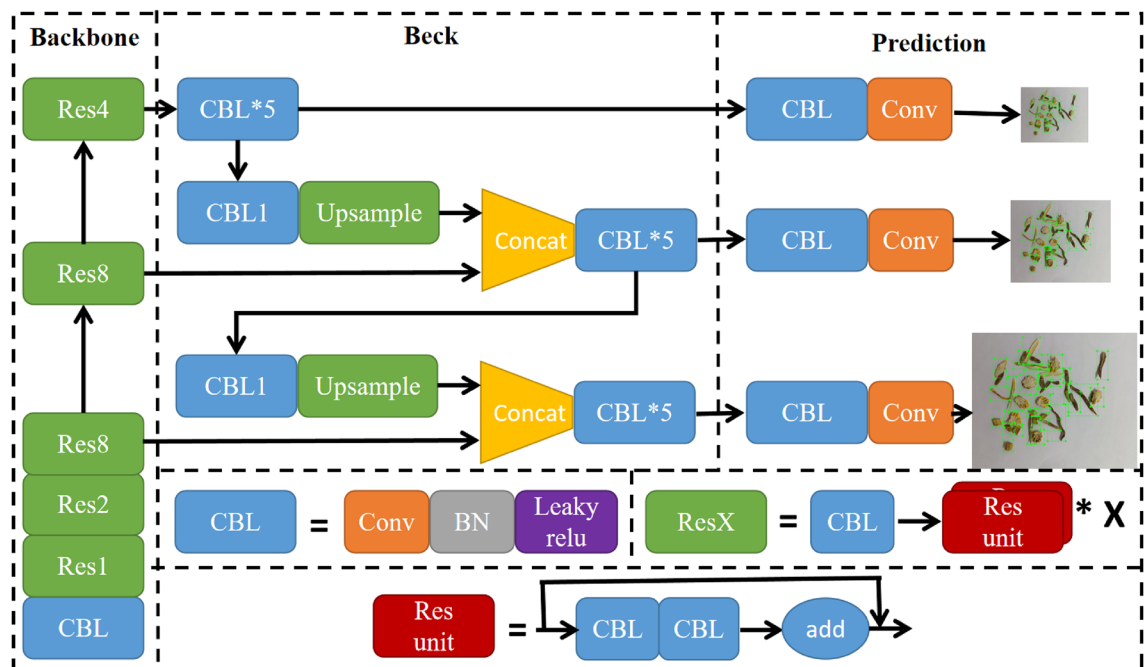


Figure 2. Framework of YOLOv5.

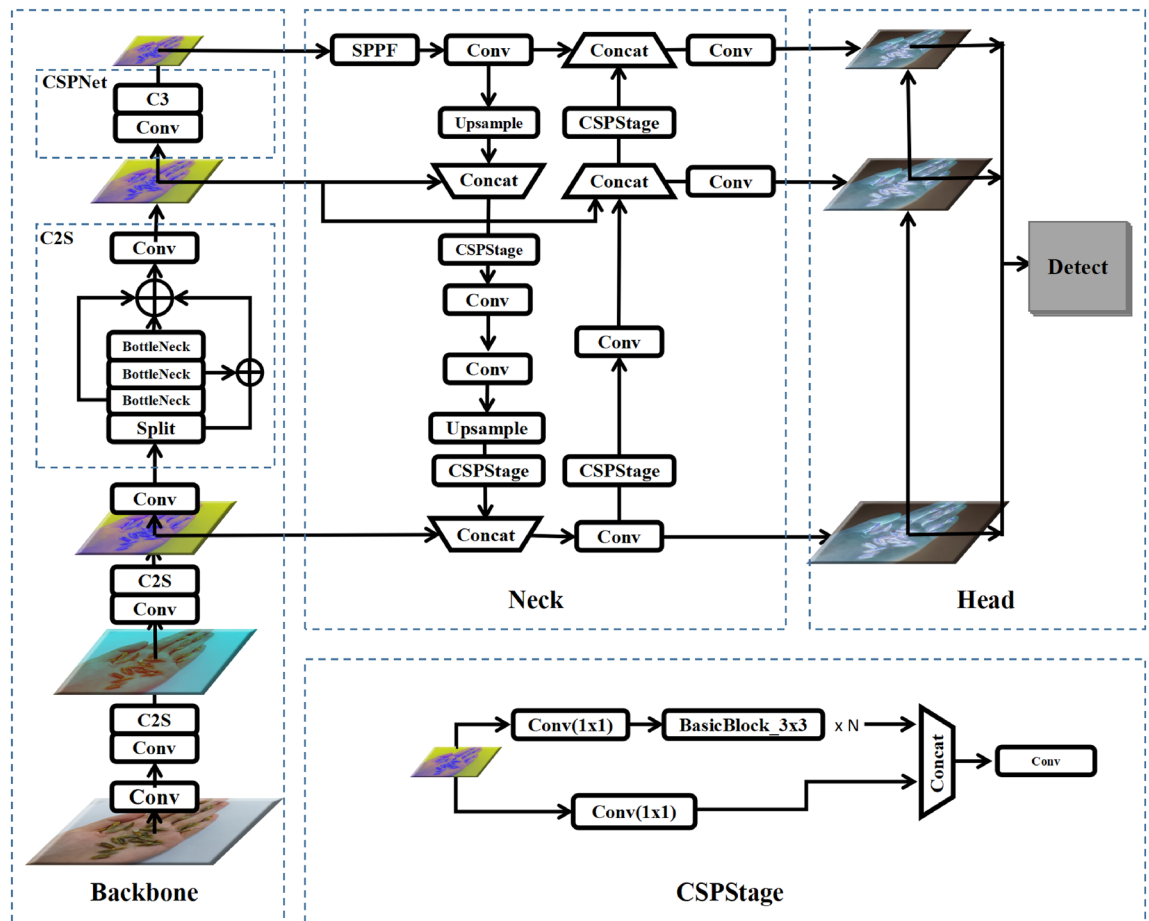


Figure 3. Improved YOLOv5 network structure.

bottom-up and top-down feature propagation along with lateral connections to fuse feature maps at different levels and scales. This enables the model to extract object features at multiple scales and detect objects of varying sizes. Generalized-FPN (GFPN) has been proposed as the neck layer, achieving higher efficiency by effectively exchanging high-level semantic information and low-level spatial information. RepGFPN⁴⁹, by flexibly controlling the depth and width of different channels using the CSPStage structure, achieves state-of-the-art performance with higher accuracy. We introduce the RepGFPN network into the neck layer of YOLOv5 to enhance performance. To reduce the computational overhead of RepGFPN without compromising the feature extraction capability of the backbone layer, we add a dimension reduction operation at the front end of the RepGFPN network. The features undergo the first convolutional operation, followed by halving the number of channels before entering the RepGFPN layer. This ultimately improves the network accuracy.

In the field of deep learning, particularly in the task of object detection, the YOLO (You Only Look Once) series of algorithms have garnered widespread attention due to their efficiency and accuracy. As an advanced variant within this series, YOLOv5's loss function design has played a crucial role in enhancing the performance of the model. The loss function of YOLOv5 is a combination of multi-task losses, which simultaneously considers the classification, localization, and confidence of the targets. Specifically, the classification loss employs cross-entropy to measure the accuracy of class predictions. The localization loss uses Mean Squared Error (MSE) to gauge the discrepancy between the predicted bounding boxes and the actual bounding boxes. The confidence loss evaluates the model's predictive confidence in the existence of the targets^{50,51}. The formula is shown as follows:

$$L_{total} = \sum_i^N \lambda_1 L_{box} + \lambda_2 L_{obj} + \lambda_3 L_{cls} \quad (1a)$$

$$= \sum_i^N \left(\lambda_1 \sum_j^{B_i} L_{LCIoU} + \lambda_2 \sum_j^{S_i \times S_i} L_{obj_j} + \lambda_3 \sum_i^{B_i} L_{cls_j} \right) \quad (1b)$$

In the given formula, N refers to the number of detection layers, B represents the number of targets assigned to the prior boxes, and $S \times S$ denotes the number of grid cells into which the scale is divided. L_{box} is the loss for bounding box regression, which is calculated for each object; L_{obj} stands for the objectness loss, which is

computed for each grid cell; L_{cls} signifies the classification loss, also calculated for each object. λ_1 , λ_2 , and λ_3 are the weights for these three respective losses.

Experiment

Data set construction

The images of *D. caulis* decoction piece were captured in Zunyi, China in November 2022. To ensure that the dataset covers a wide range of realistic lighting scenarios for *D. caulis* decoction piece identification, we considered different indoor lighting conditions as well as natural indoor lighting on sunny and cloudy days. To diversify the dataset, we included various background settings that are relevant to practical application scenarios, such as the palm of a hand, white paper, and different textured and colored tabletops. These images were captured using the cameras of the Xiaomi Note 11 at different angles and distances ranging from 0.3 to 0.5 meters. In total, 7,118 images of different *D. caulis* decoction piece were captured. Although the image sizes in the dataset are inconsistent, we applied a normalization step during deep neural network training to standardize all images to a fixed resolution of 640×640. According to the Chinese Pharmacopoeia, there are five species of *D. caulis* decoction piece, including *Dendrobium chrysotoxum*, *Dendrobium huoshanense*, *Dendrobium nobile* Lindl., and *Dendrobium nobile* Lindl., all of which are included in our dataset. Each *D. caulis* decoction piece species was photographed individually as well as in combination with other species. To train a network with enhanced discriminative capabilities, we also captured a set of photographs containing a mixture of all different species of *D. caulis* decoction piece.

Evaluation metrics

For our dataset, each detected bounding box can be categorized into three scenarios. True Positives (TP) represent the detected bounding boxes whose intersection over union (defined as the ratio of intersection area to union area) with their corresponding ground truth bounding boxes is greater than 50%. False Positives (FP) represent the detected bounding boxes whose intersection over union is less than 50% with their corresponding ground truth bounding boxes. False Negatives (FN) represent the ground truth bounding boxes that are not covered by any detected bounding box. Precision reflects the accuracy of the model among all detected bounding boxes. It is defined as the ratio of the number of TP to the total number of detected bounding boxes. Recall reflects the model's ability to cover all the ground truth bounding boxes. The formulas for precision (Prec), recall (Rec) are as follows:

$$Prec = \frac{TPs}{TPs + FPs} \quad (2)$$

$$Recall = \frac{TPs}{TPs + FNs} \quad (3)$$

The definition of mean Average Precision (mAP) is as follows:

$$mAP = \frac{1}{C} \sum_{i=1}^C Precision(i) \quad (4)$$

In the formula (4), C represents the total number of categories in the Shihu dataset. $Prec(i)$ (represented by equation a(2)) denotes the precision for each category of Shihu.

Training details

We implemented improved YOLOv5 using PyTorch with Python version 3.8.0 and Torch version 1.13.1+cu1116. The training was performed on a single GPU (Nvidia RTX 3090). The improved YOLOv5 model was executed on a computer running Ubuntu 20.04 operating system with an Intel(R) Xeon(R) Silver 4210 CPU. The initial learning rate and learning rate scaling factor were both set to 0.01. Before the actual training, there was a warm-up period of 3 epochs, and the mini-batch size was set to 64. We utilized the Adam optimizer with a SGD momentum rate of 0.937, weight decay rate of 0.005, and a warm-up initial momentum rate of 0.8. The training process lasted for 300 epochs.

In order to prevent overfitting, we carefully considered the different angles of placement for *D. caulis* decoction piece and appropriately applied data augmentation algorithms during the model training process. The dataset comprises a sufficient number of images and was divided into 70% for training, 15% for validation, and 15% for testing. Specifically, the training set contains 4,990 images, the test set consists of 1,064 images, and the validation set comprises 1,064 images. Each set includes proportional representations of single herbs image, multiple herbs image, and mixed herbs image, with backgrounds consisting of palm, white paper, and various textures and colors of tabletops. To ensure that each training iteration receives a unique set of data augmentation effects, we have implemented online data augmentation, applying augmentations in real-time during the training process, rather than pre-applying augmentations and expanding the dataset beforehand. In addition, considering that end users may use the proposed algorithm to identify the *Dendrobium* under different lighting conditions, angles, and shooting distances, we applied two data augmentation techniques to the dataset: Random Perspective and HSV adjustments. For Random Perspective, we set the random rotation angle to range from -90 to +90 degrees, random translation along the X and Y axes with a magnitude of 0.1, random scaling with a factor of 0.8, and a random perspective transformation intensity of 0.001. As for HSV adjustments, we set the values of hsv_h to 0.015, hsv_s to 0.7, and hsv_v to 0.4. All image augmentation processes were implemented using the

Albumentations library in Python. Following the convention, the network was trained for 300 epochs, during which the loss fluctuated within a small range, indicating convergence of the network.

Quantitative results

In order to demonstrate the effectiveness of improved YOLOv5, we compared it only with the detection models from the YOLO series. This is because currently, the YOLO series exhibits the best performance in various image object detection applications. The models we compared against YOLOv5⁴¹. Table 1 reports the quantitative results on our test images. Our improved YOLOv5 achieved the best performance in three out of the four metrics (Prec, mAP50, and mAP95) under the premise of relatively lower computational complexity.

In addition, to demonstrate the contribution of each module to the overall performance of improved YOLOv5, we individually integrated the OTA, RepGFPN, and C2S modules into YOLOv5 by replacing the corresponding components. Table 2 reports the quantitative results of the mentioned approaches. Both C2S, OTA, and RepGFPN outperformed YOLOv5 in terms of mAP50 and mAP50-95.

we conducted an in-depth analysis of the performance of the YOLOv5, improved YOLOv5, YOLOv4, and YOLOv3 models on a our dataset and further observed the actual performance of the models through visualization techniques. To uncover the key regions that the models rely on for the identification and localization of objects of different categories within images, we employed the XGradCAM technology, specifically visualizing the last convolutional layer of the neck of each model. Utilizing heatmaps, we illustrated the areas of interest that the models focus on during the decision-making process, where red indicates high attention from the model, and blue signifies areas of relative neglect. The visualization results are shown in Figure 4.

Through comparative analysis, we found that YOLOv5 and its improved version performed excellently in detecting *Dendrobium officinale* slices, accurately identifying and classifying all samples. In contrast, YOLOv3 and YOLOv4 exhibited omissions during the detection process. Moreover, the YOLOv5 and its improved version demonstrated a significantly higher coverage of the regions of interest in the image compared to YOLOv3 and YOLOv4, indicating their superior capability in object localization. Particularly, in the improved YOLOv5, we observed a notable increase in the confidence of predictions compared to the standard YOLOv5, and the red areas in the heatmap corresponded more closely with the *D. caulis* decoction piece. This finding further confirms the effectiveness of the improved YOLOv5 in comprehending image features and enhancing detection accuracy.

Ablation analysis

To demonstrate that YOLOv5 is the optimal combination of all the modules, we conducted a simple yet effective ablation analysis on the dataset. The results of all the ablation analyses are shown in Table 3. We compared the complete YOLOv5 model with the “YOLOv5,” “C2S+RepGFPN,” and “C2S+OTA” models using precision, recall, and mAP on the same dataset. The fully expanded YOLOv5 exhibited the best performance among the ablation comparisons.

This paper focus on the research involving the use of deep learning models to achieve high-accuracy detection or recognition of different plants or fruits. Zhou et al.⁴² used a PSPNet to detect the endpoints of the dragon fruit, including dragon fruit segmentation and position, achieved an accuracy of around 95%. On the other hand, Huang et al.⁵² designed a deep learning network that combines UAV data collection, AI embedded device, and target detection algorithm to detection citrus with an accuracy of 93.32%. Likewise, Parico et al.^{53–55} used machine learning algorithms to accurately identify plants or fruits. In our work, improved YOLOv5 achieved an average mAP of 95.73% for multiple *D. caulis* decoction piece. the accuracy and mAP of the model are up, which is improved compared with the original baseline. Together with our improved YOLOv5, the above works disclose the popularity and the broad application prospects of machine learning and deep learning on fruit and plant detection.

Methods	Prec	Rec	mAP@0.5	mAP@0.5:0.95
YOLOv3	0.908	0.892	0.933	0.535
YOLOv4	0.858	0.939	0.674	0.475
YOLOv5	0.922	0.925	0.945	0.590
Improved YOLOv5(ours)	0.928	0.960	0.965	0.621

Table 1. The quantitative comparison of several methods including YOLOv5 on the test dataset.

Methods	Prec	Rec	mAP@0.5	mAP@0.5:0.95
YOLOv5	0.922	0.925	0.945	0.590
C2S	0.926	0.931	0.955	0.593
OTA	0.923	0.954	0.963	0.625
RepGFPN	0.929	0.931	0.957	0.607

Table 2. The results of the peeling test of the network on the *D. caulis* decoction piece detection task.

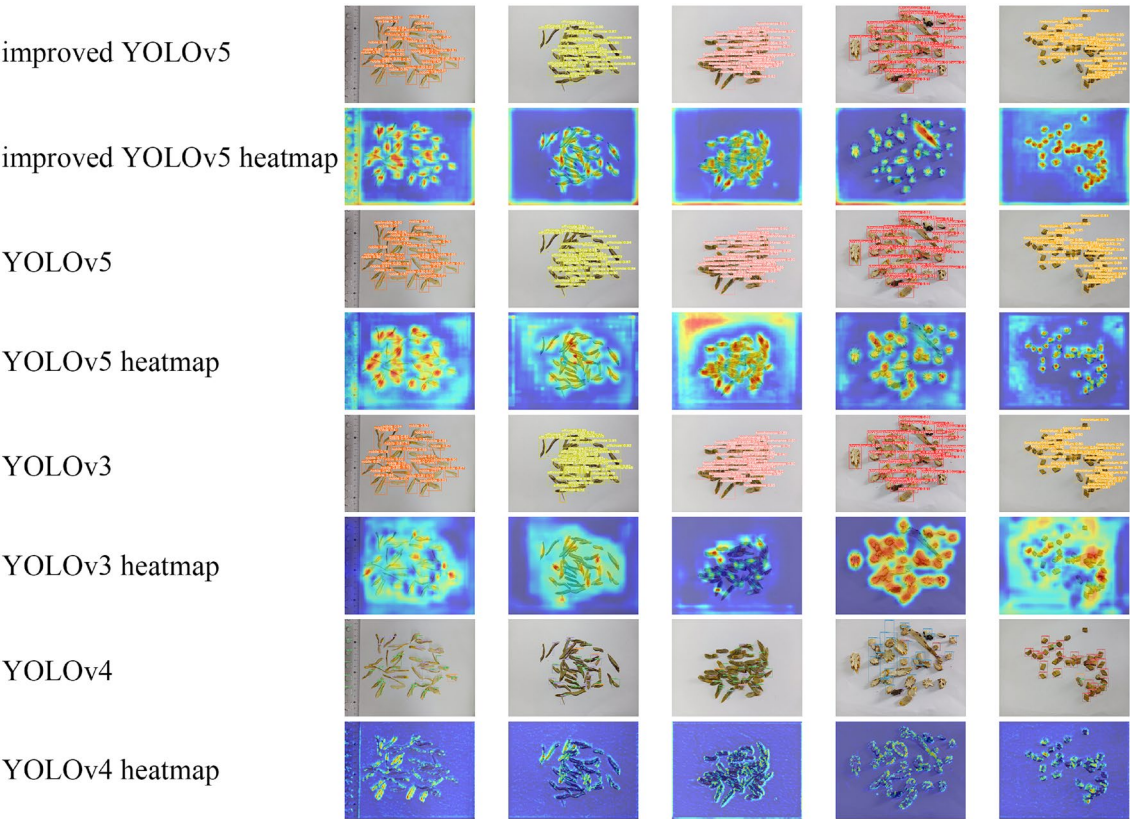


Figure 4. The visualization results of the model’s localization and classification.

Methods	Prec	Rec	mAP@0.5	mAP@0.5:0.95
YOLOv5	0.922	0.925	0.945	0.590
Imporved YOLOv5(C2S+OTA)	0.922	0.925	0.955	0.620
Imporved YOLOv5(C2S+RepGFPN)	0.928	0.928	0.953	0.605
Imporved YOLOv5(C2S+GFPN+OTA)	0.928	0.960	0.965	0.621

Table 3. Experimental results of different combination models. *The best measures are in bold.

Ethical approval

The research was approved by the Guizhou Provincial Science and Technology Support Project (Program No. Qian Science Support [2018] 2804) including the permission to collect *D. caulis*. all the methods were carried out in accordance with relevant Institutional guidelines and regulations. Informed consent was obtained from all participants.

Conclusions

This paper presents improved YOLOv5, a model for detecting and classifying *D. caulis* decoction piece, aiming to assist consumers unfamiliar with *D. caulis* decoction piece in quickly identifying the species using mobile devices such as smartphones. The network improves the capability of extracting features from small objects by introducing the C2S layer to replace the original C3 layer. It enhances the detection efficiency of the network by incorporating the OTA algorithm into the loss function. Additionally, the RepGFPN module is introduced in the feature fusion stage to better fuse shallow and deep features, achieving more effective feature fusion. We established a dataset and validated the effectiveness of the proposed method. The experiments demonstrate significant improvements in dense small object detection tasks compared to other state-of-the-art methods. The performance of the model can be attributed to the combination of learned shallow features and attention features, enabling our model to detect more small objects based on low resolution and weak features, thereby improving the recall rate of targets in dense and occluded scenes to some extent.

On one hand, for dense small objects, especially those with occlusions, our algorithm can improve their recall rate, but there are still undetected targets. Therefore, in future work, we will focus on addressing the detection of dense and occluded targets, such as using better post-processing mechanisms. On the other hand, compared to YOLOv5-Lite, we achieve better detection results but at a slightly slower speed and with higher computational

complexity. Hence, we will further investigate methods to alleviate our approach and improve real-time detection speed. For example, depth-wise separable convolutions and lighter backbones can be explored as alternatives to the backbone of our method.

Data and code availability

Data and code are available from Sen Wang upon reasonable request.

Received: 22 September 2023; Accepted: 28 May 2024

Published online: 12 June 2024

References

- Lin, Y. *et al.* Anti-inflammatory phenanthrene derivatives from stems of dendrobium denneanum. *Phytochemistry* **95**, 242–251 (2013).
- Zhou, C., Xie, Z., Lei, Z., Huang, Y. & Wei, G. Simultaneous identification and determination of flavonoids in dendrobium officinale. *Chem. Cent. J.* **12**, 1–9 (2018).
- Duan, H. *et al.* Alkaloids from dendrobium and their biosynthetic pathway, biological activity and total synthesis. *Phytomedicine* **102**, 154132 (2022).
- Zhang, X. *et al.* Polysaccharides from dendrobium officinale inhibit proliferation of osteosarcoma cells and enhance cisplatin-induced apoptosis. *J. Funct. Foods* **73**, 104143 (2020).
- Ren, G. *et al.* Bibenzyl derivatives from leaves of dendrobium officinale. *Nat. Prod. Commun.* **15**, 1934578X20908678 (2020).
- Zhou, H., Zhou, L., Li, B. & Yue, R. Anti-cyclooxygenase, anti-glycation, and anti-skin aging effect of dendrobium officinale flowers' aqueous extract and its phytochemical validation in aging. *Front. Immunol.* **14**, 1095848 (2023).
- Wang, Y.-H. Traditional uses, chemical constituents, pharmacological activities, and toxicological effects of dendrobium leaves: A review. *J. Ethnopharmacol.* **270**, 113851 (2021).
- Teixeira da Silva, J. A. & Ng, T. B. The medicinal and pharmaceutical importance of dendrobium species. *Appl. Microbiol. Biotechnol.* **101**, 2227–2239 (2017).
- Yuan, Y., Liu, X., Wang, J. & Zhang, J. Morphological and microscopic identification of three major medicinal dendrobium species in ta-pieh mountains area. *Microsc. Res. Tech.* **82**, 483–493 (2019).
- Yin, L. *et al.* A review of the application of near-infrared spectroscopy to rare traditional chinese medicine. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **221**, 117208 (2019).
- Chen, W., Chen, X., Xu, J., Cai, J. & Wang, X. Identification of dendrobium officinale using dna barcoding method combined with hrm and qpcr technology. *Food Anal. Methods* 1–11 (2022).
- Du, G. *et al.* Use of microscopic characteristics and multielemental fingerprinting analysis to trace three different cultivation modes of medicinal and edible dendrobium officinale in china. *Biol. Trace Elem. Res.* **201**, 1006–1018 (2023).
- Shu, W. *et al.* Transcriptomic and metabolomic analyses reveal differences in flavonoid pathway gene expression profiles between two dendrobium varieties during vernalization. *Int. J. Mol. Sci.* **24**, 11039 (2023).
- Liu, F. *et al.* Identification of dendrobium varieties by infrared spectroscopy. *Spectrosc. Spectr. Anal.* **34**, 2968–2972 (2014).
- Ng, P. C. & Henikoff, S. Sift: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
- Albiol, A., Monzo, D., Martin, A., Sastre, J. & Albiol, A. Face recognition using hog-ebgm. *Pattern Recogn. Lett.* **29**, 1537–1543 (2008).
- Wang, H., Li, G. & Wang, Z. Fast svm classifier for large-scale classification problems. *Inf. Sci.* **642**, 119136 (2023).
- Yusro, M. M., Ali, R. & Hitam, M. S. Comparison of faster r-cnn and yolov5 for overlapping objects recognition. *Baghdad Sci. J.* **20**, 0893–0893 (2023).
- Linker, R., Cohen, O. & Naor, A. Determination of the number of green apples in rgb images recorded in orchards. *Comput. Electron. Agric.* **81**, 45–57. <https://doi.org/10.1016/j.compag.2011.11.007> (2012).
- Wei, X. *et al.* Automatic method of fruit object extraction under complex agricultural background for vision system of fruit picking robot. *Optik* **125**, 5684–5689. <https://doi.org/10.1016/j.jleoe.2014.07.001> (2014).
- Zheng, Z., Hu, Y., Li, X. & Huang, Y. Autonomous navigation method of jujube catch-and-shake harvesting robot based on convolutional neural networks. *Comput. Electron. Agric.* **215**, 108469 (2023).
- Zheng, Z. *et al.* Aghrnet: An attention ghost-hrnet for confirmation of catch-and-shake locations in jujube fruits vibration harvesting. *Comput. Electron. Agric.* **210**, 107921 (2023).
- Mukhiddinov, M., Muminov, A. & Cho, J. Improved classification approach for fruits and vegetables freshness based on deep learning. *Sensors* **22**, 8192 (2022).
- Saleem, M. H., Velayudhan, K. K., Potgieter, J. & Arif, K. M. Weed identification by single-stage and two-stage neural networks: A study on the impact of image resizers and weights optimization algorithms. *Front. Plant Sci.* **13**, 850666 (2022).
- Chowdhury, M. E. *et al.* Automatic and reliable leaf disease detection using deep learning techniques. *AgriEngineering* **3**, 294–312 (2021).
- Liu, K.-H., Yang, M.-H., Huang, S.-T. & Lin, C. Plant species classification based on hyperspectral imaging via a lightweight convolutional neural network model. *Front. Plant Sci.* **13**, 855660 (2022).
- Teng, Y., Zhang, J., Dong, S., Zheng, S. & Liu, L. Msr-rcnn: A multi-class crop pest detection network based on a multi-scale super-resolution feature enhancement module. *Front. Plant Sci.* **13**, 810546 (2022).
- Wagle, S. A., Hari Krishnan, R., Ali, S. H. M. & Faseehuddin, M. Classification of plant leaves using new compact convolutional neural network models. *Plants* **11**, 24 (2021).
- Sarker, I. H. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput. Sci.* **2**, 420 (2021).
- Zhao, M. *et al.* Faster mean-shift: Gpu-accelerated clustering for cosine embedding-based cell segmentation and tracking. *Med. Image Anal.* **71**, 102048 (2021).
- Xie, X., Cheng, G., Wang, J., Yao, X. & Han, J. Oriented r-cnn for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, 3520–3529 (2021).
- Jiang, P., Ergu, D., Liu, F., Cai, Y. & Ma, B. A review of yolo algorithm developments. *Procedia Comput. Sci.* **199**, 1066–1073 (2022).
- Liu, W. *et al.* Ssd: Single shot multibox detector. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I* **14**, 21–37 (Springer, 2016).
- Girshick, R. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 1440–1448 (2015).
- Ren, S., He, K., Girshick, R. & Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Proc. Syst.* **28** (2015).
- He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 2961–2969 (2017).

37. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788 (2016).
38. Sang, J. *et al.* An improved yolov2 for vehicle detection. *Sensors* **18**, 4272 (2018).
39. Masurekar, O., Jadhav, O., Kulkarni, P. & Patil, S. Real time object detection using yolov3. *Int. Res. J. Eng. Technol. (IRJET)* **7**, 3764–3768 (2020).
40. Bochkovskiy, A., Wang, C.-Y. & Liao, H.-Y. M. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint [arXiv:2004.10934](https://arxiv.org/abs/2004.10934) (2020).
41. Ultralytics. ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation. <https://github.com/ultralytics/yolov5.com>, <https://doi.org/10.5281/zenodo.7347926> (2022). Accessed: 7th May, 2023.
42. Zhou, J., Zhang, Y. & Wang, J. A dragon fruit picking detection method based on yolov7 and psp-ellipse. *Sensors* **23**, 3803 (2023).
43. Talaat, F. M. & ZainEldin, H. An improved fire detection approach based on yolo-v8 for smart cities. *Neural Comput. Appl.* **35**, 20939–20954 (2023).
44. Shrivastava, A. & Gupta, A. Contextual priming and feedback for faster r-cnn. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I* **14**, 330–348 (Springer, 2016).
45. Ye, J., Yuan, Z., Qian, C. & Li, X. Caa-yolo: Combined-attention-augmented yolo for infrared ocean ships detection. *Sensors* **22**, 3782 (2022).
46. Ghiasi, G., Lin, T.-Y. & Le, Q. V. Nas-fpn: Learning scalable feature pyramid architecture for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7036–7045 (2019).
47. Wang, K., Liew, J. H., Zou, Y., Zhou, D. & Feng, J. Panet: Few-shot image semantic segmentation with prototype alignment. In *proceedings of the IEEE/CVF international conference on computer vision*, 9197–9206 (2019).
48. Tan, M., Pang, R. & Le, Q. V. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10781–10790 (2020).
49. Xianzhe *et al.* Damo-yolo: A report on real-time object detection design. arXiv preprint [arXiv:2211.15444v2](https://arxiv.org/abs/2211.15444v2) (2022).
50. Murthy, J. S. *et al.* Objectdetect: A real-time object detection framework for advanced driver assistant systems using yolov5. *Wireless Commun. Mobile Comput.* **2022** (2022).
51. Mahaur, B. & Mishra, K. Small-object detection based on yolov5 in autonomous driving systems. *Pattern Recogn. Lett.* **168**, 115–122 (2023).
52. Huang, H., Huang, T., Li, Z., Lyu, S. & Hong, T. Design of citrus fruit detection system based on mobile platform and edge computer device. *Sensors* **22**, 59 (2021).
53. Parico, A. I. B. & Ahamed, T. Real time pear fruit detection and counting using yolov4 models and deep sort. *Sensors* **21**, 4803 (2021).
54. Kromdijk, J. *et al.* Crop management impacts the efficiency of quantitative trait loci (qtl) detection and use: case study of fruit load× qtl interactions. *J. Exp. Bot.* **65**, 11–22 (2014).
55. Jin, X. *et al.* A novel deep learning-based method for detection of weeds in vegetables. *Pest Manag. Sci.* **78**, 1861–1869 (2022).

Author contributions

Conceptualization, S.W., Y.C. and Y.T.; methodology, Y.C. and S.W.; software, Y.C.; validation, S.W.; investigation, S.O.; data acquisition, D.Z. and S.O.; writing-original draft preparation, Y.C. and D.Z.; writing-review and editing, Y.C., S.W. and Y.T.; visualization, Y.C.; supervision, S.W.; funding acquisition, S.W. and S.O. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported in part by the Guizhou Provincial Science and Technology Support Project (No.Qian Science Support [2018] 2804) and the Scientific Research Project of Guizhou Dendrobium Industry Development Research Center No.QSKH[2019003]-002 and the Guizhou Province Education Department, Characteristic Region Project (No.Qian Education KY [2021] 072) and Future Master Medical Technician Talent Cultivation Program of Zunyi Medical University (20221040) and The Project of Guizhou Provincial Health Commission (No.gzkwj2021-138).

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024