Recognition of Chinese wolfberry images with windy and sandy noises using improved YOLOv8

Fengque Pei[1], Zhi Li[1,2], Song Mei[2,3*], Zhiyu Song[2], Zhigang Shi[4*], Dunbing Tang[3], Ru Wan[4]

(1. *College of Mechanical and Electrical Engineering, Hohai University, Nanjing 210000, China*;
2. *Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing 210014, China*;
3. *College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*;
4. *Chinese Wolfberry Science Institute, Ningxia Academy of Agriculture and Forestry Sciences, Yinchuan 210037, China*)

Abstract: With the nature of the high wind and sand in western China, the Chinese wolfberry recognition shows a strong relationship with the sandy noise and needs a high-accuracy algorithm. To address this issue, this study aimed to develop an algorithm for accurately detecting and recognizing wolfberries. YOLOv8, an algorithm promoted by Ultralytics, supports image classification, object detection, and instance segmentation tasks. To enhance the performance of the original YOLOv8 model, a novel YOLOv8 algorithm incorporating FasterNet, RepBiFPN, and Lightweight Asymmetric Dual-Head was proposed. Firstly, thousands of Chinese wolfberry images were collected from the Ningxia Academy of Agriculture and Forestry Science, China, and random noises were added to simulate the wind and sand conditions typical of spring. Secondly, leveraging the advantages of YOLOv8n, such as its high speed and accuracy, this research innovatively integrated the FasterNet block into the C2f module of YOLOv8 to improve the effective handling of data uncertainty and noise. Additionally, an innovative RepViT+BiFPN, a new detective head, and a Lightweight Asymmetric Dual-Head were introduced to improve the training efficiency of the YOLOv8 network. Finally, to evaluate the effectiveness of improved YOLOv8 for the recognition of wolfberry, the dataset of wolfberry images was divided into a training set, a validation set, and a testing set to assess the performances of different models. Experiment results demonstrate that the YOLOv8-FasterNet+LADH+RepBiFPN model outperforms other models in terms of mAP@0.50-0.95, achieving a 4.5% improvement on the validation set compared to the original YOLOv8n. This research addresses the high-speed and accurate recognition of the Chinese wolfberry under strong winds and sand noise through algorithmic improvements and integration, which can facilitate the automation and intelligence of Chinese wolfberry harvesting and contribute to the advancement of agricultural mechanization.

Keywords: image recognition, Chinese wolfberry, windy and sandy noise, YOLOv8, attention module
DOI: 10.25165/j.ijabe.20251802.9052

Citation: Pei F Q, Li Z, Mei S, Song Z Y, Shi Z G, Tang D B, et al. Recognition of Chinese wolfberry images with windy and sandy noises using improved YOLOv8. Int J Agric & Biol Eng, 2025; 18(2): 239–247.

1 Introduction

With the increasing market demand and the strong support from government policy aimed at promoting agriculture, the planting area of Chinese wolfberry has been expanding significantly. China, as a major exporter of wolfberry, had an approximate planting area of 29 677 hm^2 by the end of 2022. The annual production reached 86 300 t, accounting for over 95% of the global output, and the export volume amounted to 11 932.24 t. From January to March 2023, the cumulative export of wolfberries reached 3277.42 t, valued at 164.57 million RMB yuan. China has incorporated wolfberry picking and processing into more comprehensive development planning as an important part of rural revitalization and poverty alleviation.

The rapid development of machine vision technology has made object detection one of the most promising areas for application in fruit picking[1]. As a large agricultural production country, China has always been exploring and promoting the development of agricultural architecture. In recent years, with the support of new information technology and robotics, agricultural products' picking precision, accuracy, and efficiency have been greatly improved but still need to face up to their shortcomings. For example, China's wolfberry picking relies on artificial breeding, while with the increasing aging of the population, the labor force is in short supply. However, as the labor demand as well as its cost continue to increase, the competitiveness of the Chinese wolfberry industry in the international market is gradually weakened. Therefore, machine vision technology has been introduced into agricultural picking[2], and the automation of wolfberry picking has become the recognized development trend.

By employing machine vision technology, the precise recognition, accurate positioning, and growth prediction of wolfberry can be achieved, and the efficiency and quality of wolfberry picking will be significantly enhanced. The mainstream direction of visual recognition research in the current market can be

Received date: 2024-05-06 **Accepted date:** 2025-01-02

Biographies: Fengque Pei, Researcher, research interest: the application of new IT in agriculture, Email: fq_pei@hhu.edu.cn; **Zhi Li**, Master candidate, research interest: the application of new IT in agriculture, Email: 221619010043@hhu.edu.cn; **Zhiyu Song**, Associate Professor, research interest: agricultural mechanization engineering, Email: songzhiyu@caas.cn; **Dunbin Tang**, Professor, research interest: mechanical engineering, Email: d.tang@nuaa.edu.cn; **Ru Wan**, Associate Researcher, research interest: agricultural engineering, Email: Wan_ru@163.com.

***Corresponding author: Song Mei**, Associate Professor, research interest: agricultural mechanization engineering for fruits. Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing 210014, China. Tel: +86-15366092940, Email: meisong@caas.cn; **Zhigang Shi**, Professor, research interest: agricultural engineering. Chinese wolfberry science institute, Ningxia Academy of Agriculture and Forestry Sciences, Yinchuan 210037, China. Tel: +86-13099572450, Email: 237518963@qq.com.

divided into two categories: fruit recognition under strong noise and fruit recognition under occlusion and overlap, as explained below:

1) Fruit recognition under strong noise

Strong noise refers to the noise in the image that is obvious and severely affects the observation, processing, or analysis[3]. This type of noise can be generated by a number of factors, including sensor problems in the acquisition equipment itself, environmental conditions, abnormal interference in image acquisition, etc. Commonly, strong noises include light noise, rain, fog, snow, frost noise, vibration noise, etc., which can lead to blurring and distortion of image details, which has a detrimental impact on image analysis. Scholars at home and abroad have done numerous research studies in this field. For the problem of low light, a recognition method of tomato fruit at night based on improved YOLOv5 was proposed by He et al.[4] A total of 2000 tomato images in the night environment were collected as the initial training samples, the original loss function was improved by establishing a CIoU target position loss function based on intersection and union ratio, and the optimal anchor frame was generated based on the computational function anchor. An improved YOLOv5 network model was constructed, which improved the recognition accuracy of feature objects in the dark. Facing the complex and changing environment, Zhao et al.[5] proposed an improved fruit recognition algorithm based on deep learning to address the issue of low performance of traditional fruit recognition algorithms in complex environments. This algorithm incorporated residual modules and Cross Stage Partial Networks (CSP Net), integrated Spatial Pyramid Pooling (SPP) modules into the recognition network of YOLOv5, and replaced the Non-Maximum Suppression (NMS) algorithm with the Soft NMS algorithm. Finally, the algorithm was improved by constructing a joint loss function based on focal and CIoU loss, which led to a higher recognition accuracy. Aiming at the adverse effects of complex environmental factors on the detection and recognition of safflower silk, Wang et al.[6] proposed an improved model based on YOLOv7 to create a real safflower sample dataset to establish a complex picking environment aiming at the problem of inaccurate recognition of safflower mechanized harvesting caused by complex environments such as light, shade, density, and uneven number of samples, which creates a safflower sample dataset to establish the complex environment data of real picking, and increases the attention mechanism of Swin Transformer to improve the accuracy of the model in detecting each classified sample.

In addition, Li et al.[7] addressed the difficulty of picking robots to accurately locate the fruit in the case of fruit oscillation. An oscillating fruit recognition and localization method based on monocular vision and ultrasound detection was proposed and ultimately achieved a recognition and picking success rate exceeding 86%.

It is summarized that the existing strong noise is mostly based on light noise, rain, fog, snow and frost noise, and vibration noise. However, there are few noise reduction studies based on wind and sand. China's wolfberry is mainly distributed in Ningxia Hui Autonomous Region and other places where the windy and sandy weather is very common. Recognition of wolfberries with noise and sand forms the research foundation for fruit positioning, occlusion restoration, growth prediction, and precision picking. However, this specific field has not yet undergone in-depth investigation.

2) Recognition of overlapping fruit

Strong noise is the main concern of fruit picking research, but dealing with the overlapping problem of fruit has also become a major challenge. To address the challenges of missed and false detections in citrus fruit detection caused by environmental factors such as leaf occlusion, fruit overlap, and variations in natural light in hilly and mountainous orchards, Yu et al.[8] proposed a citrus detection model based on an improved YOLOv5 algorithm, which introduced receptive field convolutions with full 3D weights (RFCF). The model overcomes the issue of parameter sharing in convolution operations, and it enhances detection accuracy. Due to the unstructured characteristics of the orchard environment, Li et al.[9] proposed an ensemble U-Net segmentation model suitable for small sample datasets, in which edge structures are designed by integrating residual blocks and gated convolutions to obtain the boundary semantic information of the target image; atrous convolutions are applied to resolve the contradiction between the resolution of the feature map and the receiving field, retain more multi-scale context information, and achieve target fruit segmentation. Yang et al.[10] designed a new attention module, NCBAM, to improve the ability to extract blueberry features in response to the problems of adherence densities and severe occlusion of blueberries. Then, a small target detection layer was added to improve the multi-scale recognition ability of blueberries. Finally, the C3Ghost module was introduced into the backbone network to reduce the number of model parameters, which reduces the complexity of the model to a certain extent. A method of tomato segmentation based on RGB-D depth images and K-means optimized SOM neural network was proposed by Li et al.[11] to solve the problem of automatic recognizing and localizing difficulties caused by fruit overlapping and adherence. By extracting both the planar and depth information of fruit contour points, and employing a *K*-means optimized Self-Organizing mAP (SOM) neural network to construct clustered points that fit the position and contour shape of tomatoes, this approach enhances the accuracy of recognizing occluded fruits. Hao et al.[12] proposed a YOLO-RD-Apple orchard heterogeneous image occlusion fruit detection model based on dual inputs of RGB and Depth images. The lightweight MobileNetV2 and the lighter MobileNetV2-Lite were utilized as feature extractors for RGB and Depth images, and the new SE-DWCSP3 module was proposed to improve the PANet structure and enhance the feature extraction capability of the network for stubby apple targets. Zhao et al.[13] researched the fast-tracking and recognition of overlapping fruits, which fit and predetermine the robot's motion path based on the center of the circle of the captured overlapping images, and finally used fast normalized mutual correlation matching to match and recognize the overlapping fruits, which improved the robot's recognition of the picking timeliness.

The existing research mainly focuses on large fruits such as tomatoes and apples, and there are few studies on the recognition of small fruits like wolfberry. By using the depth information and considering the prior adjacent constraint between the fruit and the stem of cherry tomatoes, Xu et al.[14] proposed an improved Mask R-CNN for visual recognition of cherry tomatoes. The input layer was modified to achieve dual-mode data fusion of RGB and depth images, and the corresponding region generation network was constructed to indicate the integral constraint between the fruit and the stem. Meanwhile, multi-task loss balance and adaptive feature pooling were adopted to overcome the limitation caused by the size difference between fruit and stem. Liu et al.[15] proposed a fruit recognition method for winter jujube based on improved YOLOv3 (YOLOv3-SE), which strengthened effective features, weakened ineffective or invalid features, and improved the expressive ability of the feature map, thus improving the model recognition accuracy. In response to the complexity and variability of grape growth scenarios, Sun et al.[16] proposed a grape detection method based on improved YOLO-v5 for grapes with different occlusion states, which used the lightweight network MobileNetv3 as a feature extraction network and introduced the neck network in the RepVGG Block. Finally, the loss based on dynamic non-monotonic focusing mechanism (wise intersection over union loss, WIoU Loss) was used as the bounding box regression loss function to accelerate network convergence and improve the detection accuracy of the model. Wang et al.[17] improved the YOLOv4-Tiny network structure and proposed a network containing an Attention Module Target Detection Network (I. YOLOv4.Tiny), which uses the CSPDarknet53. A tiny network model as the backbone network and the Convolution Block Attention Module (CBAM) is added to the Feature Pyramid of the YOLOv4. A tiny network structure, whose network has fewer network layers and a low memory footprint, is one way to improve the accuracy of blueberry fruit detection. Zhang et al.[18] leveraged deep learning techniques to develop a coated seed recognition model named YOLO-Coated Seeds Recognition (YOLO-CSR), aiming to address the challenges posed by coated seed recognition tasks in response to the aforementioned challenges. The experimental results showed that YOLO-CSR achieved the best recognition performance on the self-built coated seed image dataset.

Nevertheless, the existing research studies on the recognition of wolfberry are relatively few. In the wolfberry's main regions of distribution, such as Gansu and Ningxia, where the climate is arid and sandy[19], the conventional recognition technology cannot be carried out effectively. Additionally, because its fruit is more dense, it is very easy for the fruit to block each other, which increases the difficulty of recognition. Therefore, research on the recognition algorithm of wolfberry under the influence of strong noise has high scientific significance and engineering urgency.

In summary, the application of target detection algorithms has been gradually popularized in fruit recognition and picking, and the strong adaptability and wide application prospects of machine vision in fruit recognition[20] have been demonstrated. However, influenced by environmental and lighting conditions, equipment vibration, fruit size, and other factors, the traditional fruit detection methods still have great limitations in detecting wolfberries. Under the environment of mature target detection algorithms and recognition techniques, there is still room for research on wolfberry recognition due to the special characteristics of its fruit. Besides, the target detection algorithm is constantly updated and iterative, and the traditional target detection algorithm no longer meets the picking process for the recognition and recognition efficiency needs. Therefore, this study focuses on the above difficulties related to the small size of wolfberry fruit, growing environment with large wind and sand, collecting picture data affected by wind and sand, and optimizing the recognition accuracy of wolfberry as well as the recognition efficiency, all to enhance the reliability of the target detection algorithm for recognition of wolfberry.

2 Improved recognition algorithm architecture for YOLOv8

2.1 YOLOv8 algorithm

YOLOv8 is the model of the YOLO series with high accuracy and speed of detection[21]. According to the depth of the network and the width of the feature map, the YOLOv8 algorithm is divided into five versions: YOLOv8-n, YOLOv8-s, YOLOv8-m, YOLOv8-l, and YOLOv8-x. The network structure of YOLOv8 is composed of four sections: Input, Backbone, Neck, and Head. The input part is the image input link, and the Backbone part references the CSPDarkNet-53 network, which consists of the convolution module Conv, C2f structure, and SPPF module[22]. Compared with YOLOv5's C3 module, YOLOv8's C2f module adds more layer-hopping connections, and the amount of computation is significantly reduced, which effectively improves the convergence speed and the convergence effect. The Focus module mainly operates for image slicing, which is capable of further extracting the target features. The PANet structure is adopted by the Neck part, which deletes the convolution operation at the sampling stage on YOLOv5, and effectively integrates and utilizes features of different scales to capture the various scales and features of the detected target more accurately. The Head part is responsible for the computation of the enhanced target features and finally getting the confidence level and position of the target.

2.2 Improvement of the backbone

To achieve faster networks, commonly used operators were revisited by Chen et al.[23], and it was demonstrated that low FLOPS are mainly caused by frequent memory access, particularly in depthwise convolutions. Consequently, a novel Partial Convolution (PConv) was proposed, enabling the more efficient extraction of spatial features by simultaneously reducing redundant computation and memory access. Building on PConv, FasterNet was further developed, achieving significantly faster performance across various devices while maintaining accuracy in a wide range of visual tasks.

Specifically, the authors proposed a simple Partial Convolution (PConv) to simultaneously reduce computational redundancy and memory access, and the working principle of PConv is illustrated in Figure 1. In light of the novel PConv and the readily available PWConv as the primary operator, FasterNet was further proposed. It is structured into four hierarchical stages as shown in Figure 2, with each stage composed of a set of FasterNet blocks, preceded by an embedding layer or a merging layer. The final three layers are used for feature classification. Within each FasterNet block, a PConv layer is followed by two PwConv layers. To maintain feature diversity and achieve lower latency, normalization and activation layers are set only after the intermediate layer.

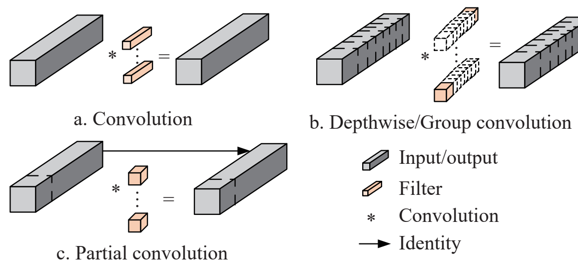


Figure 1 Architecture of the PConv used in this study

2.3 Improvement of FPN

A fusion mechanism, RepBiFPN (BiFPN + RepViT), was proposed in this study.

2.3.1 RepViT

Although lightweight Vision Transformers (ViTs)[24] have demonstrated excellent performance due to their ability to learn global representations, the architectural differences between lightweight CNNs and lightweight ViTs have not been fully explored. Therefore, the efficient architectural design of lightweight ViTs was integrated, and standard lightweight CNNs (particularly MobileNetV3) were gradually improved to create a series of novel pure CNN models, referred to as RepViT[25]. These models have been shown to perform well across various visual tasks and are more efficient than existing lightweight ViTs. Figure 3 shows the overview of RepViT.

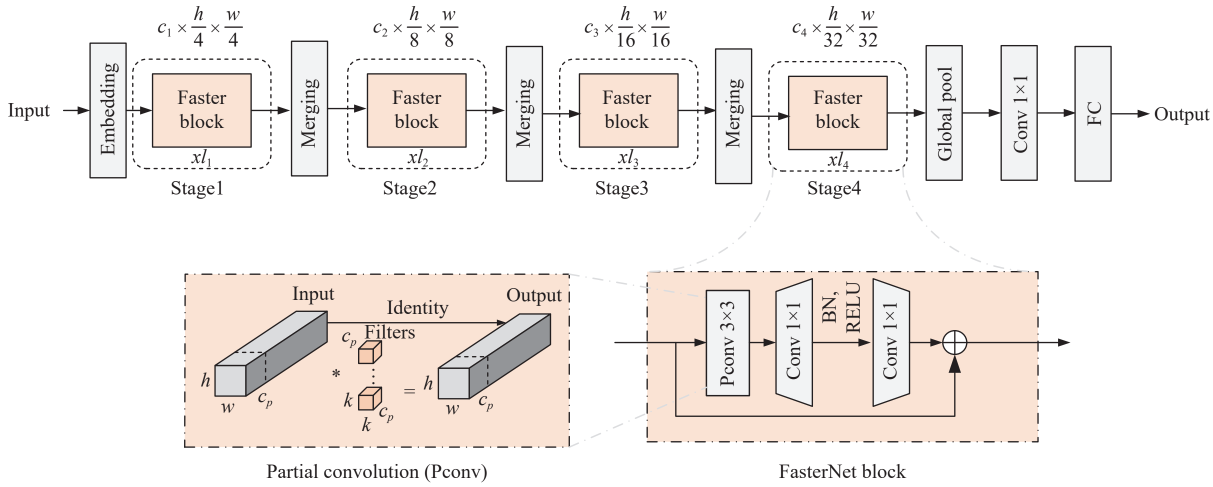


Figure 2 Overview of FasterNet

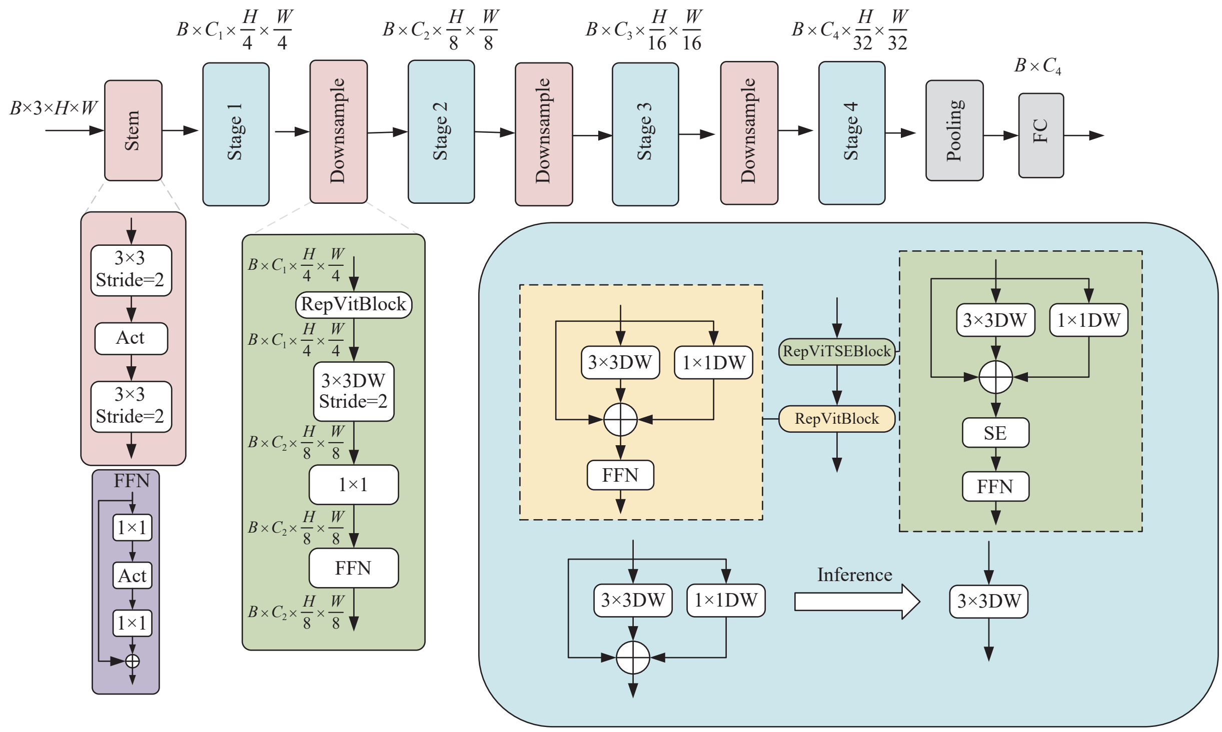


Figure 3 Overview of RepViT

2.3.2 BiFPN

BiFPN (Bidirectional Feature Pyramid Network), an efficient multi-scale feature fusion network, is optimized based on the traditional Feature Pyramid Network (FPN).

Figure 4 illustrates the specific details of the EfficientNet architecture, including the use of EfficientNet as the backbone and BiFPN as the feature network. In this architecture, BiFPN layers are employed to receive multi-scale input features from the EfficientNet backbone through their bidirectional feature fusion capability and then generate expressive features for object classification and bounding box prediction.

2.4 Improvement of head

To further improve the performance of YOLO, an innovative Lightweight Asymmetric Dual-Head (LADH, as in Figure 5)[26] detection head was proposed. The detection head utilizes asymmetric multi-stage compression technology for enhancing detection efficiency and accuracy. Traditional YOLO detectors typically use symmetric multi-stage compression structures, where the compression ratios of all detection heads are the same. However, since the complexity of objects in different categories varies, an asymmetric multi-stage compression strategy was introduced. Different compression ratios are applied to different categories to better adapt to various feature representations and target size distributions, thereby improving the generalization capability of the detector.

The LADH detection head mainly consists of two parts: AsymmetricHead and Dual-Head. AsymmetricHead is responsible for asymmetric compression of features from different categories to handle the differences in category complexity, while Dual-Head combines the outputs of two AsymmetricHeads to generate the final detection results. By incorporating these two components, LADH enhances detection performance and accuracy while maintaining a lightweight model.

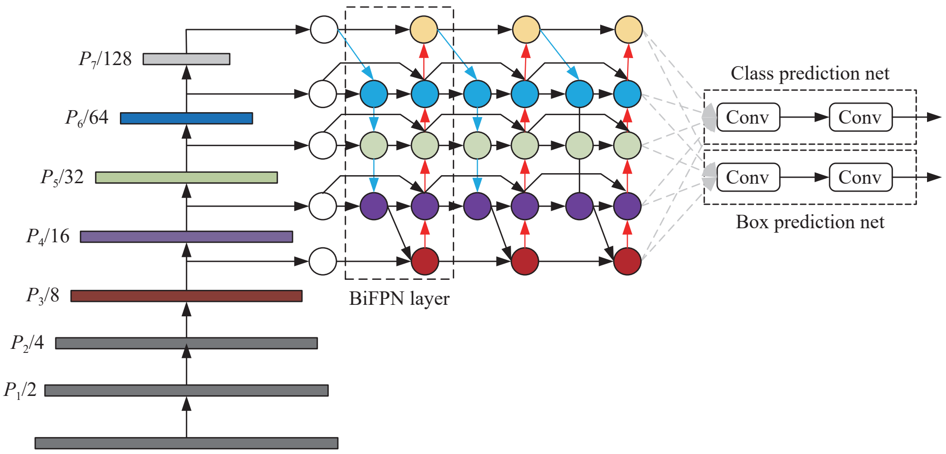


Figure 4 Overview of EfficientNet

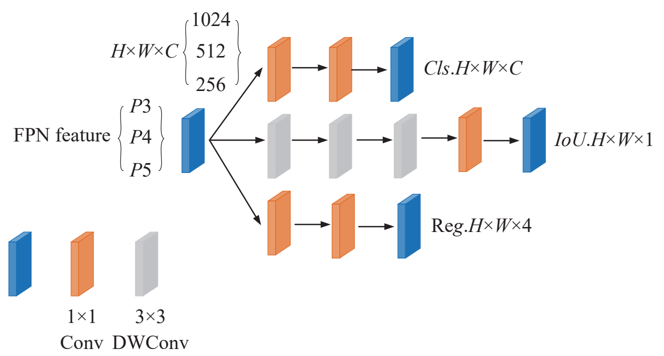


Figure 5 Architecture of LADH

In order to validate the effectiveness of the improved method, the following metrics were used to evaluate the model: precision (P), recall (R), mean accuracy (mAP), and the rate of model prediction (fps). The calculations are as follows, and the improved YOLOv8 architecture is proposed in Figure 6.

$$P = \frac{TP}{TP+FP} \tag{1}$$

$$R = \frac{TP}{TP+FN} \tag{2}$$

$$AP = \sum_{i=1}^{n=1} (r_{i+1} - r_i) p(r_{i+1}) \tag{3}$$

$$\mathrm{mAP} = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{4}$$

$$\mathrm{fps} = \frac{1000}{preprocess(ms) + \inf erence(ms) + postprocess(ms)} \tag{5}$$

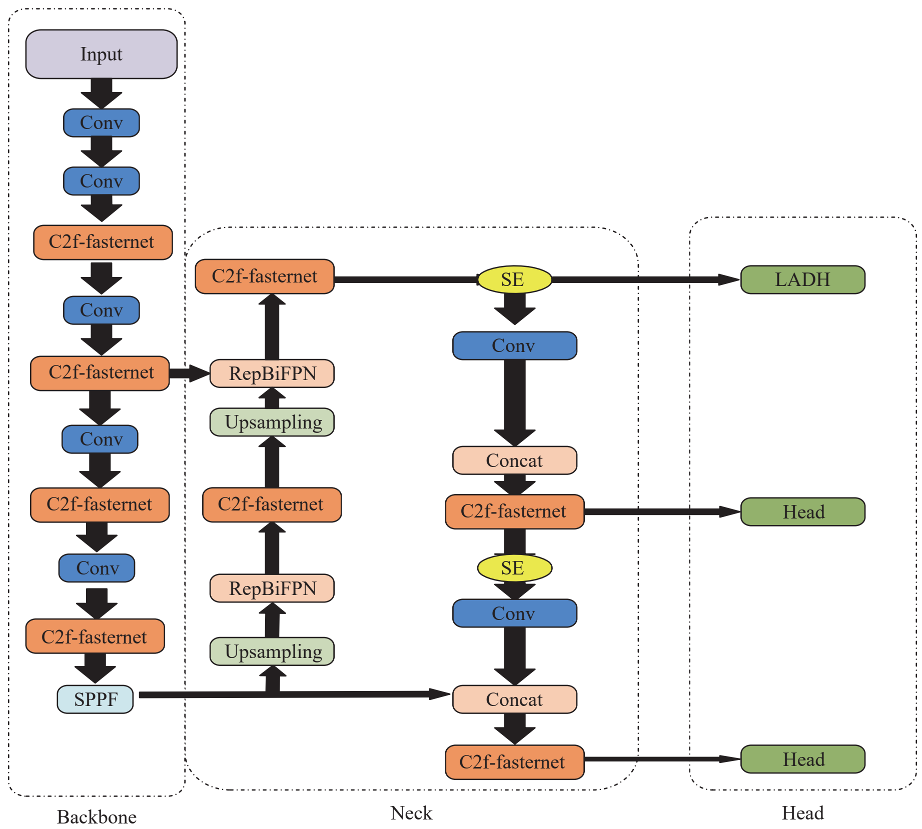


Figure 6 Improved YOLOv8 architecture

3 Data sample collection and preprocessing

3.1 Data acquisition

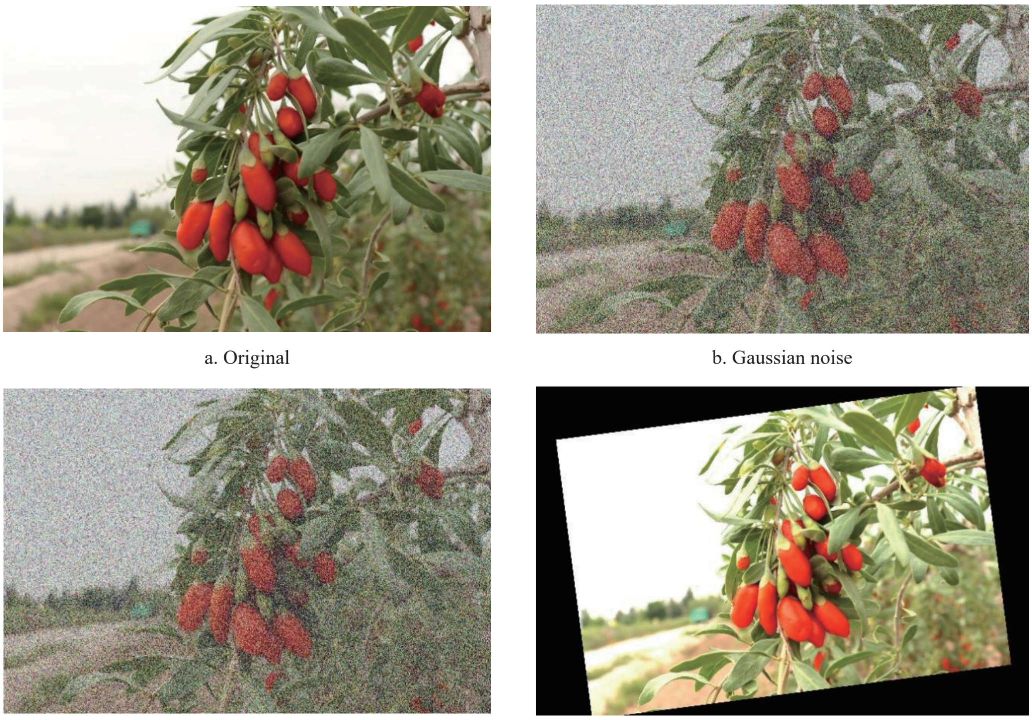
The wolfberry fruits used in this experiment were collected from the Ningxia Academy of Agriculture and Forestry Science. A total of 990 images of wolfberry were collected, which were captured by a DJI hand-held camera at a distance of 50 cm from the wolfberry fruits. The original format of the photographed images was JPG, with a resolution of 2976×1984. The wolfberry samples are shown in Figure 7, and a collection of images of wolfberry are available at: https://pan.baidu.com/s/1O4dNaoGWiL2lq_TLR5VAtQ?pwd=hck1.

Figure 7 Examples of wolfberry bush image

3.2 Preprocessing and image enhancement of sample datasets

In the collected LBP images, due to different shooting angles and different light conditions under different weather, the color of the collected images differed to some extent, which led to the difficulty of extracting the features of wolfberry. In order to obtain accurate data parameters, the screened and classified wolfberry images were manually labeled with LabelImg. To ensure the effectiveness of the training, reduce the impact of noise during shooting, and avoid overfitting of deep learning models during training, the Opencv was employed for data enhancement on the 990 images collected for training and testing, with which the samples were expanded to 3840 images. The samples of images with noises added are shown in Figure 8.

After expanding the images, labeling software was used to annotate the targets of wolfberries in the images to obtain the XML files in VOC format. During the labeling process, the size of wolfberries in the obscured situation was in accordance with the real size. The labeled wolfberries' dataset was divided into a training set, a validation set, and a test set, which were randomly assigned according to the ratio of 8:1:1 of the total dataset.



a. Original b. Gaussian noise

c. Sault noise d. Turning around

Figure 8 Image of wolfberries after data enhancement process

4 Experimental validation of YOLOv8 based on improvement of different parts

4.1 Experimental environment and parameters

Model training and testing were performed on the same computer with a hardware configuration of 12thGen Intel(R) Core (TM) i7-12700H2.30 GHz, GeForceGTX3060GPU, 16 GB of operating memory, a software environment of ubuntu18.04 operating system, and a model training framework of Pytorch1.12. The relevant training parameters are listed in Table 1.

Table 1 Model training parameters

Parameter	Value	Parameter	Value
LearningRate	0.01	Epochs	220
BatchSize	16	Momentum	0.937
ImageSize	640×640	WeightDecay	0.0005

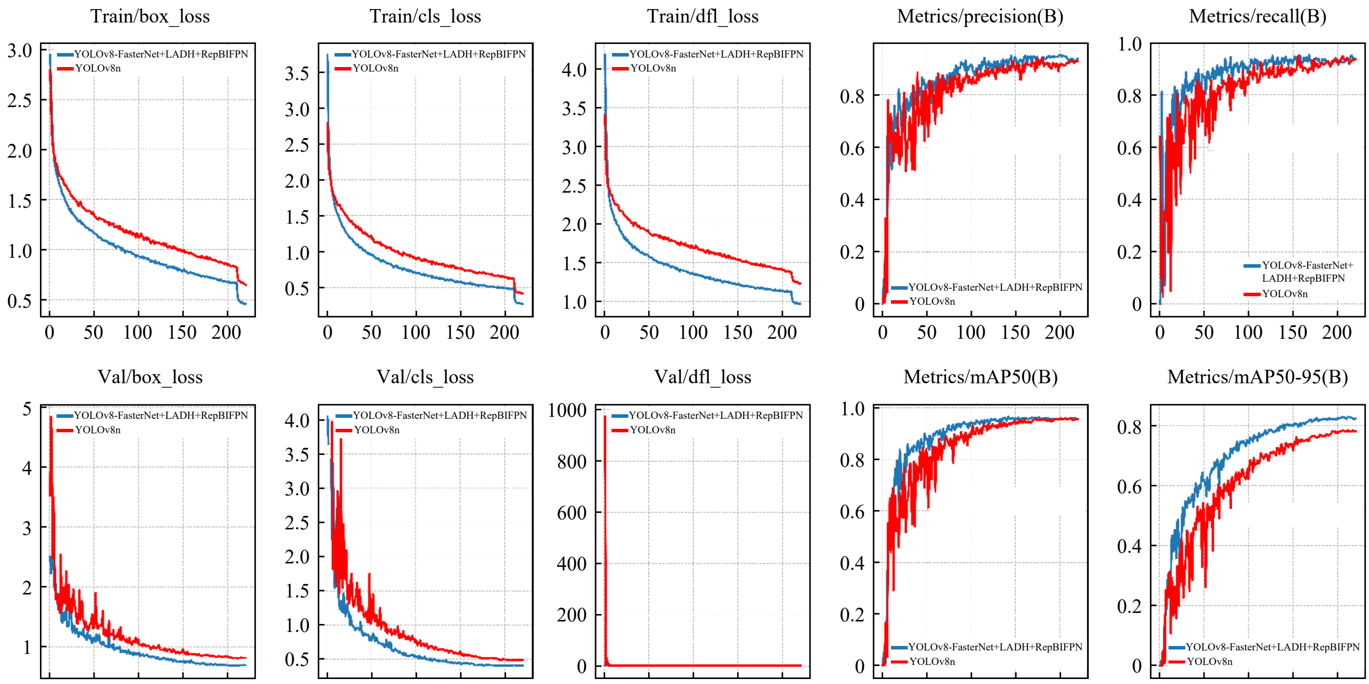
4.2 Ablation experiment

In this study, three different neural network backbones—EfficientNet, ReNLANFasterNet, and FasterNet—along with three different detection heads—LightshiftHead, LAHD, and ImplicitHead—and three different feature pyramid networks—AFPN, EffQAFPN, and RepBiFPN—were selected and integrated into the YOLOv8 network for training. The model's various performance metrics were compared. The advantages and disadvantages of each improvement module are listed in Table 2.

From the data in Table 3 and Figure 9, it can be summarized that the mean average precision under the IoU of 0.5-0.95 (mAP@ 0.5-0.95) of YOLOv8-FasterNet+LADH+RepBiFPN is the highest, at about 0.81. Although the speed of YOLOv8-FasterNet+LADH+RepBiFPN slightly decreases compared to YOLOv8n, its performance of other metrics is better than the original model. Table 3 lists some more information about the ablation experiment.

Table 2 Comparison of different YOLO models

Model	Advantages	Disadvantages
EfficientNet	Efficient parameter usage, excellent performance, strong scalability, and good generalization ability	Longer training time, high hardware resource requirements, complex architecture, limited inference speed
FasterNet[27]	Speed improvement, high efficiency, easy deployment, model compression	Increased complexity, trade-off in accuracy, limited applicability, more difficult optimization
FocalNet[28]	Saliency focus, efficient computational usage, strong adaptability	Strong data dependency, difficulty in handling complex backgrounds, complex debugging
LightshiftHead	High computational efficiency, resource-saving, strong adaptability, maintains performance	Limited applicability, complex model tuning, immature model
LADH	High efficiency, strong flexibility, resource-saving	Increased complexity, limited applicability, potential risk of overfitting
ImplicitHead	Simplified model, flexible adaptation to input features	Insufficient accuracy, weak domain dependency
AFPN[29]	Enhanced multi-scale processing capability, optimized feature fusion	Increased computational complexity, complex model tuning
EffQAFPN[30]	High accuracy, high efficiency, flexible modular design	Sensitive to noise, relatively high model complexity
RepBiFPN	Efficient feature reparameterization, enhanced multi-scale detection, bidirectional feature fusion	High data dependency, large computational resource requirements



a. Training curves of YOLOv8n and YOLOv8-Fasternt+LADH+RepBiFPN

b. Prediction of YOLOv8n

c. Prediction of YOLOv8-FasterNet+LADH+RepBiFPN

Figure 9 Comparison of YOLOv8n and YOLOv8-FasterNet+LADH+RepBiFPN

Table 3 Results of the ablation experiment

Model	*P*	*R*	mAP@0.5	mAP@0.5-0.95	Frames per second/fps	Para
YOLOv8n (original)	0.898	0.946	0.963	0.765	109	3 005 843
FasterNet	0.929	0.911	0.969	0.765	90	2 560 467
FocalNet	0.912	0.915	0.963	0.76	50	3 089 523
EfficientNet	0.933	0.889	0.96	0.755	91	3 086 275
FasterNet+LADH	0.927	0.941	0.973	0.778	85	2 343 779
FasterNet+LightshiftHead	0.926	0.908	0.962	0.764	96	2 601 059
FasterNet+ImplicitHead	0.905	0.948	0.967	0.776	83	2 557 990
FasterNet+LADH+RepBiFPN	0.921	0.944	0.963	0.81	67	2 276 867
FasterNet+LADH+AFPN	0.865	0.889	0.936	0.703	67	1 628 456
FasterNet+LADH+EffQAFPN	0.868	0.922	0.954	0.738	73	2 362 483

Figure 9a presents the comparison results of 10 indicators of the ablation experiments. The horizontal axis for all 10 indicators of results represents the number of epochs (the figure displays results from epoch 0 to 250, with convergence achieved before 250 epochs). In the figure, blue represents the results of the proposed YOLOv8-FasterNet+LADH+RepBiFPN in this study, while red represents YOLOv8n. The 10 indicators are:

(1) train/box_loss, representing the bounding box regression loss on the training set, typically referring to the error in bounding box predictions.

(2) train/cls_loss, representing the classification loss on the training set, indicating the prediction error for each category.

(3) train/dfl_loss, representing the distributed focal loss on the training set, a specific loss function related to object detection.

(4) metrics/precision(B), representing the precision on the validation set, measuring the accuracy of positive class predictions.

(5) metrics/recall(B), representing the recall on the validation set, measuring the completeness of positive class samples retrieved by the model.

(6) val/box_loss, representing the bounding box regression loss on the validation set.

(7) val/cls_loss, representing the classification loss on the validation set.

(8) val/dfl_loss, representing the distributed focal loss on the validation set.

(9) metrics/mAP50(B), representing the mean average precision (mAP) at an IoU (Intersection over Union) threshold of 50%.

(10) metrics/mAP50-95(B), representing the mean average precision at different IoU thresholds (from 50% to 95%).

For Indicators (1)-(3) and Indicators (6)-(8), which are all loss values, the smaller, the better. The results show that, whether in the training process or the testing process, the YOLOv8-FasterNet+LADH+RepBiFPN demonstrates comparable or superior loss performance compared to the YOLOv8n.

From Indicators (4) and (5) and Indicators (9) and (10), it is evident that YOLOv8-FasterNet+LADH+RepBiFPN exhibits better average recognition accuracy compared to YOLOv8n, as shown in Figures 9b and 9c.

4.3 Comparative experiment

This study aims to evaluate the performance of YOLOv5, SSD, YOLOv7, YOLOv8, and YOLOv8-FasterNet+LightshiftHead+RepBIFPN models in the task of detecting defects in wolfberries. To ensure a robust comparison, the same dataset was employed as the baseline, and all parameters were standardized prior to training. Specifically, it was ensured that the parameters across all models remained consistent. A controlled and fair testing environment was established to facilitate a more accurate evaluation of the models' performance under identical conditions. The mean average precision (mAP) scores of the five models obtained after testing are listed in Table 4.

Table 4 Comparison of accuracy across different network models

Model	mAP@0.50	mAP@0.50-0.95
YOLOv5s	0.875	0.641
YOLOv7-tiny	0.894	0.697
YOLOv8n	0.963	0.765
YOLOv8-FasterNet+LADH+RepBiFPN	0.963	0.810
SSD	0.832	0.608

As listed in Table 4, the YOLO algorithm outperforms SSD. Furthermore, among the YOLO series models, YOLOv8 demonstrates superior detection accuracy on this dataset compared to YOLOv5 and YOLOv7. Notably, the improved YOLOv8-FasterNet+LightshiftHead+RepBiFPN achieves significantly higher mAP@0.5-0.95 accuracy than the original YOLOv8 model. These comparative experiments further validate the effectiveness of the improved YOLOv8 neural network in enhancing the detection and recognition performance of Chinese wolfberries.

5 Conclusions

Based on the YOLOv8 model, this study proposed an improved YOLOv8 model for recognition of wolfberry, named YOLOv8-FasterNet+LADH+RepBiFPN, which weakened the impact of noise while collecting images of wolfberry and realized the accurate detection of wolfberries influenced by sand and wind, providing some theoretical support for the realization of automated wolfberry picking.

Based on the self-constructed wolfberry dataset, several random noises were innovatively added to the original images to imitate the noise of sandy and windy weather, which enhanced the robustness and generality of the model while training.

This study introduced several improvements to the YOLOv8 model to enhance its accuracy in wolfberry recognition. Specifically, FasterBlock was incorporated into the C2f layer of the backbone, and the original detection head was replaced with the novel LADH. Additionally, the YOLOv8 Feature Pyramid Network (FPN) was enhanced by integrating RepViT (Revisiting Mobile CNN from ViT Perspective) and BiFPN (Bidirectional Feature Pyramid Network). Ablation experiments were conducted to compare the effects of improvements made to the backbone, neck, and detection head. Furthermore, the improved model was benchmarked against various network models, including YOLOv5, SSD, and YOLOv7-tiny, through horizontal comparisons, further demonstrating the effectiveness of the proposed enhancements.

As a kind of small fruit, wolfberry grows densely and shows the phenomenon of fruit overlapping, so it is proposed to carry out the next research in the fields of fruit localization, shade repair, etc. to form a complete technological closed loop with growth prediction, accurate picking, etc. This will provide the necessary support for the development of intelligent technology in the wolfberry industry.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 32201681); the Ningxia Hui Autonomous Region Science and Technology Program (Grant 2021BEF02001); the Fruit, Vegetable, and Tea Harvesting Machinery Innovation Project of the Chinese Academy of Agricultural Sciences; the Jiangsu Agriculture Science and Technology Innovation Fund (Grant JASTIF, CX(23)3036); and the Changzhou Sci&Tech Program (Grant CJ20241131).

[References]

[1] Yang Y Y, Han Y X, Li S, Yang Y D, Zhang M, Li H. Vision based fruit recognition and positioning technology for harvesting robots. Computers and Electronics in Agriculture, 2023; 213: 108258.

[2] Zhu X, Chen R W, Xia H K, Zhang P Y. Intelligent recognition system of fruit picking robot. Application Research of Computer, 2014; 31(9): 2711–2714. (in Chinese) doi: 10.3969/j.issn.10013695. 2014.09.035.

[3] Shen W, Pang Q, Fan Y L. Study on strong noise image restoration based on adaptive stochastic resonance in bistable system. Computer Engineering and Applications, 2009; 45(15): 180–182. (in Chinese)

[4] He B, Zhang Y B, Gong J L, Fu G, Zhao Y Q, Wu R D. Fast recognition of tomato fruit in greenhouse at night based on improved YOLO v5. Transactions of the CSAM, 2022; 53(5): 201–208. (in Chinese)

[5] Zhao Z Q, Wang J, Zhao H. Research on apple recognition algorithm in complex orchard environment based on deep learning. Sensors, 2023; 23(12): 5425.

[6] Wang X R, Xu Y, Zhou J P, Chen J R. Safflower picking recognition in complex environments based on an improved YOLOv7. Transactions of the CSAE, 2023; 39(6): 169–176. (in Chinese)

[7] Li G L, Ji C Y, Gu B X. Recognition and location of oscillating fruit based on monocular vision and ultrasonic testing. Transactions of the CSAM, 2015; 46(11): 1–8. (in Chinese)

[8] Yu Y, Liu Y C, Li Y J, Xu C S, Li Y W. Object detection algorithm for citrus fruits based on improved YOLOv5 Model. Agriculture, 2024; 14(10): 1798.

[9] Li Q W, Jia W K, Sun M L, Hou S J, Zheng Y J. A novel green apple segmentation algorithm based on ensemble U-Net under complex orchard environment. Computers and Electronics in Agriculture, 2021; 180: 105900.

[10] Yang W J, Ma X X, Hu W C, Tang P J. Lightweight blueberry fruit recognition based on multi-scale and attention fusion NCBAM. Agronomy, 2022; 12(10): 2354.

[11] Li H, Tao H X, Cui L H, Liu D W, Sun J T, Zhang M. Recognition and localization method of tomato based on SOM-*K*-means algorithm. Transactions of the CSAM, 2021; 52(1): 23–29. (in Chinese)

[12] Hao P F, Liu L Q, Gu R Y. YOLO-RD-Apple orchard heterogeneous image obscured fruit detection model. Journal of Graphics, 2023; 44(3): 456–464. (in Chinese)

[13] Zhao D A, Shen T, Chen Y, Jia W K. Fast tracking and recognition of overlapping fruit for apple harvesting robot. Transactions of the CSAE, 2015; 31(2): 22–28.

[14] Xu P H, Fang N, Liu N, Lin F S, Yang S Q, Ning J. Visual recognition of cherry tomatoes in plant factory based on improved deep instance segmentation. Computers and Electronics in Agriculture, 2022; 197: 106991.

[15] Liu T Z, Teng G F, Yuan Y C, Liu B, Liu Z G. Winter jujube fruit recognition method based on improved YOLO v3 under natural scene. Transactions of the CSAM, 2021; 52(5): 17–25. (in Chinese)

[16] Sun J, Wu Z Q, Jia Y L, Gong D J, Wu X H, Shen J F. Detecting grape in an orchard using improved YOLOv5s. Transactions of the CSAE, 2023; 39(18): 192–200. (in Chinese)

[17] Wang L S, Qin M X, Lei J Y, Wang X F, Tan K Z. Blueberry maturity recognition method based on improved YOLOv4-Tiny. Transactions of the CSAE, 2021; 37(18): 170–178. (in Chinese)

[18] Zhang X W, Xuan C Z, Hou Z F. Recognition model for coated red clover seeds using YOLOv5s optimized with an attention module. Int J Agric & Biol Eng, 2023; 16(6): 207–214.

[19] Xu R, Yang J L, Liu J, Yan W X, Ma G F, Ma J B. Trends and effects of agro-climatic resources and main meteorological disasters during *Lycium barbarum* L. growing seasons in Ningxia. Chinese Journal of Eco-Agriculture, 2023; 31(10): 1645–1656. (in Chinese)

[20] Feng J, Zeng L H, Liu G, Si Y S. Fruit recognition algorithm based on multi-source images fusion. Transactions of the CSAM, 2014; 45(2): 73–80. (in Chinese)

[21] Li P, Zheng J S, Li P Y, Long H W, Li M, Gao L H. Tomato maturity detection and counting model based on MHSA-YOLOv8. Sensors, 2023; 23(15): 6701.

[22] Wang X T, Lu R T, Bi H X, Li Y H. An infrared small target detection method based on attention mechanism. Sensors, 2023; 23(20): 8608.

[23] Chen J R, Kao S-H, He H, Zhuo W P, Wen S, Lee C-H. Run, don't walk: Chasing higher FLOPS for Faster Neural Networks. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, Canada: IEEE, 2023; pp.12021–12031. doi: 10.1109/CVPR52729.2023.01157.

[24] Li N N, Chen Y R, Zhao D B. Adaptive search for broad attention based vision transformers. Neurocomputing, 2025; 611: 128696.

[25] Wang A, Chen H, Lin Z J, Han J G, Ding G G. RepViT: Revisiting mobile CNN from ViT perspective. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle: IEEE, 2024; pp.15909–15920. doi: 10.1109/CVPR52733.2024.01506.

[26] Zhang S F, Chi C, Yao Y Q, Lei Z, Li S Z. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle: IEEE, 2020; pp.9756–9765. doi: 10.1109/CVPR42600.2020.00978.

[27] Wen M, Li C Y, Xue Y S, Xu M, Xi Z H, Qiu W D. YOFIR: High precise infrared object detection algorithm based on YOLO and FasterNet. Infrared Physics and Technology, 2025; 144: 105627.

[28] Gulsoy T, Kablan B E. FocalNeXt: A ConvNeXt augmented FocalNet architecture for lung cancer classification from CT-scan images. Expert Systems With Applications, 2025; 261: 125553.

[29] Yang G Y, Lei J, Zhu Z K, Cheng S Y, Feng Z L, Liang R H. AFPN: Asymptotic Feature Pyramid Network for object detection. 2023 IEEE International Conference on Systems, Man and Cybernetics (SMC), Honolulu: IEEE, 2023; pp.2184–2189. doi: 10.1109/SMC53992.2023.10394415.

[30] Wang K H, Chu X X, Xu X M, Huang J S, Wei X M. EfficientRep: An efficient Repvgg-style ConvNets with hardware-aware neural network design. arXiv, 2023; In press. arXiv: 2302.00386. doi: 10.48550/arXiv.2302.00386.