YOLOv10-RLF:A lightweight small target grape leaf disease detection method based on improved YOLOv10

Guangda Chen
Beihua University
Shiyu Jin
Beihua University
Yanfei Jia
jia_yanfei@163.com
Beihua University

Research Article

Keywords: grapevine leaf disease, lightweight, YOLOv10, small target, disease identification

Posted Date: May 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9350270/v1

Additional Declarations: No competing interests reported.

YOLOv10-RLF:A lightweight small target grape leaf disease detection method based on improved YOLOv10

Guangda Chen；Shiyu Jin；Yanfei Jia*

College of Electrical and Information Engineering，Beihua University, Jilin 132021, China.

Corresponding Author Email:jia_yanfei@163.com

Abstract

Grapevine crops often suffer from various diseases, posing significant challenges to agricultural production. The diverse morphology and dense distribution of grape leaf diseases complicate identification efforts. As demand for intelligent devices grows, mobile platforms require more efficient neural network algorithms. To address these challenges, we propose a lightweight grape leaf disease detection model, YOLOv10-RLF, based on the YOLOv10n architecture. We introduce the C2f-RVB module, which separates token and channel mixing to enhance feature extraction and reduce redundancy, achieving a 21.74% reduction in parameters and a 19.54% reduction in Giga Floating-point Operations Per Second(GFLOPs). Additionally, we propose a lightweight detection head, v10_LSCD, utilizing shared convolution to further decrease the model's parameters and operational load. To improve detection accuracy, especially for small targets, we replace the SPPF module with a Feature Pyramid Shared Conv (FPSC) module, enhancing multi-scale feature fusion. We also optimize the loss function using SIOU to boost convergence speed and computational accuracy. Experimental results demonstrate that our model reduces parameters and operations by 26.2% and 30.4%, respectively, while compressing model size by 36.4%, all without sacrificing detection accuracy. This research provides a technical foundation for mobile detection of grapevine diseases and supports precise variable applications in agriculture.

Keywords: *grapevine leaf disease; lightweight; YOLOv10; small target; disease identification*

1. INTRODUCTION

Grapes are widely cultivated around the world, and their rich nutritional value, as well as their important medicinal and economic value, have given them an important place in the agricultural industry. The health of the grape crop is directly related to the efficiency of agricultural production and the development of related industries. However, grape crops this aspect of are often subjected to various diseases, which leads to severe yield losses and poses a great challenge to agricultural production. Accurate and timely identification of diseases can not only effectively reduce the yield loss caused by diseases, but also play a positive role in improving the current situation of pesticide abuse.At present, grape disease identification mainly relies on manual observation and empirical judgement, which not only occupies manpower and is inefficient, but also has high requirements for observers. Therefore, it is of great practical significance to carry out the research[1-2] on intelligent identification of grapevine diseases.

With the rapid development of artificial intelligence technology, there are many researchers trying to apply techniques such as machine vision and image processing to crop disease recognition[3]. For example, Zhang Chuanlei et al.[4] extracted segmentation of disease images by using a combination of Genetic Algorithm (GA) and Correlation-based Feature Selection (CFS) methods to extract colour, texture and shape, and, finally, a Support Vector Machine (SVM) classifier for disease recognition.REHMAN Z et al.[5] developed an entropy and multi-class support vector machine based selection method (EaMSVM) to select disease features. However, this traditional machine vision and image processing method will extract features from combinations of colour, shape, texture, etc., and then use classifiers on the extracted features to classify the method is inefficient and difficult to meet the needs of actual agricultural production.

With the progress and development of deep learning technology, it is able to reduce human intervention by learning feature vectors autonomously, overcoming the limitations of traditional machine learning methods in feature extraction, model generalization, etc., and has been widely used in the field of crop picking and disease identification.[6] For example, Hu Yifan et al.[7] improved the target recognition ability by replacing the YOLOv5 backbone network, increasing the attention mechanism, and modifying the feature fusion layer to achieve a recognition accuracy of 90.1% for tomato. Yue Kai et al.[8] proposed a citrus recognition model YOLOv8-MEIN based on improved YOLOv8n, designed the ME convolution module and used it to improve the C2f module of YOLOv8n, which improved the average accuracy mean value by 0.4 percentage points compared to the initial model. Although the above studies have high detection accuracy, they are aimed at large fruit recognition, and grape leaf diseases are characterized by diverse morphology, dense distribution and more small target diseases, which are not suitable for grape leaf disease recognition.

As the demand for intelligent devices expands, more and more mobile devices are applied in the field of disease detection, which also possesses higher requirements for neural network algorithms, especially the model size, number of parameters, and computational volume of the network and other indicators[9]. In terms of disease recognition, Zhang Huili et al.[10] proposed a grape leaf disease recognition model GSGF based on the YOLOv8 model, but the memory occupation is large, reaching 49.3M, which is not conducive to the deployment of mobile devices. Zheng Yuda et al.[11] used the ConvNeXtV2 structure to replace the backbone network of the YOLO v5 model and added dynamic detection heads to design a citrus disease detection model with a floating-point operation of 16.4G, which is extremely unfriendly for mobile deployment. Ma Chaowei et al.[12] proposed a lightweight wheat disease detection method based on improved YOLOv8, and the experimental results show that the improved lightweight model has a parameter count of 2.6M, a model size of 5.5M, and a floating-point capacity of 7.0G, which is in the still a large space of decrease parameter count and floating-point capacity.

For the problem of grape leaf disease morphology multi-use, dense distribution and small target difficult to detect, in order to facilitate the deployment of embedded mobile devices and reduce the occupation of storage space, this paper proposes a lightweight improvement algorithm based on YOLOv10, to reduce the complexity of the algorithm and the amount of computation under the guarantee of high accuracy.

2. Based on improved YOLOv10 network modeling

2.1 YOLOv10 network architecture

As a cutting-edge deep learning algorithm in the field of target detection, YOLOv10 not only inherits the excellent performance of the YOLO series, but also realizes significant breakthroughs in network structure and technology, greatly improving detection accuracy and speed. It takes convolutional neural network as the core, and with the help of multi-branch fusion and dense connection mechanism, it realizes the synergistic capture of target details and global features, and effectively solves the problem of detecting targets of different sizes. Meanwhile, it adopts a consistent dual allocation strategy and introduces an overall efficiency-accuracy-driven model design strategy, which greatly reduces the computational overhead.

The main structure of the YOLOv10 algorithm network model consists of three parts[13]: the backbone feature extraction network (Backbone), the feature enhancement extraction network (Neck), and the detection head (Head). In the network structure, the main modules include Conv, SCDown, C2f, C2fCIB, SPPF, PSA, Concat, Upsample and v10Detect modules. SCDown is a spatial and channel decoupling of the standard convolution to maximise information retention while reducing computational cost.The C2f module is the main module for learning residual features, which is connected across layers by more branches to give the model a richer gradient flow, forming a neural network module with stronger feature representation capability. In the structure of C2fCIB, the Bottleneck in C2f is replaced by CIB (compact inverted block), and the standard convolution in Bottleneck is replaced by deep convolution and point-by-point convolution in CIB[14].PSA is designed to address the problem of high overhead of self-attention[15]. The network actually contains two Heads (with the same structure and different parameters), corresponding to the One-to-many Head and the One-to-one Head, which are involved simultaneously during training, while only the One-to-one Head is needed for inference prediction, and the classification head is lightened from two 3×3 convolutions of YOLOv8 to two depth-separable convolutions of The classification head is lightened from two 3×3 convolutions in YOLOv8 to two deeply separable convolutions[16].

2.2 YOLOv10-RLF modeling

Aiming at the problems such as large number of model parameters and large weight files in the current grape leaf disease detection, this study proposes a lightweight grape leaf disease detection method based on improved YOLOv10, and the improved network structure YOLOv10-RLF is shown in Fig 1. Firstly, the RepViT Block module is used to replace the Bottleneck module in the C2f module, and the C2f-RVB module is constructed to replace the C2f module to realize more efficient extraction of image features and reduce redundancy. Secondly, the original detection header is replaced with LSCD detection header, which fully utilizes the advantages of Group Norm and shared convolution, and can while maintaining the effective fusion of feature information further reduce the computational amount and arithmetic complexity. Finally, the self-developed is used FPSC feature pyramid shared convolution to replace the SPPF pooling layer, which is able to capture more fine-grained features by using convolutional layers with different expansion rates that are enough to extract features at different scales. In addition, SIOU is used to optimize the loss function to improve the prediction frame regression speed and accelerate the convergence of the model, thus improving the detection accuracy.

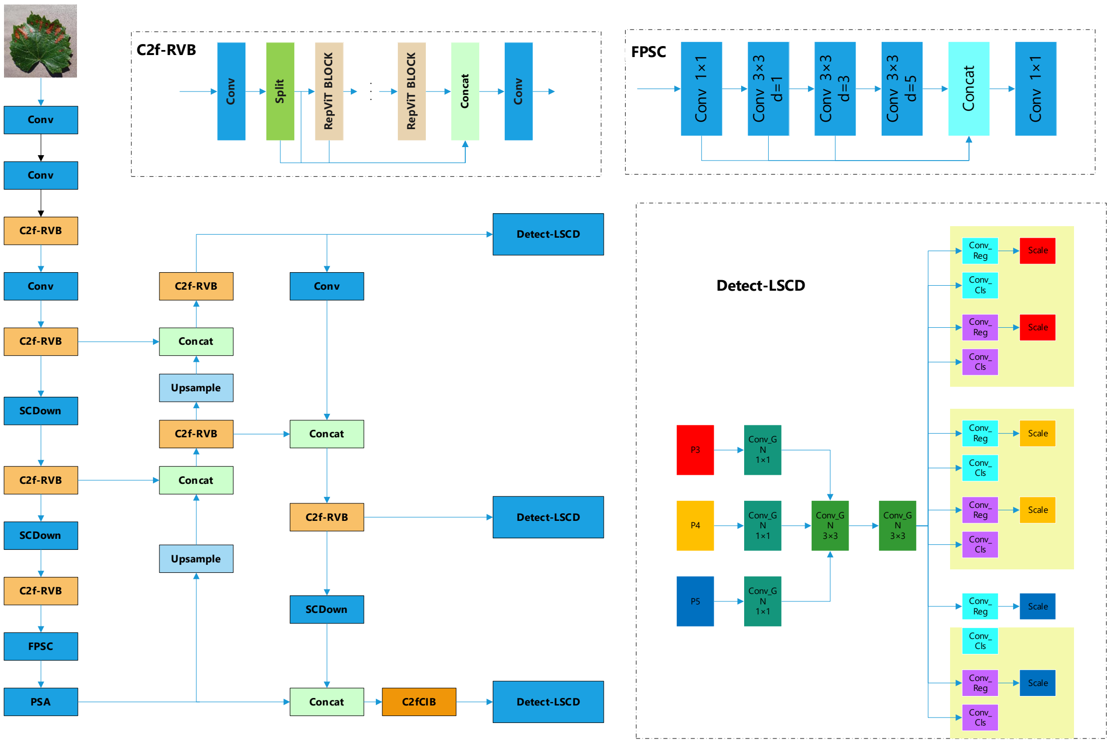
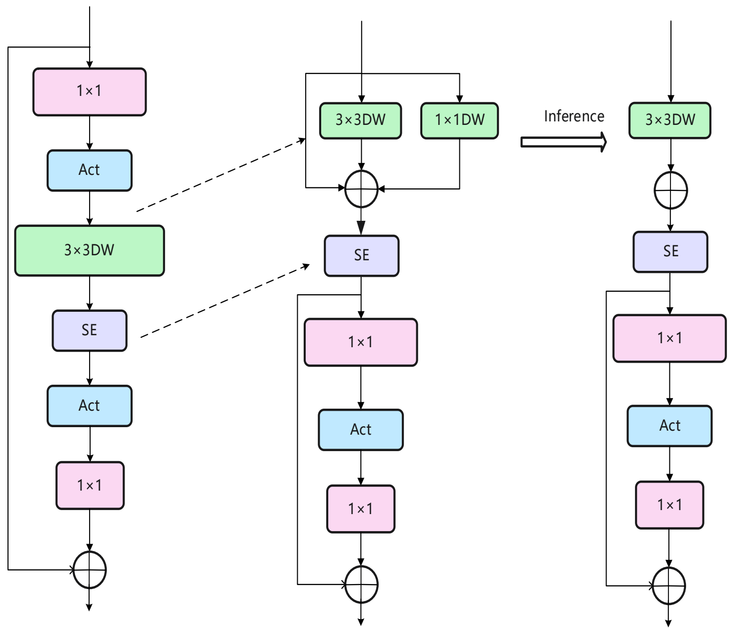


Fig 1. YOLOv10-RLF model structure diagram

2.3 C2f-RVB Lightweight Module

The network is an efficient and RepViT lightweight neural network architecture designed with the goal of without sacrificing accuracy reducing the amount of computation , especially in visual tasks[17]. The network is designed with a novel called RepViT Block module , which enables the model to more efficiently by separating the designs such as token mixer and channel mixer extract image features . This design allows RepViT to reduce the amount of computation while maintaining performance, thereby reducing the model's inference latency in resource-constrained environments such as mobile devices.

The operation flow is Fig. shown in 2, the two in MobileNetV3-L 1×1 convolutions can play the role of channel mixer, while the 3×3 deep convolution plays the role of token mixer, and this design couples the two together, which affects the feature processing. To decouple them, the 3×3 DW is moved to the top, and in addition the optional SE module is moved to the top along with it, thus separating the token mixer from the channel mixer. The structural reparameterization technique was used to enhance the learning ability of the model during the training process, and the modified results are shown in Fig 2.(b), where the right-hand side is the structure during inference can be fused together with 3 × 3 DW and 1 × 1 DW by the structural reparameterization technique, which greatly reduces the number of parameters and the amount of computation.



(a) (b) (c)

Fig 2. RepViT Block module structure

In this paper, the RepViT Block of the RepViT network is used to improve the C2f of the YOLOv10 algorithm to reduce the redundant computation and to achieve a lightweight network. The improvement measure is to introduce the RepViT Block module into the C2f module of YOLOv10, replacing the Bottleneck module in the C2f module with the RepViT Block module, so as to construct a new lightweight module named C2f-RVB module, whose network structure is shown in Fig. 3, which adopts the split operation to stratify the input feature graphs. The module uses split operation to stratify the input feature map, part of the features are retained, part of the features are processed by multiple bottleneck layers, and the multi-scale features are learnt effectively by feature vector transfer and multi-layer nested convolution within the module, and the connection between them is established by combining residuals and jumps, so as to achieve the information propagation and fusion between feature maps of different scales. Finally, the constructed C2f-RVB module is used to replace all C2f modules in the

YOLOv10 network structure to achieve a lightweight improvement of the YOLOv10 algorithm.

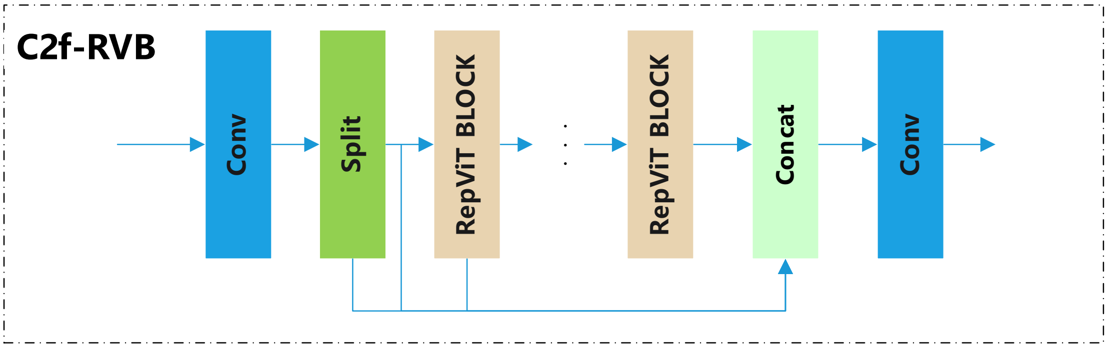


Fig 3. C2f-RVB module structure

2.4 Lightweight detection head v10_LSCD

In deep learning networks, data normalization process is crucial for accelerating model convergence, improving generalization ability and preventing gradient-related problems. Batch normalization (Batch Norm), as a commonly used normalization technique, has significant advantages. It can increase the learning rate, accelerate the model convergence to the optimal parameters, and reduce the training time; reduce the sensitivity of the model to the initial weights, so that the model performance is similar under different initial values; reduce the dependence on regularization techniques such as Dropout and reduce the risk of overfitting; and stabilize the training process by adjusting the distribution of the activation values to prevent the gradient from disappearing or exploding. However, Batch Norm is extremely sensitive to the batch size, because the mean and variance are calculated based on each batch, and if the batch size is too small, the calculation results cannot accurately reflect the overall data distribution, resulting in a decrease in model performance.

Given the limitations of Batch Norm, Group Norm is chosen as an alternative in this paper. Group Norm is more adaptable to maintain the relative relationship between channels and improve the model robustness, it reduces the dependence on small batch size and maintains stable performance with small batch size[18]. Literature shows that Group target detection task in both Norm significantly improves the performance of the detection head in localization and classification tasks. It is not constrained by batch size, and Group Normalization will play a great role in situations where computer memory resources are limited and cannot support larger batch data processing, or where the number of samples must be set to a smaller number due to specific needs[19].

In order to reduce the number of parameters and improve computational efficiency, this paper designs a lightweight detection head Lightweight Shared Convolutional Detection-v10_LSCD for YOLOv10 by adopting Group Norm and Shared Convolutional idea for the structure of YOLOv10 detection head, which is shown in Fig 4. Where the Conv_GN module represents the Group Norm + Conv operation, where the green and purple colours share the weights respectively. Both one-to-one and one-to-many heads are retained, and in the training phase, both heads are involved in training, thus enriching the supervised information of the model; in the inference phase, only one pair of heads is used, thus avoiding non-maximum suppression (NMS) post-processing, with the three one-to-one heads sharing the weights and the three one-to-many heads sharing the weights.

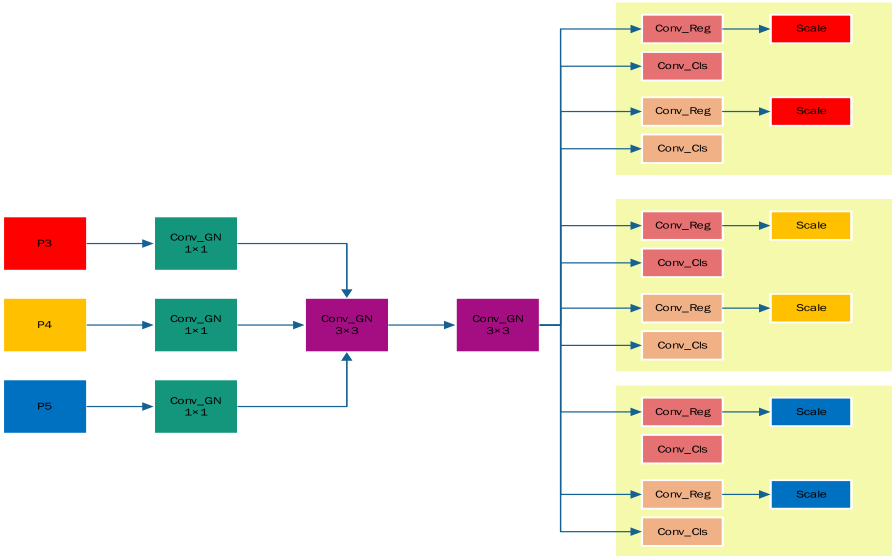


Fig 4. v10_LSCD detection head structure

2.5 FPSC feature pyramid shared convolution

Shared convolution is the key concept of Convolutional Neural Network (CNN). In CNN, the convolutional layer contains a series of convolutional kernels that perform convolutional operations with different locations of the input image, which in turn generates a feature map. The principle of shared convolution is that the same convolution kernel performs convolution operations with the same weight parameters at different locations in the image. In other words, no matter where the convolution kernel is located in the image, its weight parameters remain the same, thus achieving weight sharing. This concept has a number of core advantages. The weight sharing mechanism dramatically reduces the number of parameters that the network needs to be trained on, improves computational efficiency, and reduces the risk of over-fitting. By virtue of the consistency of the weights of the convolution kernel over the whole image, shared convolution can extract image local features more efficiently and maintain spatial structural information. Its shift invariance empowers the model to be robust to input data translation transformations, which further strengthens the generalization ability of the model. In addition, shared convolution can handle multi-channel inputs, fully exploiting the information between different channels and enhancing the model expression ability. These features make shared convolution outstanding in image processing and other fields.

SPP (Spatial Pyramid Pooling) is a technique in deep learning, especially in Convolutional Neural Networks (CNNs) for processing input images of different sizes and scales.SPPF (Spatial Pyramid Pooling- Fast) is an improved version of Spatial Pyramid Pooling. Pooling (SPPF) is an improved version of Spatial Pyramid Pooling (SPP), which achieves multi-scale feature extraction by applying the same size of pooling kernels serially instead of applying different sizes of pooling kernels in parallel in SPP, in order to increase the processing speed and efficiency, while maintaining the ability of multi-scale feature fusion[20]. In this paper, a Feature Pyramid Shared Convolution module Feature Pyramid Shared Conv, FPSC, is designed using the shared convolution idea, and its network structure is shown in Fig 5.

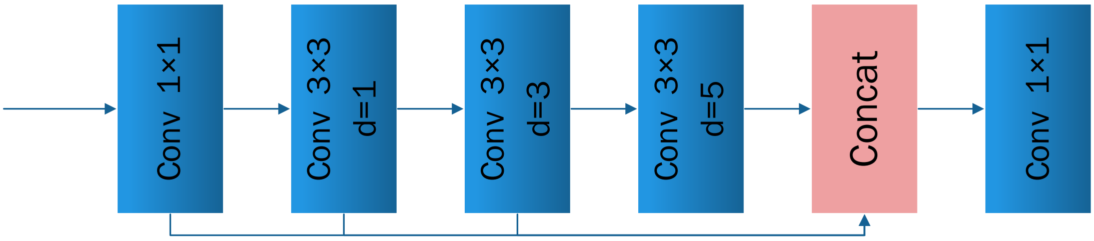


Fig 5. FPSC modular structure

This module replaces the Max Pooling operation (Max Pooling) in SPPF with a 3×3 convolution by using a serial structure. The three convolutional kernels use the same weight parameters and in this way share the weights.The pooling operation in SPPF loses the detail information, the convolutional operation has higher flexibility and expressiveness in feature extraction, and also has the learning of the parameters, which is more capable of capturing fine-grained features in feature extraction. The convolution kernel in the structure has different convolutional expansion rates and by using convolutional layers with different expansion rates, the module is able to extract features at different scales. This is advantageous for capturing information of different sizes and different contexts in the image, with low expansion rates used to capture local details and high expansion rates used to capture global context, which provides a greater advantage for multi-target small target detection. To prevent the problem of excessive amount of parameters, the convolution is made to perform weight sharing and at the same time the utilization of parameters is improved.

2.6 SIOU loss function

The loss function of the detection model consists of 3 parts[21]: which are location loss, confidence loss, and categorization loss. Due to the large number of small targets and the dense distribution of some of the categories of grapevine leaf diseases, the set of metrics for the bounding box regression metrics is considered. Compared with the CIOU of the original model, the vector angle between the centroids of the two frames when the predicted frame returns to the real frame is taken into account on this basis, which accelerates the convergence speed and improves the computational accuracy.The SIOU loss function consists of four parts: angle loss, distance loss, shape loss, and IOU loss, as shown in Figure 6.

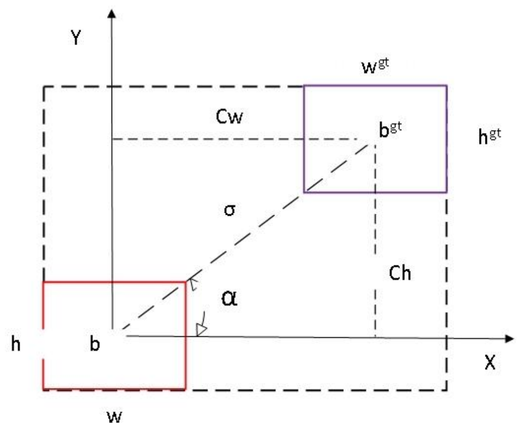


Fig 6. Schematic diagram of SIOU parameters

The angular loss calculation process is:

$$\Lambda = 1 - 2\sin^2(\arcsin(c_h/\sigma) - \pi/4)$$

included among these

$$\sigma = \sqrt{\left(\mathrm{b}_{\mathrm{cx}}^{\mathrm{gt}} - \mathrm{b}_{\mathrm{cx}}\right)^2 + \left(\mathrm{b}_{\mathrm{cy}}^{\mathrm{gt}} - \mathrm{b}_{\mathrm{cy}}\right)^2}$$

$$c_h = \max\left(b_{cy}^{gt}, b_{cy}\right) - \min\left(b_{cy}^{gt}, b_{cy}\right)$$

(b_{cx} ,b_{cy}), (b_{cx}^{gt} ,b_{cy}^{gt}) are the coordinate positions of the prediction frame and the real frame. Its makes the prediction frame approach the nearest coordinate axis quickly and then approach the real frame along the coordinate axis, thus speeding up the convergence of the distance between the two frames.

The distance loss refers to the Euclidean distance between the centroid of the prediction frame and the real frame, and in calculating the distance loss, SIOU adjusts the minimum rectangular frame enclosing the prediction frame and the real frame according to the angular loss, which is calculated by the formula:

$$\Delta = 2 - e^{-\rho_x \gamma} - e^{-\rho_y \gamma}$$

Among them.

$$\rho_x = \left(\frac{b_{cx}^{g} - b_{cx}}{c_w}\right)^2, \rho_y = \left(\frac{b_{cy}^{g} - b_{cy}}{\mathrm{c_h}}\right)^2, \gamma = 2 - \Lambda$$

c_w ,c_h are the width and height between the center distance of the real frame and the predicted frame.

3. Experiment and result analysis

3.1 Experimental data set

In this paper, selected four as the object of the recognition studycommon diseases of grape leaves, black_root, black_measles, blight and healthy were ,, and the data images were partly collected from Jiangbei Vineyard in Jilin City, Jilin Province, and partly selected from the open database Plant Village, the whole dataset totaled 4290 images, and the example diagram is shown in Fig 7. In this dataset, most of the images contain small targets of multiple diseases. In order to improve the training efficiency of the model and save computational resources and training time, the image resolution size is standardized to 640 pixels × 640 pixels. Label each leaf in the dataset using LabelImg image annotation tool, the labeling results were saved in YOLO txt text format, and the labeling was performed using rectangular picture frames. The dataset was randomly divided into training set, validation set and test set according to the ratio 7:2:1, and a total of 3003 images were obtained in the training set, 858 images in the validation set and 429 images in the test set.

Black rot leaf blight black measles health.

Fig 7. Example of the dataset images

3.2 Test environment configuration and evaluation indicators

In this paper, the experimental training environment uses Ubuntu 22.04 LTS operating system, the CPU configuration is 14 vCPU Intel(R) Xeon(R) Platinum 8362 CPU @ 2.80GHz, the GPU as shown in Table 1 below. configuration is NVIDIA RTX3090 graphics card with 24GB of video memory, and the Pytorch deep learning framework is used, the Pytorch version is 2.3.0, cuda version is 12.1, Python version is 3.12, and the training parameters are configured Pytorch version is 2.3.0, cuda version is 12.1, Python version is 3.12, and the training parameter configuration is shown in Table 1 below.

Table 1 Training hyperparameter setting

parameters	numerical value
imgsize	640×640
epoch	300
batch_size	32
optimizer	SGD
mosaic	1.0
1r0	0.01
weight_decay	0.005

In order to evaluate the superiority of the improved algorithm in this paper, the whole class average precision (mAP), the number of floating-point operations (Giga Floating-point Operations Per Second, GFLOPs), the model size (model size) and the number of parameters (Parameters) as evaluation metrics. where mAP@0.5 is the average precision mean when the IOU is 0.5 and mAP@0.5-0.95 refer to the average precision mean value of IOU from 0.5 to 0.95 with a step size of 0.05.

3.3 Ablation experiments

In order to verify the performance of the improved model, this paper on designs ablation experiments basedYOLOv10n. The YOLOv10n model is used as the base network, and the model obtained by replacing the C2f module with the C2f-RVB module in YOLOv10n is YOLOv10n-RVB; the model obtained by replacing the original YOLOv10 detection head with the LSCD detection head is The model obtained by replacing the SPPF with the self-developed FPSC module and using the SIOU loss function is YOLOv10n-FPSC, and the model obtained by applying the previous two strategies simultaneously is YOLOv10n-RVB-LSCD. The ablation experiments are shown in Table 2.

Table 2 Results of ablation experiment

mould	C2f-RVB	LSCD	FPSC	mAP@0.5/%	mAP@0.5-0.95/%	GFLOPs	model size/MB	Parameters/10^6
YOLOv10n				90.7	71.1	6.5	5.5	2.3
YOLOv10n-RVB	√			90.3	70.4	5.1	4.7	1.8
YOLOv10n-LSCD		√		90.0	69.7	6.2	4.0	1.9
YOLOv10n-FPSC			√	90.9	71.7	6.5	5.8	2.4
YOLOv10n-RVB-LSCD	√	√		88.7	68.7	4.8	3.2	1.5
YOLOv10n-RLF	√	√	√	90.4	70.9	4.8	3.5	1.6

First, comparing the YOLOv10n and YOLOv10n-RVB models, using the C2f-RVB module instead of the C2f module reduces the parameters by 21.74 %, the GFLOPs by 19.54 %, and the size of the model weights is compressed by 0.8 MB.Comparing the YOLOv10n and YOLOv10n-LSCD models The YOLOv10n and YOLOv10n-LSCD models, with 4.62 % fewer parameters, 17.39 % fewer GFLOPs, and 1.5 MB compression of model weight size, have better performance especially in the loss curve. From the above comparison results, it can be seen that both the C2f-RVB module and the LSCD detection head are able to significantly reduce the model parameters and the amount of operations, and the LSCD detection head has a lower loss of distributed focus loss, which is more accurate for the localization and identification of the disease.

To further reduce the parameters and computation of the model, the above two modules are superimposed to obtain the model YOLOv10n-RVB-LSCD.Comparing the YOLOv10n and YOLOv10n-RVB-LSCD models, the Parameters is reduced by 34.8 %, the GFLOPs is reduced by 26.2 %, and the size of the model weights is compressed by 2.3 MB. Along with the decline in the number of parameters and the amount of computation, the average accuracy of the model also declined severely, by 2.0% for mAP@0.5 and 2.4% for mAP@0.5-0.95. In order to improve the detection accuracy, as well as the recognition rate of small targets, the self-developed FPSC module is used to replace the SPPF, and in order to verify the effectiveness of the module, comparative experiments are conducted. Comparing the YOLOv10n and YOLOv10n-FPSC models, mAP@0.5 improves by 0.2% and mAP@0.5-0.95 improves by 0.6% with basically no increase in computation, which can prove that the module can improve the accuracy of disease detection.The module is applied to the YOLOv10n-RVB-LSCD model to obtain the YOLOv10n-RLF model, and it can be seen that the enhancement effect is more obvious, mAP@0.5 improves by 1.7% and mAP@0.5-0.95 improves by 2.2%, and the model has a high accuracy rate to meet the actual detection needs. Compared with YOLOv10n, the parameters is reduced by 30.4%, the GFLOPs is reduced by 26.2%, and the model weight size is compressed by 2.0 MB. Combining the above improvements, the model of this paper has a significant reduction of the number of parameter Parameters and the amount of operation to satisfy the higher detection accuracy, and the size of the model is 3.5 MB, which is suitable for mobile deployment.

In order to show the improvement effect of this paper's model more intuitively, a random part of the pictures in the test set visualize its features and display them in the form of heat map, the specific effect is shown in Fig 8. From the comparison figure, it can be seen that for small targets this model has better detection accuracy and effect.

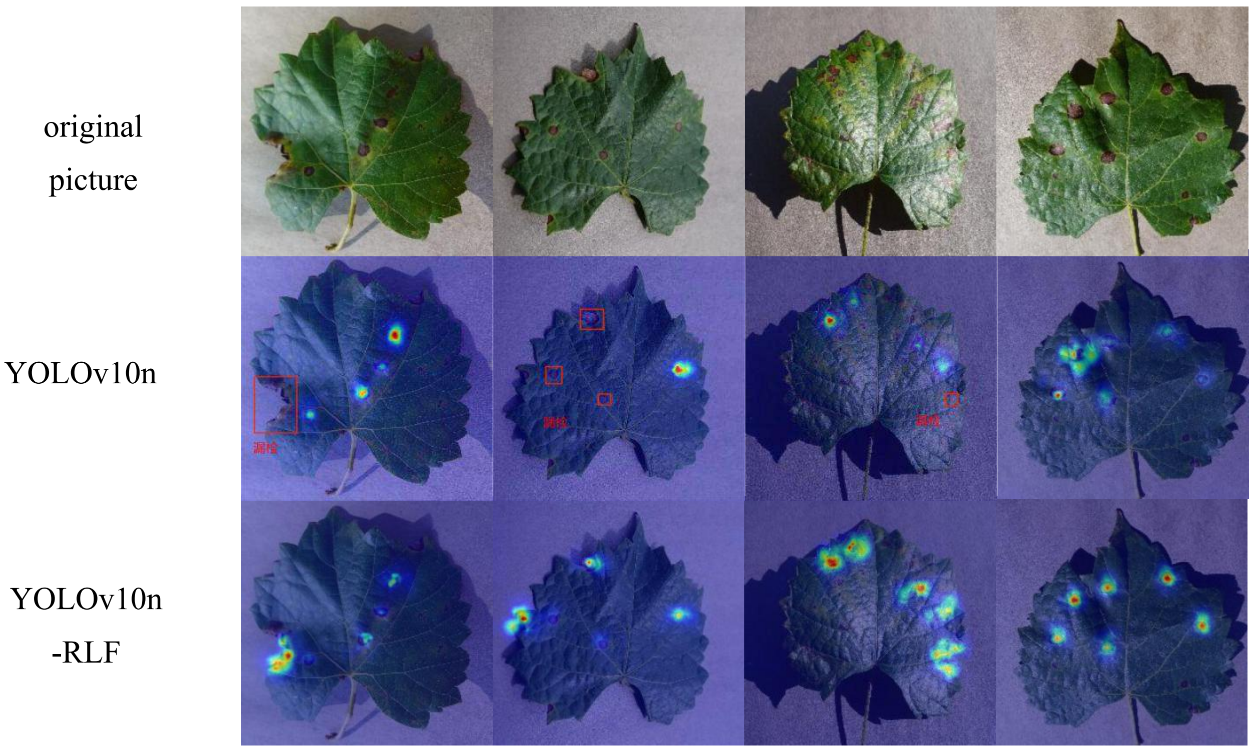


Fig 8.Visualization results of the model's ability to focus on small parts of the blade before and after improvement

3.4 Comparison of Network Layer Parameters

To further validate the effectiveness of the improved module in this paper, we use the module network parameters of the YOLOv10-RLF model and the original model for comparison, as shown in Table 3. Relative to the YOLOv10 model, YOLOv10-RLF achieves a significant reduction in the number of parameters and computation, especially in the detection head section, the parameters of the v10_LSCD module are significantly reduced from 929808 to 121779, and the computation is reduced from 3.80 GFLOPs to 1.64 GFLOPs; tThe C2f-RVB module replaces the original C2f module in several positional layers, with a reduction in parameters of about 40%-50%;and after replacing SPPF with FPSC, the number of parameters is partially improved, although the computation volume remains basically the same. These optimisations allow YOLOv10-RLF to significantly improve computational efficiency while maintaining high detection accuracy, making it more suitable for deployment on resource-constrained devices.

Table 3 Comparison of network parameters

Location Layers	Modular Structure		Parameters		GFLOPs	
	YOLOV10	YOLOv10-RLF	YOLOV10	YOLOv10-RLF	YOLOV10	YOLOv10-RLF
2	C2f	C2f-RVB	7360	4048	0.39	0.22
4	C2f	C2f-RVB	49664	22080	0.64	0.29
6	C2f	C2f-RVB	197632	85120	0.64	0.28
8	C2f	C2f-RVB	460288	233088	0.37	0.19
9	SPPF	FPSC	164608	312064	0.13	0.13
13	C2f	C2f-RVB	148224	91968	0.48	0.30
16	C2f	C2f-RVB	37248	23456	0.48	0.31
19	C2f	C2f-RVB	123648	67392	0.40	0.22
23	v10Detect	v10_LSCD	929808	121779	3.80	1.64

3.5 Comparison of the performance of different models

In order to verify the detection effect and lightweight characteristics of YOLOv10n-RLF, the grape disease detection network proposed in this paper, we selected other mainstream networks Faster R-CNN network, YOLOX-tiny, YOLOv6n network, YOLOv7-tiny network and YOLOv8n network to compare with this paper's network in this dataset, and the experimental results are shown in Table 4 shows.

From the experimental results, the Faster R-CNN dual-target detection network far exceeds the other single-target detection networks in terms of the number of operations and parameters, and is even worse in terms of the model occupancy size, which is not favourable for deployment on mobile devices. Among the single-target detection networks, the YOLOv10n-RLF network has the lowest number of operations and parameters, as well as a greater advantage in model occupancy size for mobile deployments.In terms of detection accuracy, compared to YOLOX-tiny, mAP@0.5 和 mAP@0.5-0.95 improves by 6.1% and 17.0%, respectively. Compared to other YOLO models the difference is not significant, i.e., the maximum difference is 0.4%, which provides high detection accuracy.

Table 4 Comparative experiment results with different models

mould	mAP@0.5/%	mAP@0.5-0.95/%	GFLOPs	model size/MB	Parameters/10^6
Faster R-CNN	83.7	61.3	208	315.6	41.39
YOLOX-tiny	84.3	53.9	7.6	66.2	5.0
YOLOv6n	90.6	71.3	11.9	8.7	4.2
YOLOv7-tiny	90.7	69.1	13.2	12.3	6.0
YOLOv8n	90.7	71.2	8.2	6.3	3.0
YOLOv10n	90.7	71.1	6.5	5.5	2.3
YOLOv10n-RLF	90.4	70.9	4.8	3.5	1.6

4. Conclusion

In order to cope with the characteristics of more small target diseases grape leaf diseases with diverse morphology and as well as the demand of deploying mobile devices, this paper proposes a lightweight grape leaf detection method, YOLOv10n-RLF, based on the improved YOLOv10n. Firstly, instead of the C2f module, we use the C2f-RVB module, which extracts image features more efficiently as a way of reducing the computational burden. Secondly, the thought of shared convolution is introduced in the detection header, which significantly reduces the detection header parameters and enhances the feature fusion capability using the shared convolution mechanism. Finally, in order to improve the detection accuracy as well as to improve the recognition of small targets, the self-developed FPSC module is replaced with SPPF by using the inflated convolution to extract retaining more detailed information. The experimental results show that the improved YOLOv10n-RLF model has only 1.6×10^6 model parameters and 4.8 GFLOPs of operation under satisfying higher detection accuracy, which is 30.4% less parameter numbers and 26.2% less operation compared with YOLOv10n, and has better recognition effect enhancement when targeting small targets of diseases.

The grapevine disease detection model constructed in this study, through the multi-scale feature fusion mechanism and efficient feature extraction module, achieves further lightweighting of the model and improved detection effect on small targets under the premise of guaranteeing high detection accuracy, which provides key technical support for the subsequent research on precise variable application.However, there are still limitations in the current study that need to be broken through: for the disease dataset, the training data only cover three common diseases, namely black rot, black measles and leaf blight, and there is a lack of research on other diseases and pests as well as the complexity of the background, and the data samples will be further enriched in the future.

Acknowledgements

Y.J. and G.C. were responsible for the planning of the experiments and the improvement of the models in the experiments.S.J. was in charge of the verification of the model.S.J. and G.C. were in charge of the processing and annotation of the datasets.All authors reviewed the manuscript.

Author contributions

All authors contributed to the study’ s conception and design and approved the final version of the manuscript. There is no human participate involved in this research. This article manuscript is created from collection of data set.

Funding

This work was supported by Science and Technology Research Project of the Education Department of Jilin Province (JJKH20220054KJ and JJKH20261355KJ).

Data availability

The data used to support the findings of this study are available from the corresponding author upon request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Nagi R, Tripathy S S. Deep convolutional neural network based disease identification in grapevine leaf images[J]. Multimedia Tools and Applications, 2022, 81(18): 24995-25006.
2. Zhang L, BAYIN T, Zeng Q, et al. Grape Disease Detection Algorithm Based on StyleGAN2-ADA and Improved YOLO v7 [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(01):241-252.

3. Zhou J, Li J, Wang C, et al. Crop disease identification and interpretation method based on multimodal deep learning[J]. Computers and Electronics in Agriculture, 2021, 189: 106408.
4. Zhang C, Zhang S, Yang J, et al. A leaf disease identification using genetic algorithm and correlation based feature selection method[J].International Journal of Agricultural and Biological Engineering,2017,10(2):74-83
5. Rehman Z, Khan, MA, Ahmed, F. et al. Recognizing apple leaf diseases using a novel parallel real-time processing framework based on MASK RCNN and transfer learning: An application for smart agriculture. IET Image Process. 2021;15: 2157–2168. https://doi.org/10.1049/ipr2.12183
6. Thakur P S, Sheorey T, Ojha A. VGG-ICNN: A Lightweight CNN model for crop disease identification[J]. Multimedia Tools and Applications, 2023, 82(1): 497-520.
7. Hu Y, Zhao X, Li P, et al. Tomato fruit detection innatural environment based on improved YOLOv5[J]. Journal of Chinese AgriculturalMechanization,2023,44(10):231-237
8. YUE K, ZHANG P, WANG L, et al. Recognizing citrus in complex environment using improved YOLOv8n[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(8): 152-158. (in Chinese with English abstract)
9. Zhao L, Wang L. A new lightweight network based on MobileNetV3[J]. KSII Transactions on Internet and Information Systems (TIIS), 2022, 16(1): 1-15.
10. Zhang H, Dai C, Ren J, et al. Research on grape disease identification method based on GhostNetV2 improved nnnnYOLO v8 model[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 1-11.
11. Zheng Y, Chen R, Yang C, et al. Improved YOLOv5s based identification of pests and diseases in citrus[J]. Journal of Huazhong Agricultural University, 2024，43（2）：134-143.
12. Ma C, Zhang H, Ma Xi, et al. Method for the lightweight detection of wheat disease using improved YOLOv8[J]. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39(20): 159-166.
13. Jin Z, Jin B, Song X, et al. Improved YOLOV10 Algorithm And Its Application On Pothole Detection[J]. Computer Applications and Software, 1-8.http://kns.cnki.net/kcms/detail/31.1260.tp.20250307.0824.002.html.
14. Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.
15. Mao M, Lee A, Hong M. Efficient Fabric Classification and Object Detection Using YOLOv10[J]. Electronics (2079-9292), 2024, 13(19).
16. Guan S, Lin Y, Lin G, et al. Real-time detection and counting of wheat spikes based on improved YOLOv10[J]. Agronomy, 2024, 14(9): 1936.

17. Wang A, Chen H, Lin Z, et al. Repvit-sam: Towards real-time segmenting anything [J]. arXiv preprint arXiv:2312.05760, 2023.
18. Wu, Y. and He, K. (2019) Group Normalization. International Journal of Compute r Vision, 128, 742-755.https://doi.org/10.1007/s11263-019-01198-w
19. Shen, Q, Zhang, L, Zhang, Y, et al. Distracted Driving Behavior Detection Algorith m Based on Lightweight StarDL-YOLO. Electronics 2024, 13, 3216. https://doi.org/1 0.3390/electronics13163216
20. Jooshin H K, Nangir M, Seyedarabi H. Inception-YOLO: Computational cost and ac curacy improvement of the YOLOv5 model based on employing modified CSP, SPP F, and inception modules[J]. IET Image Processing, 2024, 18(8): 1985-1999.
21. Gevorgyan Z. SIoU loss: More powerful learning for bounding box regression[J]. ar Xiv preprint arXiv:2205.12740, 2022.