

Detection of Rice Leaf Diseases Based on Improved YOLOv8n

Xue Liu¹, Shunyong Zhou^{1,*}, Ziyang Peng¹, Hangling Zhang¹, Qin Hu¹

¹ School of Automation and Information Engineering, Sichuan University of Science & Engineering, Zigong 643000, China

² Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science & Engineering, Zigong 643000, China.

* Corresponding author: Shunyong Zhou (Email: lx1352839149@163.com)

Abstract: Rice plays an important role in human food chain, but it is easily affected by related diseases in its growth, which seriously affects rice yield. At present, there are many methods to detect rice leaf diseases, but there are still problems such as large model and poor detection effect. In order to solve the limitations of traditional manual and deep learning-based models, a lightweight model based on improved YOLOv8n is proposed. Firstly, SPPD (Spatial Pyramid Pooling and Dilated) structure with different void ratios composed of GELU activation function was added to the Backbone network to increase the receptive field of the network, and Coordinate attention (CA) was combined to help the model pay attention to the characteristics of rice diseases and improve the detection accuracy. Finally, XSEPCov (Extremely Separated Convolution) is used to reduce the parameters and improve the efficiency of the model. The model was trained and tested on the self-built rice leaf disease image data set, and its mAP @0.5 reached 89.3% and FPS reached 217. The proposed SCX-YOLOv8n model is a lightweight, efficient and high-performance rice leaf disease detection model, and compared with other mainstream models, it also has certain advantages in accuracy and recall rate, which can provide an accurate and accurate rice leaf disease detection and other related fields.

Keywords: Rice leaf disease; GELU; SPPG; Coordinate attention mechanism; XSepConv.

1. Introduction

Rice is the food resource that human beings depend on, and it is the most extensive cereal plant in the world. However, in the process of rice growth, rice production is often reduced due to the influence of climate, temperature, diseases, pests and viruses, which may directly lead to rice death and seriously affect the agricultural economy [1]. This means that rice must be protected from these factors, and at the same time, specific disease types can be determined through disease detection, so as to guide the rational selection of control methods and pesticides, which will help to avoid excessive use of pesticides and reduce. In order to detect the diseases of rice plants to the maximum extent, the disease degree of rice leaves is generally detected. The traditional rice disease detection method relies on visual detection by experts or experienced farmers, but this method is not suitable for large-scale cultivated farmland, and it takes a long time and cannot feed back the changes of rice diseases in time. With the development of deep learning technology, the agricultural field has become more intelligent and accurate, which can detect the growth of rice in real time and provide a reasonable solution for the agricultural sector [2]. For example, MATIN used Alexnet neural network to detect three kinds of diseases, but in the experimental data set, the number of each image is only 40 [3]. ZHOU et al. combined FCM-KM and Faster R-CNN to detect rice diseases, which has high accuracy, but the model is too large to meet the actual needs [4]. CHEN et al. used enhanced Inception module to extract image features for SSD algorithm, and used transfer learning to optimize the model. This method can achieve the expected performance [5]. YUMANG used Tiny YOLOv3 algorithm to identify bacterial leaf blight. Among 20 test pictures, there was only one wrong prediction, with high accuracy [6]. MASYKUR F uses drones to collect images at different distances, and uses

YOLOv4 algorithm to carry out experiments, which can detect the existence of pests on rice [7]. JHATIAL used YOLOv5 algorithm to detect four kinds of rice diseases, but only the original algorithm was used, with fewer training rounds, and the model did not reach the ideal state [8]. The above-mentioned researchers put forward that some rice disease detection models have some limitations, but they have proved the potential of deep learning-based models in rice disease detection. Aiming at the diversity and complexity of rice leaf diseases, this paper realizes the efficient detection and identification of rice leaf diseases by applying YOLOv8 algorithm, mainly using the spatial pyramid pooling (SPPG) structure with different expansion rates [9] to increase the receptive field of the network; Secondly, the CA attention mechanism [10] is used to help the model better capture the important information in the image and improve the robustness of the model. At the same time, considering the complexity and parameters of the model, XsepConv structure [11] is used to replace the down-sampling structure of the network, which reduces the demand for computing resources.

2. The principle and Improvement of YOLOv8n Model

2.1. The principle of yolov8 model

YOLOv8 is a target detection algorithm based on deep learning, which combines the advantages of high accuracy and real-time performance. It is a new YOLO series algorithm proposed by Ultralytics, which mainly includes five models, such as YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l and YOLOv8x. It can quickly and accurately realize tasks such as detection, tracking, classification, case segmentation and posture estimation, and can effectively deal with rice leaf diseases.

YOLOv8n is relatively simple, contains fewer network

layers, and is easier to train and adjust. In order to maintain high accuracy and faster reasoning speed, this paper chooses YOLOv8n as the basic model and improves it. YOLOv8n is mainly composed of input layer, Backbone, Neck and output layer, as shown in Figure 1. Among them, the input layer mainly preprocesses the image and enhances the data. Backbone mainly extracts the characteristics of rice diseases through CBS, C2F and SPPF structure [12]. CBS structure is responsible for the down-sampling process of the feature map, which is composed of Convolutional Layer, Batch Normalization Layer and SiLU activation function. It is repeatedly stacked in the Backbone network to gradually extract the advanced features of the image. C2F first adjusts the number of channels through CBS structure for subsequent feature fusion. After processing the input feature map, it divides the results into two blocks according to the channel dimension, and converts the blocked results into a list. Then, the last block tensor in the list is extracted by the Bottleneck module, which uses residual connection to enhance the feature representation ability of the model. The low-level features and high-level features can be fused by adding the input features and the convolved output features. At the same time, feature information can be further extracted and enriched through multiple Bottleneck modules, thus enhancing the expressive ability of the network. The SPPF module captures the features of different receptive field sizes by sequentially pooling the maximum pool with a kernel size of 5, and provides wider context information to enhance the accuracy of detection. The Neck module of YOLOv8n is composed of FPN [13] and PANet [14] structures, which eliminates the convolution operation before upsampling in YOLOv5, and realizes the lightweight of the model while maintaining the original performance. FPN constructs a pyramid by bottom-up and horizontal connection, and PANet fuses the bottom-up feature map with the top-down feature map layer by layer to obtain a feature pyramid with rich semantic information and high resolution, which can effectively extract and utilize multi-scale features. The output layer of YOLOv8n adopts the idea of "double decoupling heads", and two independent sub-networks are used to process the target classification and bounding box regression, thus improving the effect and performance of the model.

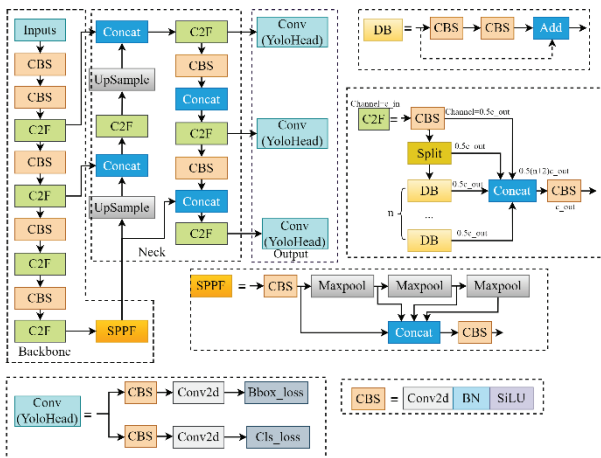


Figure 1. YOLOv8n model structure diagram

2.2. Improvement of yolov8n model

The improvement of this model mainly consists of three

parts: SPPG structure, CA attention mechanism and XSepConv lightweight structure, and the model is named SCX-YOLOv8n. The structure is shown in Figure 2.

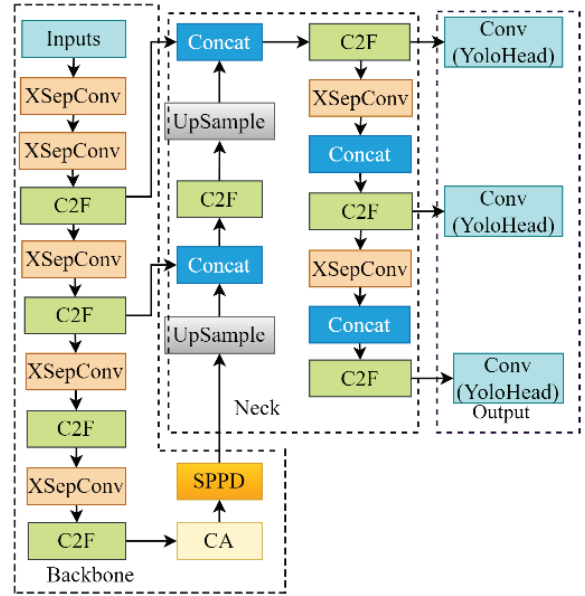


Figure 2. Structural diagram of scx-yolov8n model

2.2.1. Improve the SPPF structure

The size of the receptive field of the feature map is related to whether the network can find more global or detailed information in the image. Considering that there are many small targets in the rice disease image, and the contextual information of the receptive field can effectively help to detect small-scale targets, a GEConv convolution structure with different dilated convolution is added to the SPPF structure, which can expand the receptive field, capture more characteristic information and improve the detection accuracy without reducing the resolution of the feature map and increasing the calculation amount, as shown in Figure 3. Firstly, the feature maps generated by the backbone network are transmitted to the CBS module to realize the cross-channel interaction of information. Then, the maximum pooling with the kernel size of 3 is carried out three times in turn. The generated three receptive field feature maps are sent to the GEConv structure with different hollowing rates, and the CBS module passes through the GEConv structure with a hollowing rate of 1, and then the features of all branches are connected. Finally, the feature layer is output through CBS convolution. GEConv consists of ordinary convolution, BN layer and GELU activation function [15]. As shown in Formula (1), the GELU activation function is nonlinear in the whole real number range and smooth in the negative number range, which makes it more suitable for optimization methods such as gradient descent, and easier to deal with when propagating backward, which helps to reduce the problem of gradient disappearance, improve the representation ability of the model, and preserve the linear characteristics in more input ranges, thus better processing the input feature information.

$$GELU(x) = xP(X \leq x) = x \times \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right] \quad (1)$$

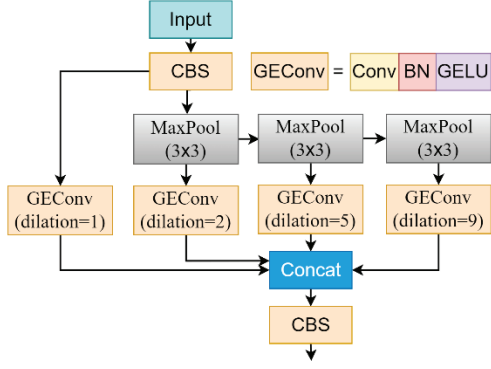


Figure 3. Structure diagram of SPPG model

2.2.2. CA attention mechanism

In order to effectively learn the target characteristics and location information of rice diseases, CA attention mechanism was added to the Backbone feature extraction network to automatically learn the correlation and feature representation between different channel locations. As shown in Figure 4, CA attention mechanism can capture the features on two channels at the same time, which is helpful to extract important feature information without increasing network parameters and model calculation.

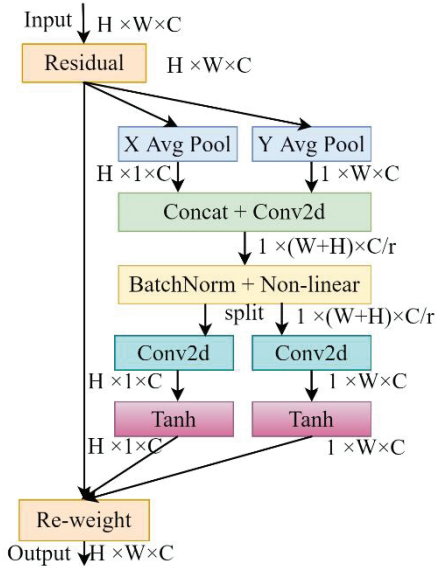


Figure 4. Structure diagram of CA attention mechanism

First of all, in order to reduce the computational complexity, CA attention mechanism will pool each channel in the input feature map through the global average operation based on width and height to obtain an information feature map containing each channel, as shown in Formula (2) and (3); Then, the feature map connected in series along the spatial dimension is transferred to the convolution layer, BN layer and activation function layer with the convolution kernel size of 1×1 , as shown in Formula (4); Then, the generated feature map is sliced along the width-height dimension to obtain two separate feature maps, and after passing through the convolution layer with Tanh activation function, the attention weight in the width-height direction is generated, as shown in Formula (5) and (6); Finally, with the input feature map, the feature value of each channel is multiplied by the corresponding attention weight through element-by-element multiplication to obtain the output feature map of CA

attention mechanism, as shown in Formula (7). Compared with Sigmoid activation function, Tanh has better output range, larger gradient (near zero) and zero centrality, which can alleviate the problem of gradient disappearance. CA attention mechanism not only considers the importance of information between different channels, but also considers the spatial location information. It connects the horizontal and vertical features to form the whole global feature, which can better help the model pay attention to the characteristics of rice diseases.

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (2)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < W} x_c(j, w) \quad (3)$$

$$f = \delta \left(F_1 \left(\left[z^h, z^w \right] \right) \right) \quad (4)$$

$$g^h = \sigma \left(F_h \left(f^h \right) \right) \quad (5)$$

$$g^w = \sigma \left(F_w \left(f^w \right) \right) \quad (6)$$

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (7)$$

2.2.3. Lightweight design

In order to reduce the amount of calculation caused by the attention mechanism of SPPG and CA, and at the same time speed up the reasoning to enhance the flexibility of model deployment, XsepConv structure is used to reduce the number of parameters and improve the efficiency of the model. XSepConv structure is a combination of depth separable convolution [16] and SE attention mechanism [17], which extends and improves the traditional convolution structure. As shown in Figure 5, it can not only capture the detailed features of the feature map, improve the perceptual ability of the model, but also reduce the calculation cost and parameter size. Firstly, the depth separable convolution with the convolution kernel size of 2×2 is used to reduce the computational complexity. Then, an improved symmetric filling strategy is used to compensate for the negative effects caused by convolution, that is, convolution with a convolution kernel size of 1×3 in the width of the feature map and a convolution kernel size of 3×1 in the height can significantly reduce the demand for computing resources and improve the operation efficiency, and on the other hand, it can better capture the context information of the input image, which is conducive to extracting more global features.

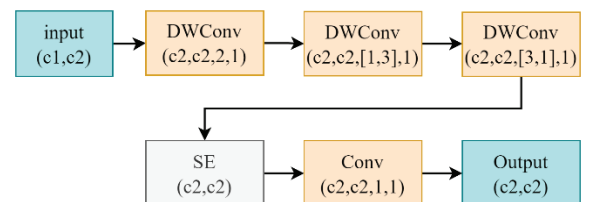


Figure 5. XsepConv convolution diagram

The XSepConv structure also introduces the SE attention mechanism, which allows the model to pay more attention to the predetermined target of the image. The SE module is to gain the attention in the channel dimension. As shown in Figure 6, firstly, Squeeze the feature map obtained by convolution to take the global spatial characteristics of each channel as the representation of the channel, then carry out nonlinear transformation Excitation to learn the dependence degree of each channel, and finally adjust the feature map by multiplying it with the number of channels on the previously unplaced feature map, as shown in the figure. Depth separable convolution can be divided into depth convolution and point-by-point convolution. An effective alternative to ordinary depth convolution.

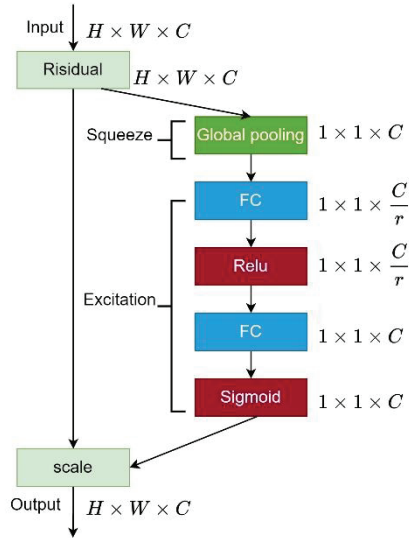


Figure 6. Structural diagram of se attention mechanism

3. Model Performance Analysis

3.1. Experimental environment and parameter setting

The model training is conducted on the windows operating system using the PyTorch framework for training and testing, and the environment configuration is shown in Table 1. The input image size is set to 320 pixels× 320 pixels, the batch size is set to 16, the number of training steps is set to 200, the learning rate is set to 0.01, the momentum is set to 0.937, the random gradient descent is optimized with the weight attenuation of 0.005, and the number of work processes is 2.

Table 1. Experimental Configuration

operating system	Microsoft Windows 10 (64位)
CPU	12th Gen Intel(R) Core(TM) i5-12400(2500 MHz)
internal storage	16.00 GB (3200 MHz)
display card	NVIDIA GeForce RTX 3060 (12288MB)
Memory	12G

3.2. Experimental data set

The detection target of this paper is rice diseases. In the process of rice growth, it will be caused by unsuitable growth environment or harmful substances and infectious diseases in the environment. According to this study, four diseases of rice, namely brown spot, rice blast, sheath blight and bacterial blight, were studied, among which rice sheath blight was very

similar to rice blast. The data set needed for the research comes from the public data set 3-class-riceleafdisease, Rice Leaf Disease and other public images on the network platform. A large number of pictures of real scenes are added to solve the problem of over-fitting of high-precision images and increase the generalization ability of the model. Firstly, all the data were manually eliminated, and labeled with labeling software according to the rich experience of rice professionals and growers. Examples of disease types and labels of rice are shown in Figure 7.

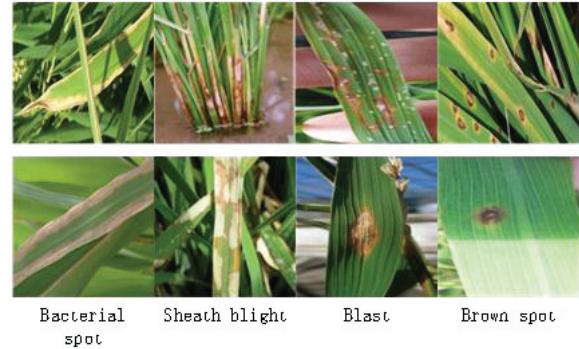


Figure 7. Types of rice diseases

In deep learning, data set preprocessing and data enhancement is a key step, which can help improve the performance and robustness of the model. Firstly, in order to ensure the quality of data, manual cleaning and screening are carried out, and then the data set is enhanced, such as image flipping, image mirroring, brightness and contrast adjustment. In the experiment, there are 3692 enlarged data sets, and the training set and verification set are divided according to the ratio of 8: 2 for training, tuning and evaluation of the model.

3.3. Evaluation indicators

In order to evaluate the effectiveness of SCX-YOLOv8n network model in rice pest detection, this study selected Precision, Recall, average precision (mAP), parameters, model size (weight) and FPS as evaluation indicators, and used mAP@0.5 with IoU threshold of 0.5 as evaluation indicators. The formula for calculating the accuracy rate is shown in Formula (8), the formula for calculating the recall rate is shown in Formula (9) and the formula for calculating the average accuracy score is shown in Formula (10).

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (9)$$

$$mAP = \frac{\sum AP}{N(class)} = \frac{\sum \int_0^1 P(R) dR}{N(class)} \quad (10)$$

Among them, TP , FP , FN represents the number of rice diseases correctly detected, the number of rice diseases and insect pests incorrectly detected and the number of rice diseases and insect pests incorrectly detected respectively, and N represents the number of categories; $Precision$ is the correct proportion of all prediction targets; $Recall$

indicates the correct proportion of all real targets. AP is the area under the PR curve (recall is the horizontal axis and accuracy is the vertical axis), which is the average accuracy of a single target. In order to compare the performance of all target categories, the average AP (mAP) is used as the evaluation index of accuracy. FPS is the speed at which the evaluation model detects rice diseases, that is, the number of pictures that can be processed per second.

3.4. Results and discussion

In order to improve the accuracy of rice disease detection, a detection model based on YOLOv8n was proposed in this study. Table 2 shows the performance of SCX-YOLOv8n and YOLOv8n in various indicators. According to the results in Table 2, SCX-YOLOv8n is superior to the original model in Precision, the mAP is increased by 3.0%. The weight file of SXC-YOLOv8n model is 6.13MB, which is 3.7 percentage points larger than that of YOLOv8n model, and the parameter quantity is increased by 3.1 percentage points. Although the improved structure has increased the calculation amount, the detection accuracy of rice leaf diseases has been obviously improved. Although the detection speed of SXC-YOLOv8n is 6 frames per second less than that of YOLOv8n, the detection speed is still 200 frames per second, which meets the requirements of rapid and efficient detection of rice diseases.

Table 2. Comparison of Results before and after Yolo V8 n Improvement

model	mAP@0.5/%	Weight/MB	FPS
YOLOv8n	86.3	5.91	223
SCX-YOLOv8n	89.3	6.13	217

In the training process, the convergence process of each evaluation index of YOLOv8n and SCX-YOLOv8n models is shown in Figures A and B in Figure 8. By comparing Figures A and B, we can see that SCX-YOLOv8n model is stronger than SCX-YOLOv8n in both convergence speed and convergence ability. Among them, SCX-YOLOv8n showed a fast convergence speed at the initial stage of training, which means that the model has made rapid progress in learning tasks. In the 15th Epoch, the model has achieved ideal results, which shows that the model can effectively learn the characteristics and patterns of data in a short time. In the subsequent training process, the SCX-YOLOv8n model continued to improve, and finally reached the optimal value of 89.3 in the 232nd Epoch. This shows that the model has achieved a higher performance level after continuous training and optimization, and further improved the accuracy and efficiency in the target detection task.

In addition to the loss function diagram, the comparison diagram of P-R curves detected by SCX-YOLOv8n and YOLOv8n models on the rice leaf disease data set is shown in Figure 9:

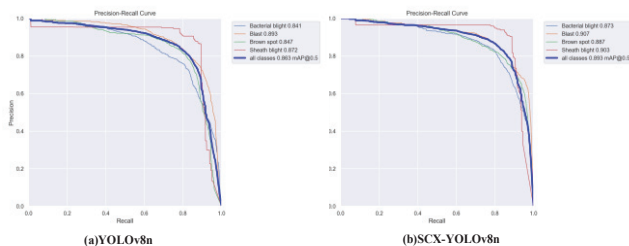


Figure 9. P-R curve

As can be seen from Figure 9, SCX-YOLOv8n shows a significant improvement in the detection of each category compared with YOLOv8n, and the area enclosed by the curve is larger, which means that SCX-YOLOv8n model is more accurate and accurate in the task of rice disease detection, and can better identify disease types and better locate rice diseases, thus improving the yield and reducing losses.

In order to compare the detection effects of the improved algorithm and the improved algorithm more intuitively, the comparison and visualization of the detection images before and after the improvement will be made, as shown in Figure 10.

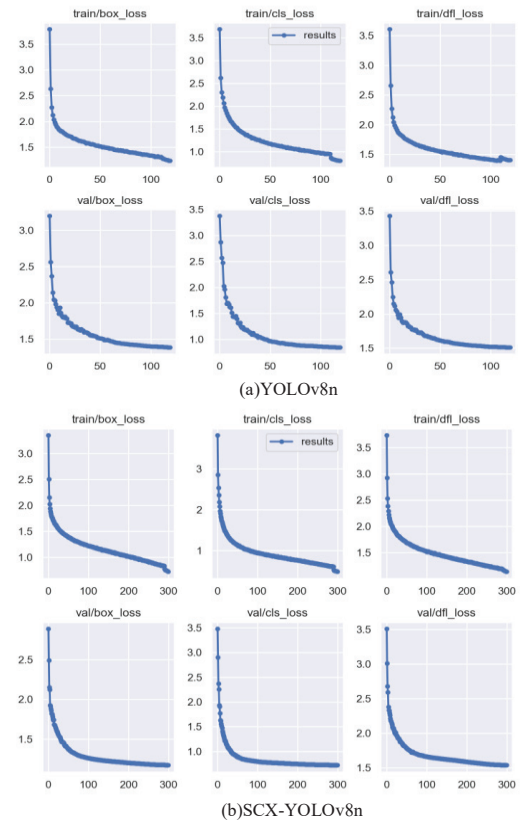


Figure 8. Types of rice diseases

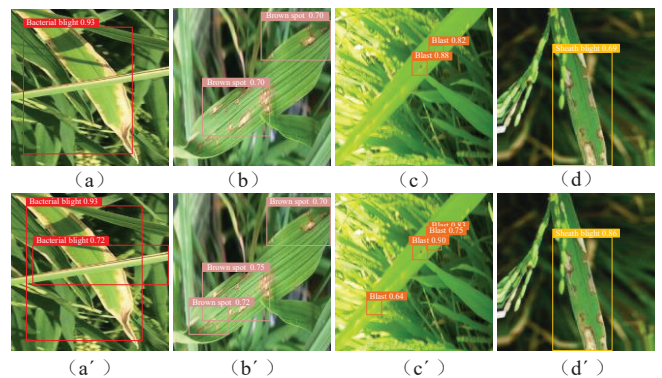


Figure 10. Detection contrast diagram

As shown in Figure 10, in which A, B, C and D are the model detection charts before improvement, and A', B', C' and D' are the improved model detection charts. From Figures A, B, C and d(a', B', C', D'), it can be seen that the overall confidence of the improved model in detection is improved, and the small target is improved.

4. Conclusion

In this paper, a rice leaf disease detection model combining SPPG structure, CA attention mechanism and XSepConv structure with different expansion rates is proposed. Through training and optimization on a large-scale rice leaf disease data set, the excellent performance of SCX-YOLOv8n in rice leaf disease detection task is demonstrated. The experimental results show that the mAP of the model reaches 89.3%, which can meet the requirements of detecting leaf diseases of rice to a certain extent, and also has the characteristics of light weight. The model is only 6.5MB and can be easily integrated into embedded devices. The future work of this research will focus on integrating the proposed model into mobile devices or drones, so that marginal farmers can use smart phones to detect leaf diseases in real time, or observe large-scale rice fields on drones. In addition, the study will explore a higher mAP model to further improve the accuracy of the detection system and provide strong support for disease diagnosis and prevention in the agricultural field.

References

- [1] MISHRA R, JOSHI R K, ZHAO K. Genome editing in rice: recent advances, challenges, and future implications[J]. *Frontiers in Plant Science*, 2018,9: 1361.
- [2] JIA L, WANG T, CHEN Y, et al. MobileNet-CA-YOLO: An Improved YOLOv7 Based on the MobileNetV3 and Attention Mechanism for Rice Pests and Diseases Detection[J]. *Agriculture*, 2023,13(7): 1285.
- [3] MATIN M M H, KHATUN A, MOAZZAM M G, et al. An efficient disease detection technique of rice leaf using AlexNet[J]. *Journal of Computer and Communications*, 2020,8(12): 49-57.
- [4] ZHOU G, ZHANG W, CHEN A, et al. Rapid detection of rice disease based on FCM-KM and faster R-CNN fusion[J]. *IEEE access*, 2019,7: 143190-143206.
- [5] CHEN J, CHEN W, ZEB A, et al. Lightweight inception networks for the recognition and detection of rice plant diseases[J]. *IEEE Sensors Journal*, 2022,22(14): 14628-14638.
- [6] YUMANG A N, VILLAVERDE J F, MC HENRY C T, et al. Bacterial Leaf Blight Identification of Rice Fields Using Tiny YOLOv3: 2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)[C]: IEEE, 2022.
- [7] MASYKUR F, ADI K, NURHAYATI O D. Approach and Analysis of Yolov4 Algorithm for Rice Diseases Detection at Different Drone Image Acquisition Distances[J]. 2023.
- [8] JHATIAL M J, SHAIKH R A, SHAIKH N A, et al. Deep Learning-Based Rice Leaf Diseases Detection Using Yolov5[J]. *Sukkur IBA Journal of Computing and Mathematical Sciences*, 2022,6(1): 49-61.
- [9] WU Z, XUE R, LI H. Real-time video fire detection via modified YOLOv5 network model[J]. *Fire Technology*, 2022,58(4): 2377-2403.
- [10] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition[C], 2021.
- [11] CHEN J, LU Z, XUE J, et al. XSepConv: Extremely separated convolution[J]. *arXiv preprint arXiv:2002.12046*, 2020.
- [12] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2015,37(9): 1904-1916.
- [13] LIN T, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection: Proceedings of the IEEE conference on computer vision and pattern recognition[C], 2017.
- [14] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation: Proceedings of the IEEE conference on computer vision and pattern recognition[C], 2018.
- [15] HENDRYCKS D, GIMPEL K. Gaussian error linear units (gelus)[J]. *arXiv preprint arXiv:1606.08415*, 2016.
- [16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks: Proceedings of the IEEE conference on computer vision and pattern recognition[C], 2018.
- [17] CHOLLET F. Xception: Deep learning with depthwise separable convolutions: Proceedings of the IEEE conference on computer vision and pattern recognition[C], 2017.