



Enhanced Detection of Soybean Leaf Diseases Using an Improved Yolov5 Model

Shiqin Peng¹, Guiqing Xi^{1*}, Yongshun Wei¹, Ling Yu²

¹ School of Information and Electrical Engineering, Heilongjiang Bayi Agricultural University, 163319 Daqing, China

² BGP INC., China national petroleum corporation, 163319 Daqing, China

* Correspondence: Guiqing Xi (xiguiqing@byau.edu.cn)

Received: 02-09-2024

Revised: 03-13-2024

Accepted: 03-22-2024

Citation: S. Q. Peng, G. Q. Xi, Y. S. Wei, and L. Yu, "Enhanced detection of soybean leaf diseases using an improved Yolov5 model," *Int J. Knowl. Innov Stud.*, vol. 2, no. 1, pp. 45–56, 2024. <https://doi.org/10.56578/ijkis020105>.



© 2024 by the author(s). Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: To facilitate early intervention and control efforts, this study proposes a soybean leaf disease detection method based on an improved Yolov5 model. Initially, image preprocessing is applied to two datasets of diseased soybean leaf images. Subsequently, the original Yolov5s network model is modified by replacing the Spatial Pyramid Pooling (SPP) module with a simplified SimSPPF for more efficient and precise feature extraction. The backbone Convolutional Neural Network (CNN) is enhanced with the Bottleneck transformer (BotNet) self-attention mechanism to accelerate detection speed. The Complete Intersection over Union (CIoU) loss function is replaced by EIoU-Loss to increase the model's inference speed, and Enhanced Intersection over Union (EIoU)-Non-Maximum Suppression (NMS) is used instead of traditional NMS to optimize the handling of prediction boxes. Experimental results demonstrate that the modified Yolov5s model increases the mean Average Precision (mAP) value by 4.5% compared to the original Yolov5 network model for the detection and identification of soybean leaf diseases. Therefore, the proposed method effectively detects and identifies soybean leaf diseases and can be validated for practicality in actual production environments.

Keywords: Soybean leaf disease; Improved Yolov5; Bottleneck transformer (BotNet); Disease detection

1 Introduction

Soybean is one of the important grain crops in China. During the growth of soybeans, the occurrence of diseases can cause the plants to weaken and become infected, affecting the yield and quality of soybeans. Therefore, it is extremely important to quickly detect and carry out early prevention and control tasks to avoid the economic losses caused by diseases in soybean cultivation each year [1].

Traditional detection methods mainly fall into two categories. One is manual detection and identification, which requires a large amount of manpower, material resources, and time costs, and the detection results are susceptible to human subjective consciousness, leading to misjudgments. The other is based on image-based machine learning methods. Shrivastava and Hoodal [2] proposed a method based on digital image processing technology to detect and classify soybean leaf blight and gray spot disease, with identification accuracies of 70% and 80%, respectively. This method extracts the shape feature vectors of leaf images and uses the K-Nearest Neighbors (KNN) classifier for detection and classification. However, the recognition accuracy of this method is not enough, and the extraction of image shape feature vectors is relatively simple, which cannot distinguish leaves with complex backgrounds and deformation features. Araujo and Peixoto [3] proposed a digital image processing technique combining color moments, Local Binary Patterns (LBP), and Bag of Visual Words (BoVW) models, using the extracted image features as inputs for a Support Vector Machine (SVM) to achieve disease classification. However, the recognition rate of this method only reached 75.8%, which is not sufficient for application in real environments. Traditional machine learning requires a series of complex data processing steps, and generally uses simpler function forms, lacking the expressive power of complex models, leading to overfitting and low recognition accuracy in real environment disease detection.

Currently, researchers both domestic and international mainly focus on deep learning for the detection and identification of soybean diseases. For example, Li et al. [4] proposed combining the feature pyramid model with the

Faster R-CNN model, which achieved an average precision mean of 82.48% for the detection of five types of apple leaf diseases. However, this method is not accurate enough for disease detection and the model detection has certain biases. He et al. [5] used an improved Yolov5 model based on weighted bidirectional feature fusion technology to detect pests in economic forests, with an average precision mean reaching 92.3%. However, the complex background of the dataset limits the extraction of feature targets in this method.

This paper focuses on whether soybean disease detection can achieve high accuracy and be applied to actual agricultural production environments. It proposes to improve the SPP structure based on the original Yolov5s network model, enhance the model's data feature extraction capabilities, make the model training more efficient, improve the CNN architecture in the backbone network to further enhance the model's detection accuracy, replace the CIoU loss function, improve NMS, and improve the detection of occluded targets. This study investigates the improved Yolov5 model's detection and identification rates for two types of soybean leaf diseases, aiming to improve the accuracy of soybean disease detection and various identification schemes.

2 Yolov5 Network Model and Improvements

2.1 Yolov5 Network Structure

Yolov5 is a one-stage object detection network, which can be further subdivided into several different versions based on the size of the algorithm model and computational complexity: Yolov5s, Yolov5m, Yolov5l, Yolov5x, and Yolov5n. As the depth and width of the network model increase, the model's detection accuracy further improves, but at the cost of slower detection speeds. Therefore, this paper chooses the Yolov5s model, which has lower model complexity. It better meets the real-time requirements of this study, consuming less computing power to maximize recognition speed [5–9].

The Yolov5s model structure primarily consists of the Input, Backbone, Neck, and Prediction segments. The Input part uses the Mosaic data augmentation method, which randomly scales, crops, redistributes, and stitches the input data, adding many small targets and enhancing the robustness of the trained model. The Backbone is the feature extraction part of the Yolov5 network, where the feature extraction capability directly affects the entire network's performance, it includes the Focus, Conv, C3, and SPP modules. The Focus module slices the image, transferring the image's width (W) and height (H) information to the channel space, allowing for 2x downsampling without losing any information. The Conv module performs convolution, batch normalization (BN), and activation function operations on the input feature map. The C3 module is used for part of the feature map extraction, where one part goes through block calculations, and another part through a convolutional shortcut, both parts are then combined using concat. The SPP module is designed to fuse feature maps of different resolutions by reducing the input channels by half with a standard convolutional module, followed by pooling operations with kernel sizes of 5, 9, and 13, and then concatenating the three max pooling results with the unpooled data, finally doubling the channel number [10–15]. The Neck is composed of FPN+PAN; the FPN structure downsamples feature maps of different resolutions to obtain a set of feature maps with high semantic content, then the PAN upsamples these feature maps, enlarging their dimensions to detect small targets with large-sized feature maps and large targets with small-sized ones, merging high and low-level feature information to output prediction feature maps. The Prediction part mainly uses the loss function (CIoU) Loss and NMS for post-processing and target prediction box handling [16–18].

2.2 Improvements to the Yolov5s Model

2.2.1 SimSPPF structure

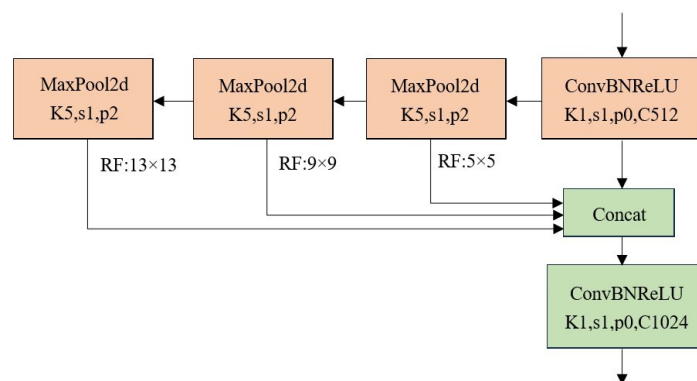


Figure 1. SimSPPF structure

The YOLOv5 model uses a SPP structure, and subsequently introduced the SPPF, which replaces the parallel Maxpool of the original SPP with a more efficient, faster serial Maxpool. The SimSPPF further builds on this by replacing the SiLU activation function with ReLU, and uses different sized pooling kernels across various scales to enhance detector performance. In the feature parsing process, the nodes in SimSPPF are divided into different layers by scale, with each layer's node scale being twice that of the previous layer. In each layer, pooling technology is used to reuse already allocated nodes, thus reducing memory usage. This method decreases spatial occupancy and improves parsing performance [19, 20]. The structure of SimSPPF is shown in Figure 1.

2.2.2 BotNet structure

In YOLOv5, the backbone feature extraction network is a CNN network, which has translational invariance and locality but lacks the capability for global and long-distance modeling. BotNet is a simple yet powerful backbone that, unlike ResNet50, uses Multi-Head Self-Attention (MHSA) to replace the 3×3 spatial convolution in the Bottleneck [21–24]. The BotNet structure is shown in Figure 2.

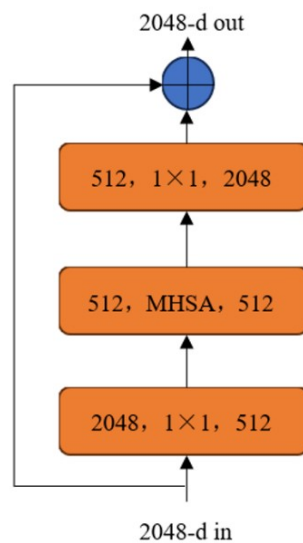


Figure 2. BotNet structure diagram

Similar to traditional attention mechanisms, MHSA can focus more on key information in the input. It runs multiple Self-Attention layers in parallel and synthesizes the learning outcomes of each “head”, capturing information from the input sequence across different subspaces, thereby enhancing the model’s expressive capacity. The structure of MHSA is shown in Figure 3.

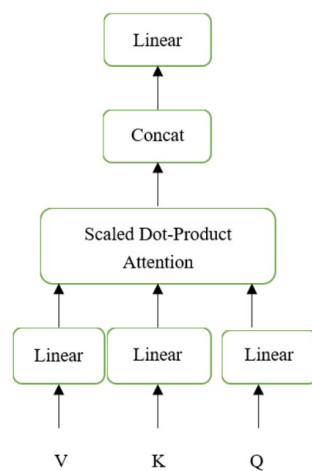


Figure 3. MHSA structure

MHSA splits the input's query, key, and value matrices into H heads, computes attention independently within each head, then concatenates these heads' outputs and applies a linear transformation. This enables simultaneous capture and integration of multiple interaction information across different representational subspaces. The specific formulas are as follows [25, 26]:

$$Head_i = \text{Attention}(Q_i, K_i, V_i) = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) * V, i \in [1, H]$$

$$MHSA(Q, K, V) = \text{Concat}(head_1, head_2, \dots, head_n) W^o$$

In Self-Attention, Q, K, and V are matrices obtained from the same input through three different linear transformations, where QK^T is a similarity matrix. Applying softmax to this matrix row-wise yields the Attention matrix. The output matrices $Head_i$ are concatenated along the feature dimension (dim) to form a new matrix, which is then multiplied by the matrix W^o to produce the output $MHSA(Q, K, V)$ [27, 28].

2.2.3 EIou loss function

This paper proposes an improved loss function to enhance the model's recognition accuracy. The original Yolov5 model uses the CIoU loss function during training, which should include coverage area, distance between centers, and aspect ratio of the detection data. The CIoU loss, building on the Distance Intersection over Union (DIoU) loss, adds a measure of the aspect ratio v between the predicted box and the ground truth (GT) box, which can accelerate the regression speed of the prediction box to some extent. However, there are still significant issues, as the model detection can sometimes be blurry. Based on the gradient formulas for predicted box width (w) and height (h), it is evident that when one value increases, the other must decrease; they cannot increase or decrease simultaneously. To address this, EIou proposes direct penalties on the predictions of w and h , where C_w^2 and C_h^2 are the width and height of the smallest enclosing rectangle around the prediction box and GT box. The calculation formula for EIou is as follows [29]:

$$L_{EIou} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(\omega, \omega^{gt})}{C_w^2} + \frac{\rho^2(h, h^{gt})}{C_h^2}$$

Considering the issue of sample imbalance in bounding box regression tasks, EIou is combined with Focal Loss. From the perspective of gradients, this approach separates high-quality anchor boxes from low-quality ones, i.e., reducing the optimization contribution of numerous anchor boxes that overlap less with the target box, focusing the regression process on high-quality anchor boxes. The calculation formula for EIou Loss is as follows [30]:

$$L_{EIouLoss} = IoU^\gamma * L_{EIou}$$

2.2.4 NMS

In recent years, common object detection algorithms (such as RCNN, SPPNet, Faster-RCNN, etc.) typically identify many potential object bounding boxes from a single image, assigning each a probability of belonging to a certain category [31, 32]. NMS filters out the boxes within a certain area that have the highest score for the same category. Through iterative processing, it continually uses the highest scoring box to perform IoU operations with other boxes, filtering out those with high IoU values to retain the best result. In Yolov5, NMS only considers the overlap between the predicted boxes and true boxes and does not account for distances between centers or aspect ratios. Therefore, this paper proposes EIou-NMS, which considers the distance between the centers of two boxes, resulting in a model that performs better with EIou-NMS. The calculation formula for EIou-NMS is as follows [33]:

$$S_i = \begin{cases} S_i, IoU - R_{EIou}(M, B_i) < \varepsilon \\ 0, IoU - R_{EIou}(M, B_i) \geq \varepsilon \end{cases}$$

2.3 Improved Network Model

This study proposes four improvements based on the Yolov5s model. First, the SPP part of the backbone network is improved by introducing SimSPPF to replace the original SPP layer, which increases the efficiency of model training. Next, the BotNet self-attention mechanism is introduced, enabling the model to locate and identify disease target features more accurately [34]. Lastly, the EIou Loss function and the NMS (EIou-NMS) are improved, enhancing the model's prediction accuracy for similar categories. These improvements have enhanced the overall recognition rate of the model. The structure of the improved Yolov5s network model is shown in Figure 4.

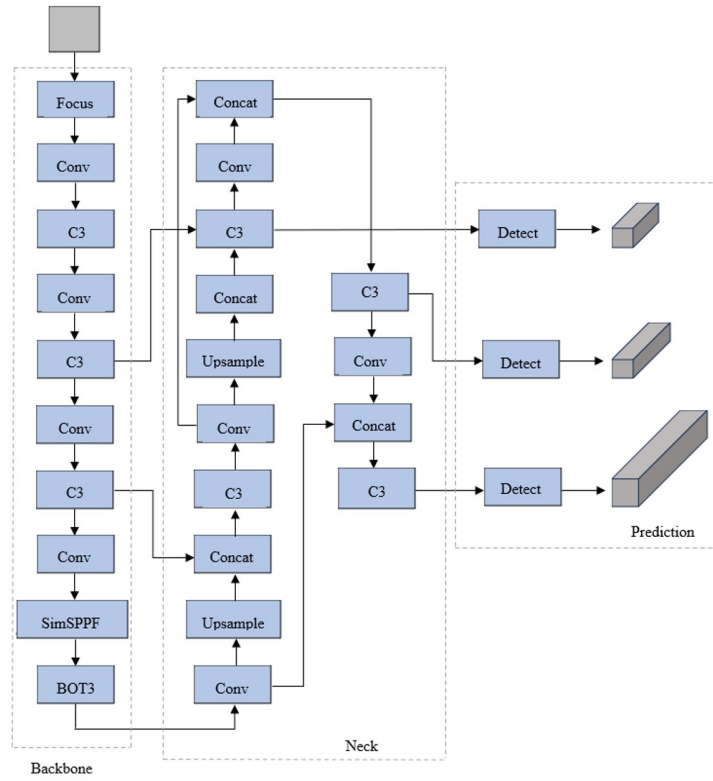


Figure 4. Improved Yolov5s network model structure

3 Experiment

3.1 Dataset Construction

This study focuses on two soybean diseases: Bacterial Spot disease and Brown Spot disease. The dataset was constructed using two methods: First, by collecting images of soybean leaf diseases in different environments in the field using a smartphone; second, through web scraping, Google searches, and various open-source websites to gather images of soybean leaf diseases. The images collected have complex backgrounds, matching real-world application conditions. The characteristics of the disease images are shown in Figure 5.

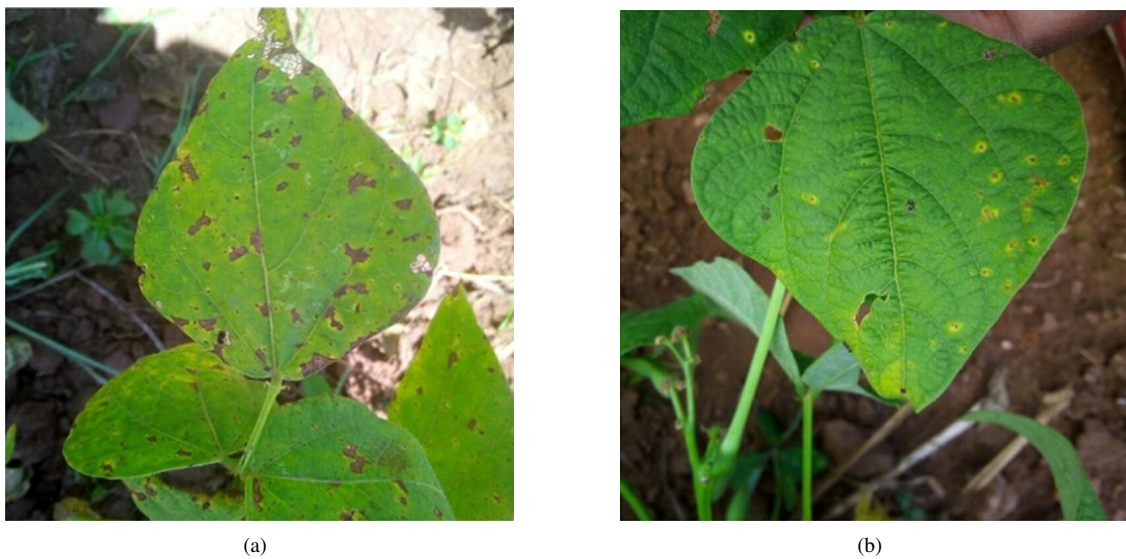


Figure 5. Soybean disease characteristics: (a) Brown spot disease; (b) Bacterial spot disease

For this experiment, over 600 images of soybean leaf diseases were collected. To avoid image redundancy, 600

images were manually selected. Due to the limited number of original disease images, which could not effectively train the network model, the dataset was augmented to five times the number of original images to enhance model stability and reduce overfitting. The augmentation techniques used included adding Gaussian noise, rotating (at 90° and 180°), mirroring, and adjusting brightness. A total of 3000 effective dataset images were selected, and examples of the augmented images are shown in Figure 6.

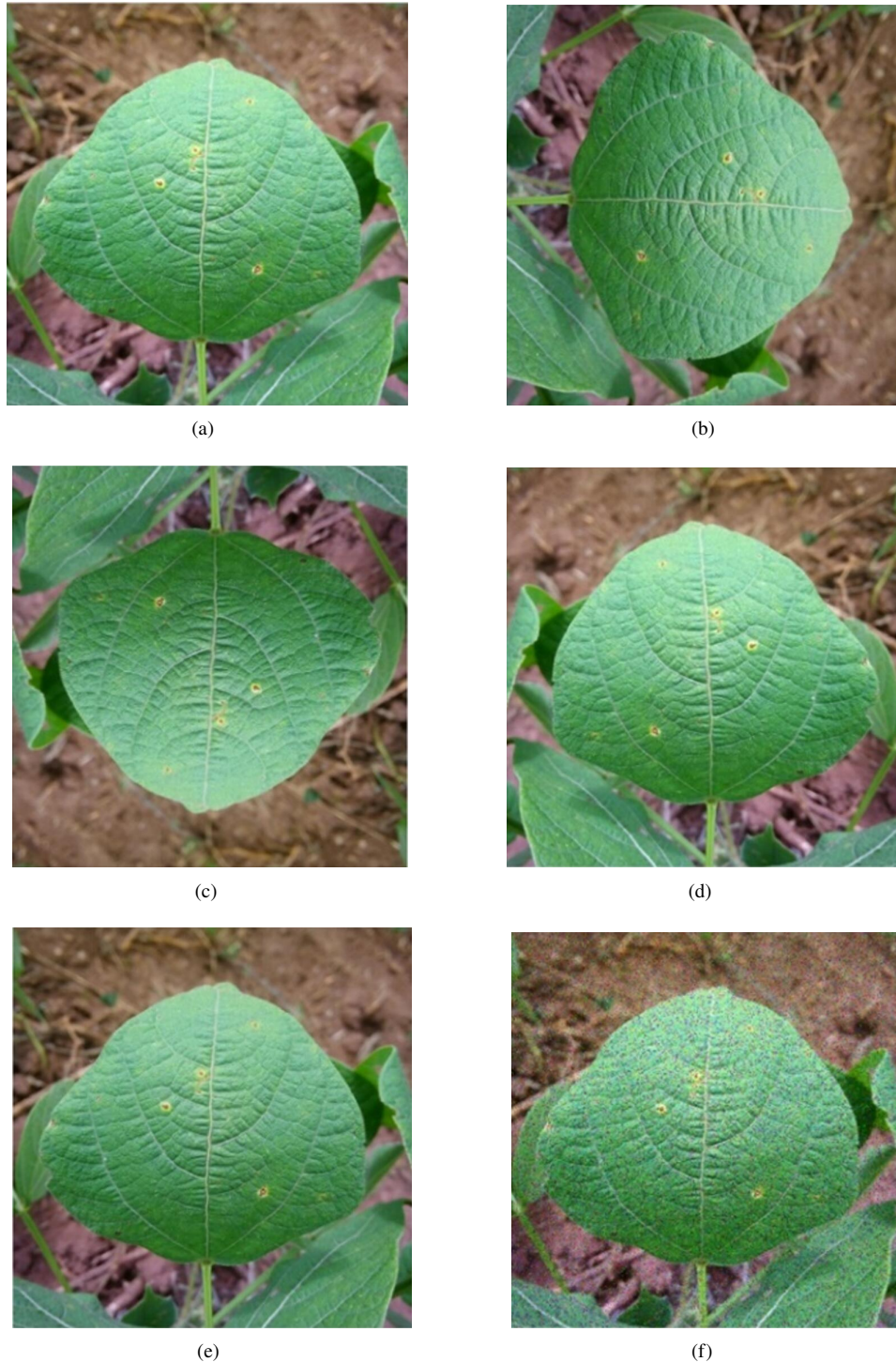


Figure 6. Augmented images: (a) Original image; (b) Rotated 90°; (c) Rotated 180°; (d) Mirrored; (e) Adjusted brightness; (f) Gaussian noise

The dataset was randomly split into a training set of 2100 images, a test set of 600 images, and a validation set of 300 images, following a 7:2:1 ratio. The Labelimg software was used to manually annotate the two types of soybean leaf diseases in the dataset to obtain the coordinates and dimensions of the disease spots on each image, with the

annotation information saved into TXT files. An example of image annotation using Labelimg is shown in Figure 7.

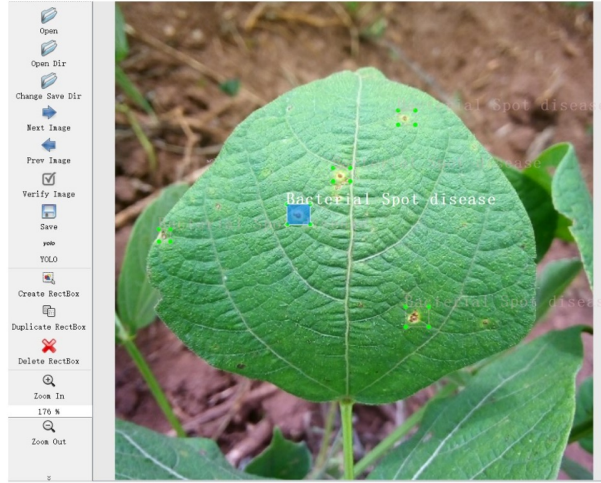


Figure 7. Labelimg annotation

3.2 Experimental Setup

3.2.1 Experimental environment

All experiments were conducted under the YOLOv5s deep learning framework for training and testing the network model. The hardware configuration of the experimental server included: an Intel(R) Core(TM) i5-10400F CPU @ 2.90GHz, NVIDIA GeForce RTX 2060 SUPER graphics card, and a computer with 16GB of memory running on a Windows 10 system. The software environment included Pycharm + Python 3.8, Conda 23.1.0. Images were input at 640×640 pixels, with a batch size of 32, undergoing 300 Epochs, and the best model was saved in the logs.

3.2.2 Evaluation metrics

To evaluate the performance of the target detection algorithm of the model, two metrics are commonly used: recall and precision. Both metrics, precision (p) and recall (r), simply judge the model's quality from one aspect and range between 0 and 1, where closer to 1 indicates better performance and closer to 0 indicates poorer performance. To comprehensively evaluate the target detection performance, mAP is generally used to further assess the model's quality. By setting different confidence threshold levels, p and r values calculated at different thresholds can be obtained. Generally, p and r values are inversely related. Each target in the target detection model can have an AP value calculated, and averaging all AP values yields the mAP value of the model. The training mAP of the improved YOLOv5 model is shown in Figure 8.

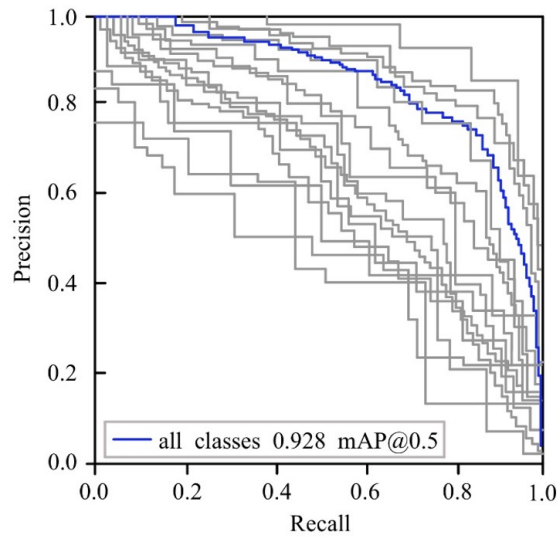


Figure 8. mAP curve

3.3 Comparative Experiments of the Improved Model

3.3.1 Comparison of adding BotNet to the backbone network

To improve the model's accuracy in detecting disease characteristics, this study explored adding the BotNet self-attention mechanism to the backbone network of the Yolov5s. The experiment involved replacing the last C3 module in the backbone network with the self-attention mechanism BotNet (BOT3 module), which yielded the best model recognition performance. Four comparative experiments were designed, adding currently popular attention mechanisms such as CA, SE, and CBAM under the same basic network and experimental data conditions. The comparative results are shown in Table 1.

Table 1. Comparison of different attention mechanisms

Model Scheme	$r(\%)$	$p(\%)$	$mAP(\%)$
Yolov5s + CA	88.5	90.2	91
Yolov5s + SE	88.2	90	90.9
Yolov5s + CBAM	86.7	88.6	89.8
Yolov5s + BOT3	88.4	90.3	91.9

Analysis from Table 1 indicates that adding BOT3 to the original Yolov5s network provides the best improvement in recall, precision, and mAP. Compared to adding the CA attention mechanism, mAP improved by 0.9%; compared to SE, it improved by 1.9%; and compared to CBAM, it improved by 2.1%. This demonstrates that adding the BotNet self-attention mechanism can better identify disease characteristics, achieving a higher disease detection rate.

3.3.2 Ablation study

To further verify the effectiveness of the proposed improvements, an ablation study was conducted by adding only one improvement at a time to the model while keeping training parameters and the dataset the same. The results are shown in Table 2.

Table 2. Comparison of ablation study results for the improved model

Model Scheme	$r(\%)$	$p(\%)$	$mAP(\%)$
Yolov5s	84.5	87.7	88.3
Yolov5s+SimSPPF	85.8	88.9	89.8
Yolov5s+ SimSPPF+ BOT3	87	90.3	91.9
Yolov5s+ SimSPPF+ BOT3+ EIoU-Loss	87.2	90.5	92.4
Yolov5s+ SimSPPF+ BOT3+ EIoU-Loss +EIoU-NMS	87.9	90.9	92.8

Analysis from Table 2 shows that compared to the original Yolov5s algorithm, the improved Yolov5s model has a 3.4% increase in recall (r), a 3.2% increase in precision (p), and a 4.5% increase in mAP. The experimental results indicate that replacing the SimSPPF module, adding the BotNet attention mechanism, improving the EIoU-Loss function, and the EIoU-NMS have made the improved Yolov5s network model perform better in detecting and identifying the two types of soybean leaf diseases.

3.3.3 Comparison of different network models

To evaluate the superiority of the improved Yolov5s network model proposed in this study, popular target detection networks such as Faster R-CNN, Yolov4, and MobileNetV2 were selected for comparative experiments. The results are shown in Table 3.

Table 3. Comparison of different network models

Model Scheme	$r(\%)$	$p(\%)$	$mAP(\%)$
The Proposed Improved Model	87.9	90.9	92.8
Faster R-CNN	77	80.3	82.5
Yolov4	82.3	85.8	87.2
MobileNetV2	85.1	88.5	90.6

As can be seen from Table 3, in terms of the mAP evaluation metric, the improved model shows a 10.3% increase compared to the two-stage target detection algorithm Faster R-CNN, a 5.6% increase compared to the Yolov4 network, and a 2.2% increase compared to the lightweight network MobileNetV2. In terms of recall (r) and average precision (p), it also shows good improvement compared to other network models, indicating that the improved model has superior detection performance.

3.4 Model Visualization and Analysis

In this study, the expanded dataset was imported into the improved Yolov5s model for training, with the training labels set as Bacterial Spot disease and Bean Rust. The model first identifies the type of disease, and each identification result provides a confidence score for the category. The best weight results generated during the training process are shown in Figure 9.



Figure 9. Best weight results during model training

The system interface detects and identifies images from the validation set. First, an image is selected for recognition, and after recognition, the type of disease and confidence score are displayed. The average recognition speed for a single image is 0.09 seconds. The system’s recognition results are shown in Figure 10.

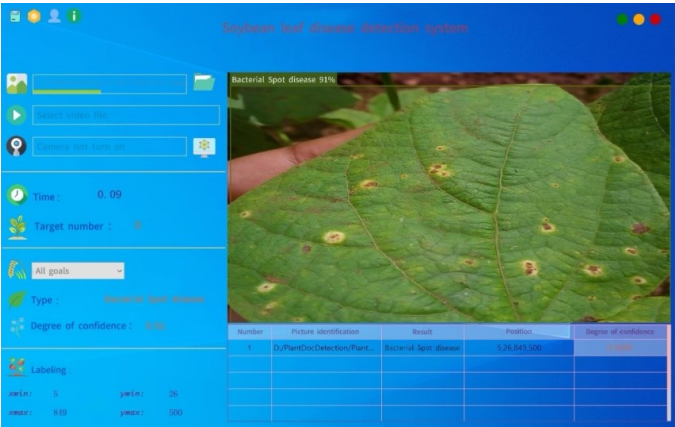


Figure 10. Soybean disease detection system recognition results

4 Conclusion and Future Work

This paper proposes an improved Yolov5s model for the detection and identification of soybean leaf diseases. The dataset was expanded through data augmentation, and the Yolov5s model was enhanced by using a superior SimSPPF structure, reducing the loss of dataset feature information. The addition of the BotNet structure allows the network to better learn the features of leaf diseases, enhancing the network model's precision in extracting target features. Improvements to the loss function and NMS further optimize the model's detection and identification rates. Final experimental results show that the improved network model has generally increased recall, precision, and mAP by 3.4%, 3.2%, and 4.5%, respectively, compared to the original Yolov5s model. Therefore, the model effectively accomplishes the task of detecting soybean leaf diseases, and the disease detection system studied in this paper has practical reference value for actual agricultural applications. Future research will focus on developing lightweight models and expanding the types of soybean leaf diseases to achieve faster model detection rates and more comprehensive disease system detection.

Data Availability

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

Acknowledgements

This research was supported by the “Three Longitudinal” Foundation Cultivation Plan of Heilongjiang Bayi Agricultural University, a provincial university in Heilongjiang Province (ZRCPY202016).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] Z. Q. Shang, D. F. Yang, and Z. P. Ma, “Automatic identification of soybean leaf diseases based on UAV image and deep convolution neural network,” *Soybean Sci.*, vol. 40, no. 5, pp. 662–668, 2021. <https://doi.org/10.11861/j.issn.1000-9841.2021.05.0662>
- [2] S. Shrivastava and D. S. Hooda, “Automatic brown spot and frog eye detection from the image captured in the field,” *Am. J. Intell. Syst.*, vol. 4, no. 4, pp. 131–134, 2014. <https://doi.org/10.5923/j.ajis.20140404.01>
- [3] J. M. M. Araujo and Z. M. A. Peixoto, “A new proposal for automatic identification of multiple soybean diseases,” *Comput. Electron. Agric.*, vol. 167, p. 105060, 2019. <https://doi.org/10.1016/j.compag.2019.105060>
- [4] X. R. Li, S. Q. Li, and B. Liu, “Apple leaf diseases detection model based on improved faster R-CNN,” *Comput. Eng.*, vol. 47, no. 11, pp. 298–304, 2021. <https://doi.org/10.19678/j.issn.1000-3428.0059290>
- [5] Y. He, D. Chen, and L. Peng, “Research on object detection algorithm of economic forestry pests based on improved YOLOv5,” *J. Chin. Agric. Mech.*, vol. 43, no. 4, pp. 106–115, 2022. <http://zgjnjhxb.niam.com.cn/CN/Y2022/V43/I4/106>
- [6] J. Wang, H. Li, Z. Ma, J. Hao, and W. Liu, “Analysis and countermeasures of mechanization development of domestic and foreign facility vegetables,” *J. Chin. Agric. Mech.*, vol. 44, no. 1, pp. 124–130, 2023. <http://zgjnjhxb.niam.com.cn/CN/Y2023/V44/I1/124>
- [7] L. Y. Shi, S. Y. Tong, T. Wu, Y. Feng, and H. H. Liu, “Farmland pest image recognition based on improved YOLOv5 attention model,” *Mod. Inf. Technol.*, vol. 7, no. 10, pp. 70–73, 2023. <https://doi.org/10.19850/j.cnki.2096-4706.2023.10.018>
- [8] L. Lu and X. Yu, “Recognition and classification of deep learning in soybean leaf image data management,” *J. Libr. Inf. Sci. Agric.*, vol. 35, no. 2, pp. 87–94, 2023. <https://doi.org/10.13998/j.cnki.issn1002-1248.21-0188>
- [9] H. X. Zuo, Q. C. Huang, J. H. Yang, F. J. Meng, S. E. Li, and L. Li, “In situ identification method of maize stalk width based on binocular vision and improved YOLOv8,” *Smart Agric.*, vol. 5, no. 3, pp. 86–95, 2023. <https://doi.org/10.12133/j.smartag.SA202309004>
- [10] J. K. Su, X. H. Duan, and Z. B. Ye, “Research on corn disease detection based on improved YOLOv5 algorithm,” *J. Front. Comput. Sci. Tech.*, vol. 17, no. 4, pp. 933–941, 2023. <https://doi.org/10.3778/j.issn.1673-9418.2210066>
- [11] Z. Y. Zhang, M. Y. Luo, S. X. Guo, G. Liu, S. P. Li, and Y. Zhang, “Cherry fruit detection method in natural scene based on improved YOLOv5,” *Trans. Chin. Soc. Agric. Mach.*, vol. 51, no. 13, pp. 221–226, 2023. <https://doi.org/10.3969/j.issn.0517-6611.2023.13.049>
- [12] A. Srinivas, T. Lin, N. Parmar, J. Shlens, P. Abbeel, and A. Vaswani, “Bottleneck transformers for visual recognition,” in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2021, pp. 16 514–16 524. <https://doi.org/10.1109/CVPR46437.2021.01625>

- [13] B. Richey and M. V. Shirvaikar, "Deep learning based real-time detection of northern corn leaf blight crop disease using YOLOv4," in *Real-Time Image Processing and Deep Learning*, 2021, pp. 39–45. <https://doi.org/10.1117/12.2587892>
- [14] B. Yang, J. He, and L. Zhang, "Identification and detection of rice leaf diseases by YOLOv5 neural network based on improved SPP-x," *J. Chin. Agric. Mech.*, vol. 44, no. 9, pp. 190–197, 2023. <http://zgjnxb.niam.com.cn/CN/Y2023/V44/I9/190>
- [15] S. F. Zhou, X. L. Xiao, Z. Y. Liu, and L. Lu, "Improved apple leaf disease detection based on YOLOv5s," *Jiangsu Agric. Sci.*, vol. 51, no. 22, pp. 212–220, 2023.
- [16] Y. L. Zeng, Y. T. He, Y. Lin, J. J. Fei, Q. Li, and Y. Yang, "A detection method for apple leaf diseases based on BCE-YOLOv5," *Jiangsu Agric. Sci.*, vol. 51, no. 15, pp. 155–163, 2023. <https://doi.org/10.15889/j.issn.1002-1302.2023.15.022>
- [17] W. Y. Fang, Y. G. Guo, F. C. Guan, W. Zhang, Q. Q. Liu, S. W. Wang, Z. C. Zhang, and H. R. Yu, "Identification of wormholes in soybean leaves based on improved YOLOv5s algorithm," *J. Hunan Agric. Univ. (Nat. Sci. Ed.)*, vol. 49, no. 1, pp. 127–132, 2023. <https://doi.org/10.13331/j.cnki.jhau.2023.01.018>
- [18] H. X. Ma, K. B. Dong, Y. F. Wang, S. Wei, H. W. G., and G. J. P., "Research on a lightweight plant recognition model based on improved YOLOv5s," *J. Agric. Mach.*, vol. 54, no. 8, pp. 267–276, 2023. <https://doi.org/10.6041/j.issn.1000-1298.2023.08.026>
- [19] J. J. Hao, S. Y. Fang, S. Xiao, and M. Li, "Research on improving YOLOv5's leaf black spot detection algorithm," *South. Agric. Mach.*, vol. 53, no. 16, pp. 1–4, 2022. <https://doi.org/10.3969/j.issn.1672-3872.2022.16.001>
- [20] C. Y. Wang, H. M. Liao, Y. Wu, P. Chen, J. Hsieh, and I. H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, 2020, pp. 1571–1580. <https://doi.org/10.1109/CVPRW50498.2020.00203>
- [21] Y. F. Zhang, W. Q. Ren, Z. Zhang, J. Zhen, W. Liang, and T. N. Tan, "Focal and efficient IOU loss for accurate bounding box regression," *ArXiv abs/2101.08158*, 2021. <https://arxiv.org/pdf/2101.08158.pdf>
- [22] W. Zhang, X. Zhao, R. Ding, Z. Zhang, H. Jiang, and D. Wang, "A detection and recognition method for tomato on faster R-CNN algorithm," *J. Shandong Agric. Univ.*, vol. 52, no. 4, pp. 624–630, 2021. <https://doi.org/10.3969/j.issn.1000-2324.2021.04.017>
- [23] D. Guo and L. Ge, "Image recognition method of smart agricultural diseases and insect pests based on transfer learning," *J. Smart Agric.*, vol. 2, no. 14, pp. 7–9, 2022. <https://doi.org/10.20028/j.zhnydk.2022.14.003>
- [24] R. Luo, H. Yin, and W. Liu, "Identification of bergamot pests and diseases using YOLOv5-C algorithm," *J. South China Agric. Univ.*, vol. 44, no. 1, pp. 151–160, 2023. <https://doi.org/10.7671/j.issn.1001-411X.202203012>
- [25] P. Shaw, J. Uszkoreit, and A. Vaswani, "Self-attention with relative position representations," 2018. <https://arxiv.org/pdf/1803.02155.pdf>
- [26] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 7132–7141. <https://doi.org/10.1109/CVPR.2018.00745>
- [27] W. Hu, H. Wang, and Y. Du, "Identification method of tomato diseases and pests based on SE module and resnet," *Agric. Eng.*, vol. 12, no. 9, pp. 33–40, 2022. <https://doi.org/10.19998/j.cnki.2095-1795.2022.09.007>
- [28] J. Cen and J. Liu, "Improved YOLOv5 for rice pest detection," *Inf. Technol. Informatization*, vol. 2023, no. 7, pp. 165–171. <https://doi.org/10.3969/j.issn.1672-9528.2023.07.042>
- [29] S. Y. Chen, H. Y. Zhu, Z. X. Wang, R. Qiao, W. X. Song, and T. Yu, "Research on soybean disease image detection technology based on python," *Soybean Sci.*, vol. 42, no. 3, pp. 360–366, 2023. <https://doi.org/10.11861/j.issn.1000-9841.2023.03.0360>
- [30] M. E. Haque, A. Rahman, and I. Junaid, "Rice leaf disease classification and detection using YOLOv5," *CS.CV*, 2022. <https://doi.org/10.48550/arXiv.2209.01579>
- [31] S. Woo, J. Park, J. Lee, and I. Kweon, "CBAM: Convolutional block attention module," in *Computer Vision – ECCV 2018*, Springer, Cham, 2018, pp. 3–19. https://doi.org/10.1007/978-3-030-01234-2_1
- [32] J. H. Zhang, H. K. Zhao, Y. H. Zhou, J. Y. Jiang, and X. W. Wei, "Soybean leaf spot identification method based on YOLOv5 model," *J. ISS*, vol. 35, no. 4, pp. 163–165, 2023. <https://doi.org/10.3969/j.issn.1003-9767.2023.04.049>
- [33] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE, 2021, pp. 303–338. <https://doi.org/10.1109/CVPR46437.2021.01350>
- [34] Z. Zhang, X. Lu, G. Cao, Y. Yang, L. Jiao, and F. Liu, "ViT-YOLO: Transformer-based YOLO for object detection," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. Montreal,

BC, Canada: IEEE, 2021, pp. 2799–2808. <https://doi.org/10.1109/ICCVW54120.2021.00314>