Crop Recognition Based on Improved YOLOv9

Li Xiang, Jiao Xiangjian, Jia Maotang
School of Intelligence Science and Technology, Xinjiang University, China

Abstract: To overcome challenges associated with crop species identification in unmanned aerial vehicle (UAV) imagery—specifically missed detections of small-sized targets, disturbances caused by intricate environmental elements, and suboptimal computational efficiency of detection models—this study introduces a streamlined detection framework that integrates YOLOv9 with Convolutional Block Attention Module (CBAM) and Spatial Channel Reconstruction Convolution (SCConv). Initial modifications involve embedding CBAM within the YOLOv9 backbone architecture. This enhancement exploits the collaborative functioning of channel-wise and spatial attention mechanisms to amplify the model's sensitivity toward critical localized crop characteristics, such as panicle morphological structures and foliar texture patterns, while concurrently mitigating noise interference from soil surfaces and weed vegetation. Subsequent improvements involve substituting conventional convolutional layers with Spatial Channel Reconstruction Convolution. This substitution capitalizes on the module's adaptive feature reorganization capabilities to achieve parameter reduction without compromising feature representation capacity, thereby substantially enhancing operational efficiency during edge device deployment. Empirical evaluations conducted on a proprietary dataset comprising 1,000 UAV-captured images representing seven distinct crop types demonstrate that the optimized model attains a mean Average Precision (mAP) of 94.5% when applying a 50% Intersection over Union threshold—a performance gain of 2.2 percentage points compared to the baseline YOLOv9 architecture. These results confirm the system's capability to meet real-time recognition demands within complex agricultural landscapes. The proposed methodology presents a cost-effective, high-accuracy monitoring solution for UAV-based precision agriculture applications, holding significant practical value for advancing intelligent management systems in modern agricultural practices.

Keywords:Unmanned Aerial Vehicle; Crop; Yolov9; Attention Module; Spatial Channel Reconstruction

1. Introduction

1.1 Research Background and Significance

With the rapid development of precision agriculture and smart agriculture, drone remote sensing technology has become an important tool in crop monitoring due to its flexibility, efficiency, and low cost[1]. High-resolution images obtained by drones enable agricultural practitioners to perform key tasks such as monitoring crop growth, pest and disease warnings, and yield estimation. However, in complex farmland scenarios, different crop varieties (such as wheat and corn) exhibit high inter-species morphological similarity, dense spatial distribution, and are easily affected by light changes and obstructions, leading to low efficiency in traditional manual identification methods. Meanwhile, automated recognition technologies based on deep learning still face challenges such as missed detection of small targets, feature redundancy, and complex background interference.

In recent years, convolutional neural network (CNN) [2]-based object detection models (such as the YOLO series) have shown potential in the agricultural field [3], but their performance in fine-grained crop variety recognition remains insufficient: First, traditional models have limited perception capabilities for local crop features (such as ear morphology and leaf texture) [4]; Second, background noise in drone images (such as soil and weeds) is highly mixed with target areas, reducing model robustness; Additionally, the computational complexity of existing models is high, making it difficult to deploy in real-time on drone edge devices. Therefore, designing a high-precision, lightweight, and adaptable crop variety recognition model for complex farmland environments has become a critical need for advancing intelligent agricultural management.

1.2 Existing Problems and Challenges

At present, the research on crop variety identification based on UAV image has the following limitations:

1. Insufficient feature extraction: Traditional target detection models (such as YOLOv5/YOLOv8) have limited receptive field [5] in dense crop areas, and it is difficult to distinguish varieties with high inter-class similarity.
2. Background interference sensitivity: The complexity of light change, occlusion and soil background in farmland scenes leads to the increase of model false detection rate.
3. Bottleneck of computing efficiency: The existing model has a large number of parameters and high computing cost, which is difficult to meet the real-time processing requirements of UAV.

Despite existing research attempts to optimize model [7] through attention mechanisms like [6] or lightweight convolutions (such as MobileNetV3), these methods still fall short in adapting to agricultural scenarios: attention mechanisms often focus solely on channel dimensions, overlooking the importance of spatial features [8]; lightweight strategies may excessively compress model capacity, leading to a significant drop in accuracy. Therefore, there is an urgent need for a solution that enhances multi-dimensional features and optimizes efficient computation simultaneously.

1.3 Innovation and Contribution of this Paper

In view of the above problems, this paper proposes a lightweight crop variety recognition model based on improved YOLOv9. By integrating convolution block attention module (CBAM) and spatial and channel reconstruction convolution (SCConv), the balance between accuracy and efficiency is achieved. The specific contributions are as follows:

1. Feature enhancement guided by multi-dimensional attention: CBAM module [9] is embedded in the YOLOv9 backbone network, and the synergistic effect of channel attention and spatial attention is used to enhance the model's ability to locate key areas of crops (such as ears and leaf edges) and suppress background noise interference.
2. Lightweight design of dynamic feature reconstruction: SCConv is used to replace part of the standard convolution layer [10], and the cascade structure of spatial reconstruction unit (SRU) and channel reconstruction unit (CRU) is adopted to adaptively reduce feature redundancy, so as to reduce the calculation amount while maintaining the representation ability of the model.
3. Optimization of agricultural scene adaptability: According to the characteristics of UAV images (such as the density of small targets in low-altitude shooting), data enhancement and targeted data collection are carried out to some extent, so as to improve the robustness of the model to scale change and occlusion [11].

2. Related Work

Target detection technology has gained widespread attention in agricultural scenarios in recent years, especially in the field of drone remote sensing. Researchers have proposed various solutions for tasks such as crop recognition and pest monitoring [12]. This section reviews relevant research progress from three dimensions: the evolution of target detection algorithms, challenges in drone agriculture applications, and data augmentation strategies, and summarizes the technical improvements for this study.

2.1 Development of Target Detection Algorithm

Deep learning-based object detection models are primarily divided into two-stage detectors (such as Faster R-CNN and Mask R-CNN) and single-stage detectors (such as YOLO series and SSD). In agricultural scenarios, single-stage detectors are favored due to their real-time advantages. For example, Gong Xulian et al. proposed a lightweight detection method for small target diseases in apple leaves using YOLOv5s, called [13], to address the complexity of detecting leaf diseases in natural environments, the high difficulty of detecting small target diseases, and the large model parameters that make deployment on mobile devices and embedded systems challenging. Jia Xueying et al. introduced an improved YOLOv7 algorithm, incorporating a CA module into the backbone network of the YOLOv7 model and adding a contextual transformer module at the network head, [14]. This module achieves online monitoring of citrus surface defects with an average precision of 91.1%, demonstrating excellent detection performance. However, existing models still suffer from missed detections and false positives in scenarios involving dense small targets in farmland and similar morphological patterns between classes.

2.2 Challenges of Uav Agriculture Application

Low-altitude UAV images have the characteristics of high resolution and multi-scale coexistence of targets, but they also face the following challenges:

Small target detection: crop seedlings or key parts (such as rice panicles) occupy only a very small pixel area (<50x50 pixels) in the image, and traditional detectors are easily affected by background interference;

Complex light and noise: uneven illumination in farmland environment, leaf reflection and sensor noise (such as Gaussian noise) will reduce the robustness of the model;

Inter-class similarity: Some crops (such as corn and sorghum) are highly similar in morphology, so texture or local features are needed to distinguish them.

In view of the above problems, researchers have proposed multi-scale feature fusion, attention mechanism (such as CBAM, SE-Net) and synthetic data enhancement, but few works systematically combine image characteristics with the need for lightweight models.

2.3 Data Enhancement Strategy

In order to improve the adaptability of the model to complex farmland environment, the following enhancement methods are adopted:

1. Image scaling: Keep the aspect ratio and scale the image to the specified size, or only shrink the image when the image size exceeds the target size, while adjusting the position and size of the corresponding boundary box (bounding boxes).
2. Flip: Randomly flip the image horizontally or vertically, and adjust the position of the boundary box accordingly, so as to ensure that the labeling after flipping is still accurate.
3. Center cropping: The center part of the image is cropped into a square, and the boundary boxes within the new image range are screened and adjusted to remove the targets outside the cropping range.
4. Color transformation: including random changes in brightness, contrast and saturation, as well as the addition of Gaussian noise, salt noise (white pixels) and pepper noise (black pixels), to simulate images under different lighting and noise conditions.

Through the above data enhancement methods, a more complex and suitable detection dataset for farmland environment is obtained to prepare for subsequent experiments

3. SC-YOLOv9 model

This paper uses a self-made drone dataset for crop photography, including seven common types of crops: lettuce, scallions, corn, cabbage, radishes, leeks, and Chinese cabbage. To achieve more precise crop recognition, this paper employs the YOLOv9 model and introduces the SCConv convolution module along with the CBAM attention mechanism to improve the model.

3.1 Scconv Module

Traditional convolution layers suffer from feature redundancy when extracting features, especially when processing crop areas with similar textures in drone images (such as densely planted corn and leeks). Redundant computations significantly increase model complexity. This study replaces the YOLOv9 backbone part of the traditional 3x3 convolution layer with a SCConv convolution module, achieving a balance between lightweight and efficient feature representation through dual spatial and channel reconstruction mechanisms.

The SCConv module consists of a spatial reconfiguration unit (SRU) and a channel reconfiguration unit (CRU). When in use, the feature map first passes through the spatial reconfiguration unit to reduce redundancy in the spatial dimension, then the refined feature map is fed into the channel reconfiguration unit to reduce channel redundancy [15]. The structure of the SCConv is shown in Figure 1.

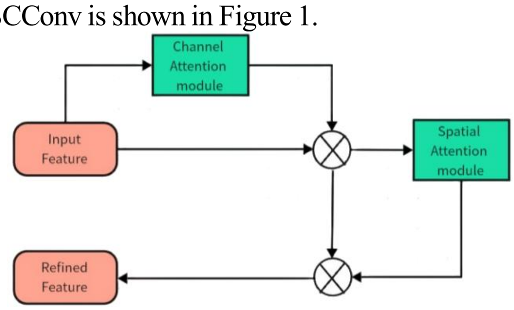

Figure 1. SCConv Module

3.2 CBAM Module

CBAM is a lightweight attention module primarily composed of channel attention (CAM) and spatial attention modules (SAM). The structure of CBAM is shown in Figure 2. After the feature map enters the CBAM module, it first goes through the channel attention module, which compresses the spatial dimension while keeping the channel dimension unchanged, extracting the feature map information of interest. This extracted information is then multiplied with the original feature map to form a weighted feature map. The weighted feature map serves as the input for the spatial attention module, maintaining the spatial dimension while compressing the channel dimension, allowing the network to better focus on the positions of the desired targets. The output of the spatial feature map is then multiplied with the weighted feature map to obtain the final feature map [16]. The structure of the CBAM module is illustrated in Figure 2.

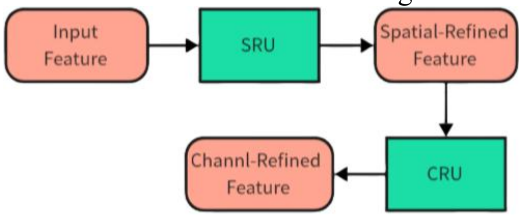

Figure 2. CBAM Module

3.3 SC-YOLOv9 Model

The SC-YOLOv9 model is composed of the YOLOv9 model, CBAM module and SCConv module. The structure of the SC-YOLOv9 model is shown in Figure 3. Since the improvement does not involve the auxiliary branch of the YOLOv9 model, it is omitted.

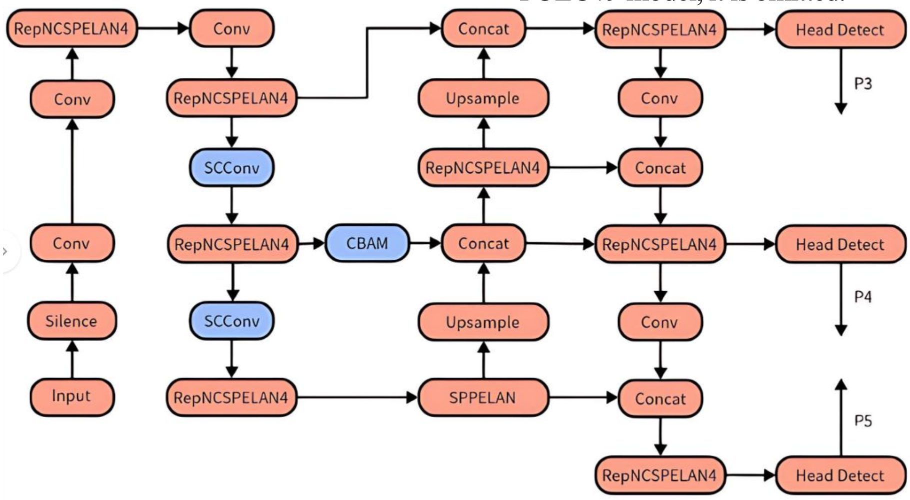

Figure 3. SC-YOLOv9 Model

4. Experiment and Analysis

4.1 Experimental Environment

This experiment was conducted in the following configuration: The hardware environment uses a NVIDIA GeForce RTX 2080 Ti*2 GPU with 22GB of video memory, a nine-core Xeon (R) Platinum 8255C CPU, and 48GB of RAM; The software environment is as follows: The operating system is Ubuntu 20.04, the deep learning framework is PyTorch 2.0, CUDA version is 11.8, and Python version is 3.9.0. All experiments were completed in this environment to ensure reproducibility and stability of the results.

4.2 Data Set and Its Preprocessing

This paper uses a self-made drone dataset for crop photography, which includes seven common types of crops: lettuce, scallions, corn, cabbage, radish, leek, and Chinese cabbage, totaling 1,000 images. The image resolution is 2250*4000 pixels, capable of clearly capturing the detailed features of the crops. Annotation method: boundary box annotation, where each target object is marked with a rectangular box to indicate its position along with corresponding category information. Preprocessing methods: (1) Data augmentation: random cropping, flipping, rotating, and color jitter are used to enhance data diversity and improve the model's generalization ability. (2) Size adjustment: all images are uniformly resized to 640*640 pixels.

4.3 Evaluation Indicators

The following indicators are used to evaluate the model:

Precision (Precision): the proportion of samples predicted as positive by the model that are actually positive. The calculation formula is:

$$Precision = \frac{TP}{TP + FP}$$

Among them, TP represents the true example (predicted to be positive and actually positive), and FP represents the false positive example (predicted to be positive but actually negative).

Recall rate (Recall): the proportion of positive samples predicted by the model in the actual positive samples, the calculation formula is:

$$Recall = \frac{TP}{TP + FN}$$

Among them, FN represents the false negative (predicted as negative but actually positive).

mAP50: The mean average precision (mAP) of the model in IoU, with a threshold of 0.5, is used to measure the comprehensive performance of the model on different categories.

$$mAP = \frac{\sum_{k=0}^{C} AP_i}{C}$$

Where: N is the total number of categoriesAP_i. It is the average accuracy (Average Precision) of the i-th category, which is calculated by the formula:

$$AP_i = \int_0^1 Precision(t)dt$$

Among them, the accuracy is when $Precision(t)dt$the confidence threshold is t.

mAP50-95: The mean average accuracy of the model in IoU, with thresholds ranging from 0.5 to 0.95 (step size of 0.05), is used to more comprehensively evaluate the performance of the model.

4.4 Ablation Test

This paper sets up four groups of experiments to analyze the influence of SCConv module and CBAM module on SC-YOLOv9 model. The experimental setup is as follows:

Experiment 1: The original YOLOv9 model, without any improvement module, as the baseline model.

Experiment 2: Replace part of the traditional 3x3 convolution layer with SCConv module in the backbone of YOLOv9 to verify the optimization effect of SCConv module on feature extraction.

Experiment 3: Add CBAM module in YOLOv9's backbone to verify the improvement effect of CBAM module on feature attention mechanism.

Experiment 4: The SCConv module and CBAM module are introduced simultaneously in the backbone of YOLOv9 to verify the comprehensive improvement effect of the synergistic effect of the two modules on the model performance.

Table 1. Results of Ablation Experiments

	P	R	mAP50	mAP50-95
Experiment 1	0.896	0.862	0.923	0.828
Experiment 2	0.906	0.852	0.918	0.722
Experiment 3	0.868	0.914	0.933	0.840
Experiment 4	0.890	0.927	0.945	0.872

4.5 Experimental Results Are Shown

The performance of Experiment 1 (the original YOLOv9 model) is good, but there are some misjudgment and missed detection when dealing with crops with complex background and similar texture.

The accuracy of experiment 2 (introducing SCConv module) is improved, but the recall rate is slightly decreased, which indicates that SCConv module may not capture some edge features sufficiently while reducing feature redundancy.

The recall rate of experiment 3 (CBAM module introduced) is significantly improved, but the accuracy rate decreases, which indicates that CBAM module can better focus on the target area, but may introduce some misjudgment.

Experiment four (simultaneously introducing the SCConv module and CBAM module) performed best, with both mAP50 and mAP50-95 higher than other experimental groups. This indicates that the synergy between the SCConv module and CBAM module can effectively enhance the overall performance of the model, reducing feature redundancy while increasing focus on target regions. The results of the SC-YOLOv9 model experiment are shown in Figure 4.

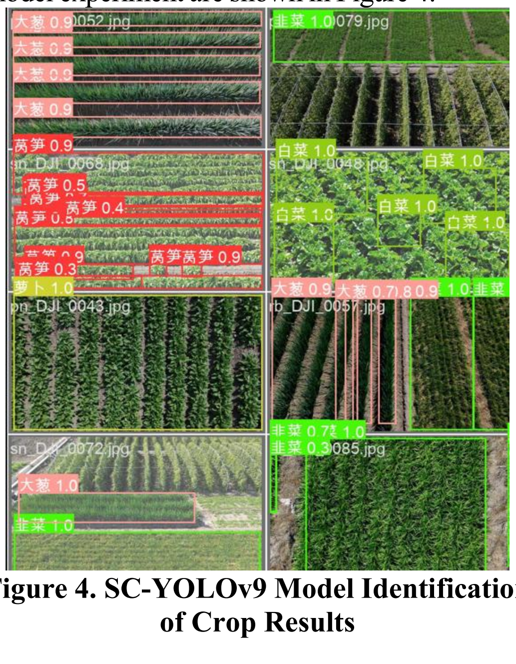

Figure 4. SC-YOLOv9 Model Identification of Crop Results

5. Conclusion

This study proposes a lightweight detection framework based on improved YOLOv9 to address the challenges of dense small target detection in UAV agricultural scenarios. Combined with dynamic data enhancement and loss function optimization strategies, the detection accuracy and real-time performance are significantly improved. The main contributions and conclusions are as follows:

Model Performance Improvement: By introducing lightweight SCConv modules and CBAM attention mechanisms, the model's ability to extract multi-scale features is enhanced while maintaining inference speed. Experiments show that the improved model achieves an mAP50 of 94.5% in dense small object scenarios, which is 2.2% higher than the original YOLOv9.

Data Augmentation Strategy Optimization: The proposed dynamic down-sampling and noise blending injection method effectively alleviates label misalignment issues caused by scale diversity in drone imagery. It also enhances model robustness by simulating complex lighting and sensor noise. Ablation experiments show that this strategy leads to better attention on small targets and strengthens the focus on target areas.

Application value and limitations: The research results can be applied to crop monitoring, pest and disease early warning and yield estimation in precision agriculture, provide technical support for real-time inspection of drones, and help agricultural automation and efficient utilization of resources.

The current research still has the following limitations: 1) The model is not adaptable to extreme light conditions (such as strong backlight); 2) The generalization ability across crops needs further verification; 3) Data enhancement depends on manual labeling, and semi-automatic labeling methods can be explored in the future.

Expand multi-modal data fusion (such as infrared and visible light image combination) to improve detection stability in complex environment;

Explore the deep adaptation of lightweight models and edge computing devices to reduce deployment costs;

A cross-regional and multi-crop agricultural detection benchmark data set was constructed to promote the research on algorithm generalization ability.

This study provides an efficient solution for UAV agricultural target detection through algorithm improvement and scenario-based strategy design, which is both theoretically innovative and has application potential, laying a technical foundation for the large-scale implementation of smart agriculture.

Acknowledgments

This paper is supported by Crop variety identification based on UAV imagery (NO.S202310755112)

References

[1].N. Falco et al., "Remote Sensing to Uav-Based Digital Farmland," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 2018, pp. 5936-5939,

[2].Y. Pratama, I. Iskandar and P. T. P. Giawa, "Implementation of Convolutional Neural Network on Farming Robots for Detecting Broccoli," 2022 International Conference on ICT for Smart Society (ICISS), Bandung, Indonesia, 2022, pp. 01-07,

[3].Chen Jinlin, Wu Yiquan, Yuan Yubin. Research progress on YOLO series algorithms for target detection from UAV perspective [J/OL]. Journal of Beihang University, 1-33[2025-03-24].

[4].Sun Daming, Xing Lulu, Zhao Wei, et al. Application of YOLO series object detection algorithm in agricultural production [J]. Information and Computer (Theoretical Edition), 2024,36(12):23-26.

[5].D. Zhu, L. Dai and P. Du, "CCE-YOLOv5s: An Improved YOLOv5 Model for UAV Small Target Detection," 2023 IEEE 5th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Dali, China, 2023, pp. 824-829,

[6].Wang Ru, Lin Haojie, Huang Wei, et al. Detection method of ancient building blue brick damage with integrated attention mechanism [J]. Journal of Safety and Environment, 2024,24(11):4244-4252.

[7].Wenxia B, Xinghua Y, Dong L, et al.Lightweight convolutional neural network model for field wheat ear disease identification [J]. Computers and Electronics in Agriculture, 2021,189 (Lightweight convolution)

[8].K. Sangaiah, F. -N. Yu, Y. -B. Lin, W. -C. Shen and A. Sharma, "UAV T-YOLO-

Rice: An Enhanced Tiny Yolo Networks for Rice Leaves Diseases Detection in Paddy Agronomy," in IEEE Transactions on Network Science and Engineering, vol. 11, no. 6, pp. 5201-5216,Nov.-Dec. 2024 (Agricultural drones)

[9].Yuchen L ,Fuzheng L ,Mingshun J .Research on YOLO algorithm for lightweight PCB defect detection based on MobileViT[J/OL].Optoelectronics Letters,1-8[2025-03-25].

[10].Guo Xin, Ma Xing, Mu Chunyang, et al. Lightweight steel wire rope defect detection method based on YOLOv8-SLA [J]. Combined Machine Tool and Automation Processing Technology, 2025, (03):140-144.

[11].Cao Jiayu, Qiao Guifang, Chen Mengyuan, et al. Improvement strategies of YOLO algorithm for small target detection from high altitude perspective [J]. Computer Applications, 2024,44(S2):280-285.

[12].K. Kiropoulos, F. Dimaraki, D. C. Tsouros and S. Bibi, "UAV saffran monitoring process," 2022 11th International Conference on Modern Circuits and Systems Technologies (MOCAST), Bremen, Germany, 2022, pp. 1-4 (drone field monitoring)

[13].Gong Xulu, Zhang Shujian. A light quantitative detection method for small target diseases in apple leaves based on improved YOLOv5s [J]. Journal of Agricultural Engineering, 2023,39(19):175-184.

[14].Jia Xueying, Zhao Chunjiang, Zhou Juan, et al. Online detection of citrus surface defects based on improved YOLOv7 model [J]. Journal of Agricultural Engineering, 2023,39(23):142-151.

[15].Yena, Wu Chenwen, Jiang Jialin. A lung sound classification model with spatial-channel reconstruction convolution module [J]. Journal of Southern Medical University, 2024,44(09):1720-1728.

[16].Hu Chengcai, Chen Yuyang, Du Xinlin, et al. A significant object detection method based on composite attention mechanism [J/OL]. Software Guide, 1-11[2025-03-24].