2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI) | 979-8-3503-6164-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICETCI61221.2024.10594197

SPD-YOLOv7: A New Method for Corn Pest Detection

Shipeng Xu*,+
Guangdong Ocean University
Zhanjiang, China
2622317147@qq.com
+Both of two authors have made equally significant contributions to the work and share equal responsibility and accountability for it.

Zishen Huang+
Guangxi Arts University
Nanning, China
2417801256@qq.com

Tzu-Ying Cheng
University of Washington
Washington, USA
biytc3nottingham@gmail.com

Zhehao Li
University of Washington
Washington, USA
timgm.lzh@gmail.com

***Abstract*—Aiming at the problems of small pests, blurred image, low resolution and large species difference in different growth periods of corn crops. In this paper, an accurate and efficient method for the detection of maize crop pests, SPD-YOLOv7, was proposed. In this paper, target background analysis, data enhancement, Gaussian noise and brightness processing are used to improve the generalization and robustness of the model. In this paper, SPD-Conv module is introduced into the framework of SPD-YOLOv7 to replace part of the step convolutional layer in the traditional system backbone and head network, so as to realize the prominent feature and location information of small target. In addition, the ELAN-W module is combined with the CBAM attention mechanism to extract more efficient features. The accuracy of the improved YOLOv7 network model reaches 98.38%, and the average accuracy reaches 99.4%.**

Keywords—yolov7, SPD-Conv module, CBAM attention mechanism, pest detection, small target detection

I. Introduction

As a major agricultural nation, corn stands as a pivotal staple crop in our country. The level of its output is directly related to the national economy and social livelihood [1]. Maize crop pests, such as maize aphids, grass armyworm, Pelagic bug, rice locust, and Lucifer bifasciata, are the main factors threatening maize food security [2]. According to literature [3], in the absence of drug control measures, the loss of corn yield due to Armyworm reached as high as 48.35%, with the ear damage rate peaking at 98.91%. The infestation of laminaria biculata will lead to a large area of corn field yield, with a yield reduction rate of 10% to 30%[4]. Traditional maize pest monitoring mainly relies on agricultural knowledge and experience of professional agricultural personnel and growers [5]. Although this method is suitable for small planting areas, it is inefficient and inaccurate. Accuracy heavily relies on the expertise of personnel involved, and the identification process is intricate, time-consuming, and labor-intensive, posing challenges in meeting the actual production demands for large-scale and rapid pest detection.

With the continuous advancement of artificial intelligence technology, machine learning and deep learning methods have become extensively utilized in the realm of plant pest detection [6-7]. For example, Wu Xiang et al. [8] proposed a pest identification method based on image feature extraction by adopting the technical route of image segmentation, feature extraction and classifier design. Dai Ning et al. [9] conducted research on maize weevil, and successfully extracted three key features of maize weevil, such as perimeter, area and complexity, by using image segmentation technology based on K-means hard clustering algorithm.

The deep learning method extracts multiscale features from training datasets, thereby significantly enhancing the accuracy and generalization capability of the model. In the domain of plant pest detection, this approach has demonstrated outstanding results [10]-[12]. The current mainstream target detection networks include SSD[13], Faster R-CNN[14-15] (Faster region with CNN) series and YOLO[16] (You only look once) series. He Hao et al. [17] proposed an improved SSD method for rice pest identification, and its average accuracy reached 79.3%. Zhou Yizhe et al. [18] used the improved Faster R-CNN algorithm to identify and locate maize pests in the video, and the accuracy rate reached 89.9%. Duan Xintao et al. [19] constructed the YOLOv4-Corn model, which effectively solved the problem of difficult identification caused by factors such as small size and easy overlap of maize pests.

Aiming at the characteristics of small size, high similarity and low detection image resolution of maize pests, an improved YOLOv7 algorithm was proposed in this paper. In addition, by adding the hybrid attention mechanism CBAM [20] to the ELAN-W module, the improved neural network can capture more feature information in channel and space respectively, significantly improving the detection accuracy of the model under complex background. This improved model can quickly and accurately detect maize crop pests in complex natural environment and provides effective technical support for maize pest detection.

II. YOLOv7 Network structure

YOLOv7[21] is a network model launched in the YOLO series in the past two years. This model is the YOLO model with the fastest inference speed and the best recognition effect on PASCAL VOC dataset [22]. The YOLOv7 network architecture comprises three fundamental components: Input, Backbone, and Detection Head. The input typically processes 640×640×3 size images, which are preprocessed and fed into the backbone network. Based on YOLOv5 architecture, the backbone network integrates Efficient Layer Aggregation

Networks (ELAN) structure and Max Pooling 1 (MP1) structure and combines CBS module to extract the features of input images. The detection head network incorporates structures such as Spatial Pyramid Pooling and Cross Stage Partial Channel (SPPCSPC), ELAN-H and MP2 feature extraction structures, and RepConv structures. The three feature layers of the backbone network output are further trained in the detection head network to integrate these outputs to achieve multi-scale detection of the target, thus output three different size prediction results. This paper takes YOLOv7 as the basic detection model and improves it.

III. YOLOv7 DETECTION ALGORITHM IMPROVED

Spd-conv module was first proposed by Raja et al [23]. Primarily, it comprises two convolution operations: space-to-depth (SPD) and non-strided convolution layers. It is employed as a replacement for traditional step convolutions to mitigate the loss of detailed information that occurs when only a small fraction of pixels are occupied during small object detection. By employing SPD-Conv, it becomes feasible to significantly enhance the accuracy of small object detection while preserving more detailed information.

The specific steps involve dividing the feature map N times along the channel direction, where each segmented sub-feature map has dimensions of S/N in both length and width, while keeping the number of channels unchanged, and then merging these sub-feature maps into the original feature map X1 along the channel direction., that is, the operation is carried out according to formula (1). This process helps retain the information in the feature map effectively, thereby improving the model's performance.

$$X\left(S\times S\times C_1\right)\rightarrow X_1\left(\frac{S}{N}\times\frac{S}{N}N^2C_1\right) \tag{1}$$

A non-step convolution An layer (step length =1) is used to convolve feature graphs to obtain features and retain as much feature information as possible, i.e., formula (2).

$$X_1\left(\frac{S}{N}\times\frac{S}{N}\times N^2C_1\right)\rightarrow X_2\left(\frac{S}{N}\times\frac{S}{N}\times C_2\right) \tag{2}$$

Among them. This operation has the advantage of downsampling the feature map while retaining the distinguishing feature information. Figure 1 shows the operation flow of SPD-Conv. Although the feature map is downsampled, it retains information on all channel dimensions.

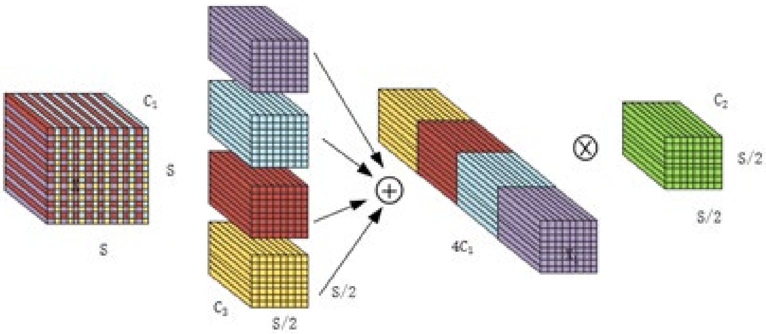

Figure 1 Structure of SPD-Conv

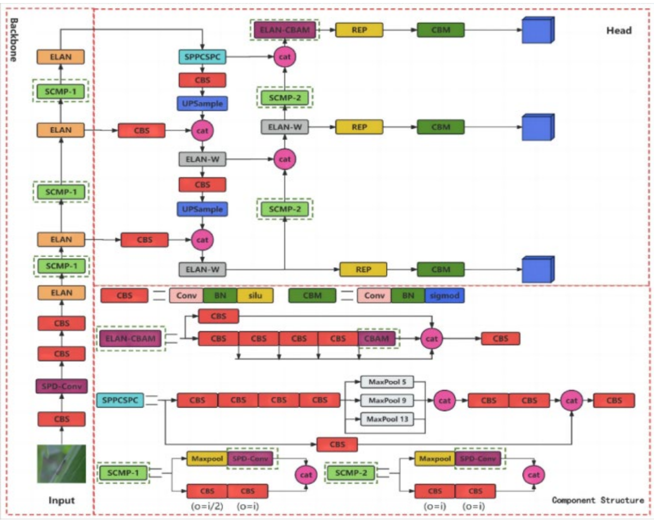

Figure 2 SPD-YOLOv7 Network model

As the convolutional neural network increases in depth, it tends to lose details of small target pests due to the convolutional stride and pooling layers. To extract effective features more fully, this paper improves the structure of MP-1 and MP-2 in Backbone and Head of YOLOv7 and replaces the CBS module with the SPD-Conv module. This improvement significantly improves the detection accuracy of small target pests. In addition, to enhance the feature extraction and learning ability of maize crop pests, the last convolutional layer of the ELAN-W module is replaced by a CBAM attention mechanism. By adjusting attention in both channel and spatial dimensions, key feature information of maize crop pests, such as color, shape and texture, is highlighted, and redundant features are suppressed to ensure accurate extraction of key features in pest images. The improved SPD-YOLOv7 model is illustrated in Figure 2, where the green dotted box highlights the enhanced components of the model. These improvements render the network more suitable for target detection tasks in complex environments, thereby enhancing the model's robustness and accuracy.

IV. EXPERIMENTAL SETUP

The experimental data of this study came from Harbin Intelligent Agriculture Industrial Park located in the Northern wilderness of Heilongjiang Province. To ensure data authenticity and universality, we selected approximately 1 mu of experimental field for data collection. Camera equipment adopts Canon D600 digital camera, using 35mm~135mm medium telephoto lens and 100mm macro lens to take close shots of corn crop pests. The data were collected from mid-June to early August 2023 and mainly covered three major maize pests: locust, Armyworm and Lucifer bimaculate. During the data screening process, we eliminated images with poor image quality due to issues such as shaking, overexposure, etc., resulting in a dataset of 1,340 photos, with about 400 samples per insect species. Figure 3 shows sample examples, with lines 1 to 3 showing samples of locust, Armyworm and Firefly beetle, respectively. Through these carefully collected and screened data, we can accurately identify and analyze maize pests.

Figure 3 Maize pest dataset

Experimental environment: The experimental platform uses Windows11 operating system, the CPU is AMD Ryzen 7 7735H with Radeon Graphics, the GPU is NVIDIA GeForce RTX 4060 (8G), and the deep learning framework is PyTorx-1.11.0. The programming language is Python-3.8.18. The integrated development environment is PyCharm2022.3.3. The setting of hyperparameters in the experimental training is shown in Table I.

TABLE I EXPERIMENTAL HYPERPARAMETER SETTINGS

Argument	Parameter value
Learning rate	0.01
Momentum	0.937
Weight	0.0005
Iteration cycle	300
Lot size	16
Image size	640×640
Learning rate attenuation method	Cosine annealing algorithm

V. EXPERIMENTAL RESULTS AND ANALYSIS

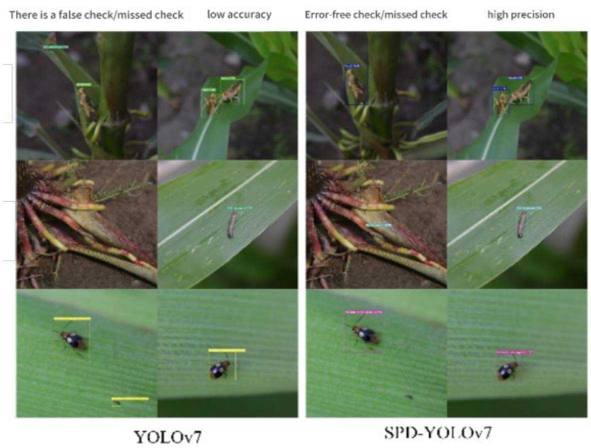

Figure 4 Test effect picture

To visually demonstrate the detection performance of the model, YOLOv7 and SPD-YOLOv7 were respectively used to detect the same pest images. As can be seen from Figure 4, the improved SPD-YOLOv7 model reduced the problems of false detection and missing detection of the original model, and the detection accuracy of the SPD-YOLOv7 model was higher than that of the original model. Experiments show that the improved SPD-YOLOv7 model has better detection effect and stronger detection ability.

To verify the effect of the improved part on the maize pest detection model, the ablation experiment was used to verify the effectiveness of each improved strategy. This study designed a total of four groups of ablation experiments using different models, and the experimental results are presented in Table II.

This study employs commonly used evaluation metrics in target detection, including precision (P), recall (R), mean average precision (mAP), detection speed (Frames Per Second or FPS), and loss function, to comprehensively assess the model's performance. These metrics objectively quantify the accuracy, recall rate, multi-class average accuracy, and real-time detection speed of the model in target detection tasks. Precision (P) denotes the ratio of correctly identified objects to the total number of objects identified (correctly and incorrectly).Its mathematical expression is shown in equation (3).

$$P = \frac{TP}{TP + FP} \tag{3}$$

In formula:

TP represents the count of corn pest images successfully detected, while FP represents the count of pest images incorrectly detected.

R represents the ratio between the correct target detected by the model and all targets present, including those not detected by the model. Its mathematical expression is shown in equation (4).

$$R = \frac{TP}{TP + FN} \tag{4}$$

In formula:

FN indicates the number of images of pests on corn crops that were not detected.

mAP refers to the calculation of accuracy under different recall thresholds and the average of these accuracies to evaluate the performance of the model at different recall levels. This is a standard evaluation metric used to objectively assess the performance of target detection models. The mathematical expression of mAP is shown in equation (5).

$$mAP = \frac{1}{N}\sum_{k=i}^{N} AP(k) \tag{5}$$

N is the number of maize crop pest classes, k is the threshold value, and AP(k) is the AP value of the detected class k pest.

FPS refers to the number of frames per second that the model can process or detect pest images, serving as a measure of the model's detection speed.

TABLE II COMPARISON OF ABLATION PERFORMANCE

Test NO.	model	P	R	mAP	FPS
1	YOLOv7	95.92	92.27	96.21	71.7
2	YOLOv7-SPD	98.50	94.32	97.89	68.3
3	YOLOv7-CBAM	96.41	98.23	99.20	72.0
4	TOLOv7-SPD-CBAM	98.38	99.51	99.40	69.0

As evidenced by the data in Table II, Experiment 1 utilized the original YOLOv7 model and demonstrated effective performance in maize pest identification. The accuracy rate is 95.92%, the recall rate is 92.27%, the mAP is 96.21%, and the FPS is 71.7. In experiment 2, based on the initial model, SPD-Conv modules are added to Backbone and Head respectively. The results indicate that compared to Experiment 1, the mAP has increased by 1.68 percentage points, reaching 98.50%, and the accuracy has improved by 2.58 percentage points. This shows that SPD-Conv module can reduce the loss of detail information and enrich the feature information of small target objects. In experiment 3, we introduced CBAM attention mechanism into the initial model. The results showed that compared with experiment 1, mAP increased by 2.99 percentage points to 99.20%, and the recall rate increased by 3.91 percentage points. This indicates that CBAM attention mechanism can allocate attention in two dimensions of channel and space and improve the recall rate and accuracy of detection of small target pests. Experiment 4 integrated the improvement methods of experiment 2 and experiment 3. Despite a slight reduction in detection speed, the accuracy rate, recall rate, and mAP increased by 2.46, 7.24, and 3.19 percentage points, respectively. Thus, the SPD-YOLOv7 model proposed in this paper exhibits superior overall performance, achieving an accuracy of 98.38%, a recall rate of 99.51%, a mAP of 99.40%, and operating at a FPS of 69.5. The ablation experiment proves that this study has positive significance for the improvement of YOLOv7 target detection model.

VI. CONCLUSION

In view of the problems in the detection of maize pests, such as small size, low resolution, easy to miss detection, false detection, and large differences in different growth periods of the same species, this study proposed an improved maize pest detection algorithm based on YOLOv7. The SPD-Conv module is introduced to replace step convolution, effectively reducing the loss of detailed information, enriching feature data for small target objects, and enhancing the accuracy of small target pest detection.

DECLARATIONS

There is no conflict of interest among the authors of this paper.

REFERENCES

[1] Shao Fei. China's corn economy: Analysis of supply and demand [D]. Xianyang: Northwest A&F University, 2011.

[2] Zha Rongqi. Common insect pests in maize and key points of comprehensive control [J]. New Agriculture,2020, (3):34-35.

[3] Ma Jiao, Zhao Xueqing, Tan You, et al. Determination of damage loss of Armya armya on maize yield in Kaiyuan City [C]// Proceedings of Green Control of Armya Armya in Yunnan Province 2019. 2019:40-42

[4] Fu Songping. A preliminary report on the damage and control of firefly bifasciata in the corn field of Tianshui City [J]. China Plant Protection Guide, 202, 42(7):69-70+75.

[5] Zhou Suhua. Identification and control technology of Armyworm in maize field in Shangqiu City [J]. Henan Agriculture, 2019, (35): 28-29.

[6] Pan Zhiguo. Research and Application of machine vision technology in Crop pests and diseases [J]. Electronic Testing & Testing, 2014,21(5): 57-58.

[7] Wang J, Li Y, Feng H, et al. Common pests image recognition based on deep convolutional neural network[J]. Computers and Electronics in Agriculture, 2020, 179: 105834.

[8] Wu Xiang. Research on pest recognition method based on machine vision [D]. Hangzhou: Zhejiang University, 2016.

[9] Dai Ning. Application of image processing in maize pest identification system [J]. Journal of Mudanjiang University, 2019, 28(2):104-107.

[10] Kiobia D O, Mwitta C J, Fue K G, et al. A review of successes and impeding challenges of IoT-based insect pest detection systems for estimating agroecosystem health and productivity of cotton[J]. Sensors, 2023, 23(8): 4127.

[11] WANG F Y, WANG R J, XIE C J, et al. Fusing multi-scale context-aware information representation for automatic in-field pest detection and recognition[J]. Computers and Electronics in Agriculture, 2020, 169(2):105222.

[12] A Setiawan A, Yudistira N, Wihandika R C. Large scale pest classification using efficient Convolutional Neural Network with augmentation and regularizers[J]. Computers and Electronics in Agriculture, 2022, 200: 107204.

[13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot Multi Box Detector[C]//European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[14] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2014:580-587.

[15] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C] //Advances in Neural Information Processing Systems, 2015: 91-99.

[16] REDMON J, DIVVALAS, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE,2016:779-788.

[17] She Hao, Wu Ling, Shan Luquan. Improved rice pest identification method based on SSD network model [J]. Journal of Zhengzhou University (Science Edition), 2019, 52(3):49-54.

[18] REN S Q, HE K M, SUN J, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017,33(6):6,1137-1149.

[19] Duan Xin-Tao, Wang Shen, Zhao Qing, et al. Research on detection method of main pests of summer maize based on improved YOLOv4 [J]. Shandong Agricultural Sciences, 2023,55 (10):167-173.

[20] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[21] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 7464-7475.

[22] Tong K, Wu Y. Rethinking PASCAL-VOC and MS-COCO dataset for small object detection[J]. Journal of Visual Communication and Image Representation, 2023, 93: 103830.

[23] Yang Z, Wu Q, Zhang F, et al. A New Semantic Segmentation Method for Remote Sensing Images Integrating Coordinate Attention and SPD-Conv[J]. Symmetry, 2023, 15(5): 1037.