MULTI-TARGET DETECTION METHOD FOR MAIZE PESTS BASED ON IMPROVED YOLOv8
/
基于改进 YOLOv8 的玉米害虫多目标检测方法

Qiuyan LIANG[1)], Zihan ZHAO[1)], Jingye SUN[1)], Tianyue JIANG[2)], Ningning GUO[1)], Haiyang YU[1)], Yiyuan GE[*1)]
[1)] School of Mechanical Engineering, Jiamusi University, Jiamusi, Heilongjiang / China;
[2)] College of Information and Electronic Technology, Jiamusi University, Jiamusi, Heilongjiang / China
Tel: 13089679130; E-mail: geyiyuan@qq.com
Corresponding author: Yiyuan Ge
DOI: https://doi.org/10.35633/inmateh-73-19

Keywords: *object detection; maize pests; yolov8; DAttention; SCConv*

ABSTRACT
When maize is afflicted by pests and diseases, it can lead to a drastic reduction in yield, causing significant economic losses to farmers. Therefore, accurate and efficient detection of maize pest species is crucial for targeted pest control during the management process. To achieve precise detection of maize pest species, this paper proposes a deep learning detection algorithm for maize pests based on an improved YOLOv8n model: Firstly, a maize pest dataset was constructed, comprising 2,756 images of maize pests, according to the types of pests and diseases. Secondly, a deformable attention mechanism (DAttention) was introduced into the backbone network to enhance the model's capability to extract features from images of maize pests. Thirdly, spatial and channel recombination convolution (SCConv) was incorporated into the feature fusion network to reduce the miss rate of small-scale pests. Lastly, the improved model was trained and tested using the newly constructed maize pest dataset. Experimental results demonstrate that the improved model achieved a detection average precision (mAP) of 94.8% at a speed of 171 frames per second (FPS), balancing accuracy and efficiency. The improved model can be deployed in low-computing-power mobile devices to achieve real-time detection, and in the future, more types of maize pests can be detected by adding multi-category datasets and training with new models with more computational power, which is important for the healthy development of maize agriculture.

摘要
玉米发生病虫害时会导致产量骤减使农民遭受重大经济损失。因此，准确、高效的玉米害虫种类检测在病虫害防治过程中可进行针对性防治。为了获得准确玉米害虫种类检测，本文提出了一种基于改进 YOLOv8n 的玉米害虫深度学习检测算法：首先，根据玉米病虫害种类，构建了玉米害虫数据集，共计2756 张玉米害虫图像；其次，在主干网络中引入具有可变形注意力机制（DAttention）来提高算法模型对玉米害虫图像的特征提取能力；然后，针对不同尺度玉米害虫，在特征融合网络中引入空间和通道重组卷积（SCConv），降低小目标害虫的漏检率；最后，基于自建玉米害虫数据集对改进后的模型进行训练和测试。实验结果表明，改进后的模型在 171 帧每秒（FPS）的速度下实现了 94.8％的检测平均精度（mAP），平衡了准确性和效率，可将改进的模型部署在低算力移动设备中实现实时检测，未来可通过添加多类别数据集和使用计算能力更强的新模型进行训练，实现更多种类的玉米害虫检测，对玉米农业健康发展具有重要意义。

INTRODUCTION
Maize is one of the most important cereal crops in the world, yet its yield is severely threatened by pests and diseases. These afflictions can lead to reduced yield, quality degradation, and even total crop failure. Consequently, early detection and targeted control of these pests and diseases are crucial for ensuring the quantity and quality of maize production. Early warning and forecasting are foundational to the effective control of maize pests and play a significant role in the agricultural management and decision-making processes related to maize production. Currently, the detection of maize pests primarily relies on manual inspection by agricultural technicians, which is time-consuming and labor-intensive (*Tian et al., 2020; Qi et al., 2021*).

---

[1] *Qiuyan Liang, Assoc. Prof., Ph.D.; Zihan Zhao, master degree; Jingye Sun, master degree; Tianyue Jiang, master degree; Ningning Guo, master degree; Haiyang Yu, master degree; Yiyuan Ge, Prof., Ph.D.*

Additionally, the acquisition of pest information in some remote areas is delayed. Therefore, exploring a rapid, efficient, cost-effective, and accurate real-time detection method to identify maize pests, which helps reduce the use of chemical pesticides, lower production costs, and protect the ecological environment, holds significant practical value (*Xinlu et al., 2023; Xu et al., 2023*).

Convolutional Neural Networks (CNNs) possess richer feature extraction capabilities compared to traditional image detection algorithms, improving both accuracy and speed of detection. Representative CNN models include AlexNet (*Yuan and Zhang, 2016*), VGGNet (*Jun et al., 2018*), GoogLeNet (*Al-Qizwini et al., 2017*), ResNet (*Targ et al., 2016*), DenseNet (*Zhu and Newsam, 2017*), Faster RCNN (*Ren et al., 2017*), and the YOLO series (*Redmon et al., 2016; Redmon and Farhadi, 2017, 2018; Bochkovskiy et al., 2020; Ge et al., 2021; Wang et al., 2023*). With the advancement of artificial intelligence technology, deep learning methods have been extensively applied to crop pest detection (*Maican et al., 2023; Xintao et al., 2023; Zongwang et al., 2023; Zhang et al., 2023; Ronghua et al.*). For instance, literature (*Hao et al., 2020*) utilized an improved SSD network for detecting rice pests, enhancing normalization and activation functions, which improved the identification rate and detection speed of rice pests, achieving an mAP of 79.3%.

Zhang et al., (2022), proposed an improved model based on YOLOv4, utilizing contextual information and a multi-scale mixed attention mechanism to enhance the biological features of pests and integrate feature information, thereby improving pest detection accuracy, with the improved model achieving an accuracy of 80.16%. *Xiaoyu et al., (2023*), introduced an improved YOLOv5 model designed to identify and locate small target pests by adding additional feature extraction layers at the cost of some detection speed, achieving a detection accuracy of 92%. *Hui et al., (2023*), presented a pest identification method based on an improved YOLOv7, incorporating Transformer and CBAM modules and making enhancements to the loss function to boost pest detection performance, with an mAP of 91.6%. *Zhu et al., (2023*), proposed a new object detection model, Poly-YOLOv8, which features a loss calculation algorithm insensitive to order and introduces a loss ratio factor based on the perimeter of polygons, outperforming other models in accurately and effectively detecting areas infected by maize leaf pests.

The aforementioned deep learning methods are capable of effectively learning target features from training data and can detect maize pests with certain levels of accuracy and efficiency. Among these, the YOLOv8 algorithm exhibits high detection precision. However, there are still some issues concerning the detection of maize pests that remain unresolved.

In recent years, research has rarely focused on the identification of maize pests; even when it has, it has only identified and detected a limited number of pest categories. Consequently, the lack of a comprehensive maize pest dataset hinders precise identification. Additionally, due to uncertain factors such as lighting and occlusion, the accuracy of existing methods in complex backgrounds is not high. Therefore, in this paper, a large-scale maize pest dataset was first constructed, which includes 2,756 images of five types of maize pests. Subsequently, the Deformable Attention mechanism (DAttention) (*Zhu et al., 2020*) was incorporated based on the YOLOv8n framework to capture richer feature information of maize pests and enhance the model's feature extraction capability.

Using Spatial and Channel Recombination Convolution (SCConv) (*Li et al., 2023*), feature fusion was performed for different scales of pest categories, reducing the miss rate of small target pests. Finally, the proposed method, embedded into the YOLOv8n algorithm, achieved accurate detection of multiple categories of maize pests. The improved YOLOv8n network model can be deployed on mobile embedded platforms, such as patrol robots, and holds promising prospects for applications in mobile target detection based on video streams.

MATERIALS AND METHODS

YOLOV8 ALGORITHM PRINCIPLES

The YOLOv8 algorithm is divided into five versions based on model size: v8n, v8s, v8m, v8l, and v8x. As the model size increases, its accuracy improves, while the detection speed gradually decreases. This allows for the selection of network models of varying depths and widths depending on the requirements of tasks such as object detection, image classification, instance segmentation, and keypoint detection. Given the hardware limitations of the deployment devices for maize pest detection, this paper utilizes the YOLOv8n model, which is compact yet highly precise. The network structure of YOLOv8n, as shown in Figure 1, comprises three main components: the Backbone, the Neck, and the Detection Head.

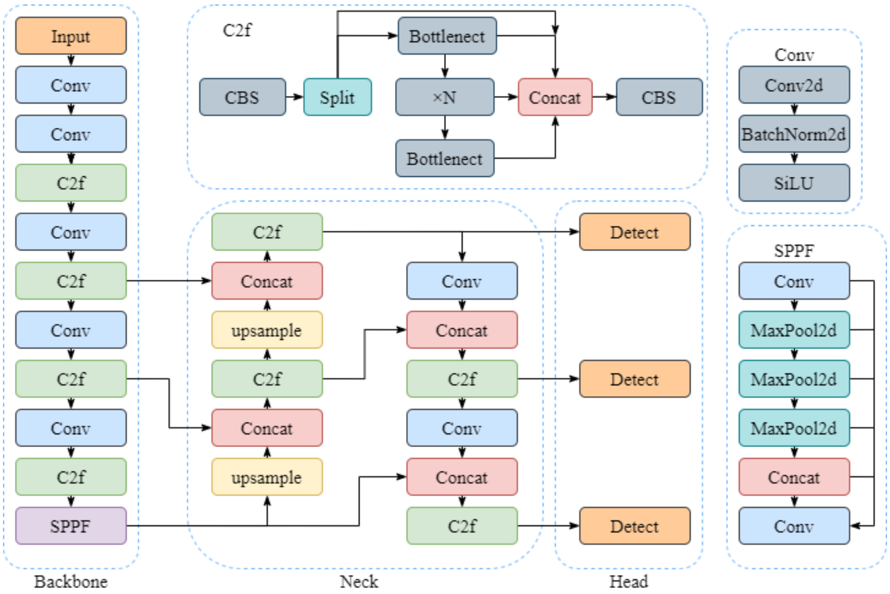

Fig. 1 - YOLOv8n network structure diagram

Backbone

The Backbone component primarily focuses on feature extraction from maize pest targets, utilizing the CSPDarknet-53 framework, which is composed of CBS, C2F, SPPF, and other modules. Specifically, the C2F module draws on the design philosophy of ELAN from YOLOv7 (*Wang et al., 2023*) and the C3 module from YOLOv5, replacing all the original C3 modules with C2F (CSPLayer_2Conv) modules for residual learning. Additionally, more branches are added within this module. Inspired by CSP and ELAN, it employs numerous skip connections and additional Split operations, integrating gradient variations from beginning to end into the feature map. The Conv convolution modules and C2F modules are serially stacked four times, with each stack referred to as a stage. Ultimately, the SPPF (Spatial Pyramid Pooling Fusion) module, used in architectures like YOLOv5, is employed to fix the vector size of different scale feature maps.

Neck

The Neck component primarily handles the feature fusion of extracted maize pest characteristics, employing the Feature Pyramid Network (FPN) (*Lin et al., 2017*) and Path Aggregation Network (PAN) (*Liu et al., 2018*) from YOLOv5 to achieve both bottom-up and top-down feature pyramids. Similar to the Backbone, the C2F module replaces the C3 module in the Neck. YOLOv8n eliminates the 1×1 convolution used in YOLOv5 and YOLOv6 before upsampling, directly performing upsampling operations on features output at different stages of the Backbone.

Head

The Head component primarily converts feature maps of maize pests into prediction results, accurately predicting the location, category, and bounding box information of maize pests. It is an improvement on the YOLOv5 model, replacing the Anchor-Based coupled head with an Anchor-Free decoupled head. The Decoupled-Head structure is used, with the classification branch and the regression branch that employs Distribution Focal Loss (DFL), thus separating the classification and detection heads. The Head has three scales - large, medium, and small - of feature map detection heads, enabling the detection of maize pests of varying sizes.

To further enhance the detection speed and accuracy of the algorithm, improvements were made to the YOLOv8n model. Based on YOLOv8n, DAttention was introduced to enhance the model's feature fusion capabilities. Additionally, SCConv was employed to improve feature fusion capabilities and increase the detection precision for small targets, thereby enhancing the detection performance for maize pests.

Optimization of Feature Extraction Network Based on DAttention

In the YOLOv8n model, DAttention is inserted prior to the SPPF network in the backbone to optimize the process. This helps the model to better extract features of maize pests at different scales, thereby allowing the model to better focus on the crucial parts of the maize and pests, improving the accuracy and robustness of detection.

DAttention is an attention mechanism used in neural networks. In traditional attention mechanisms, weights are calculated based on a fixed-location attention model. However, in deformable attention, the shape and size of the attention model can be dynamically adjusted to better adapt to the characteristics of different tasks and input data, enhancing the feature extraction capability of the backbone network. The model of the deformable attention mechanism is shown in Figure 2.

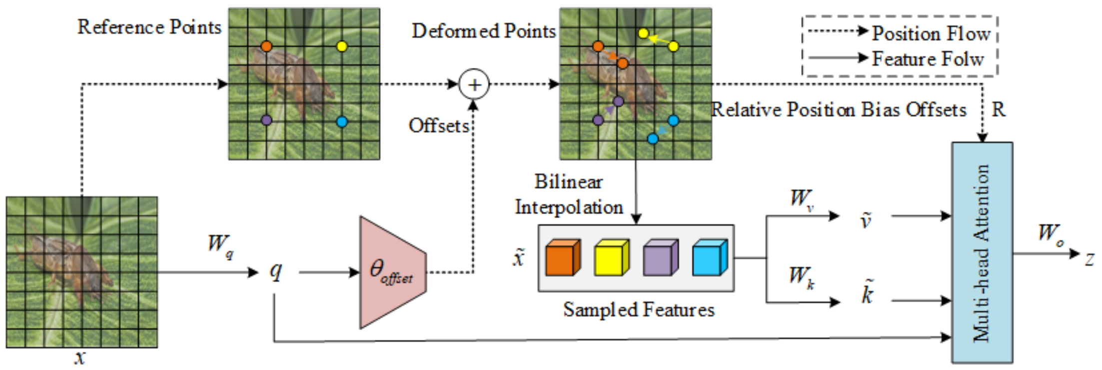

Fig. 2 - Deformable attention module

(1) Given a feature map of maize pests, $x \in R^{H\times W\times C}$, generate a uniform grid of points over $p \in R^{H_G\times W_G\times 2}$ as a reference.

(2) Linearly project the feature map onto query token $q = xW_q$, then input it into a lightweight network, θ_{offset}, to generate offsets $\Delta\theta_{offset(q)}$.

(3) Sample at the positions of the deformed points, which serve as key and value, and together with the query, pass them into the multi-head attention mechanism.

(4) Concatenate the features from each head and project them through W_o to obtain the final output.

When processing images of maize pests, DAttention focuses only on a small key area of the image, maintaining good performance while significantly reducing computational load. In the selection of sampling points, DAttention does not process the entire maize pest image rigidly; rather, it uses a dynamic selection mechanism that allows the model to focus more on the regions that are most relevant to the current task. Additionally, DAttention can adapt to different sizes of maize pests, making it effective in the identification tasks involving various types of maize pests.

Optimization of Feature Fusion Network Based on SCConv

SCConv combines spatial and channel information by performing concatenation operations across channel and spatial dimensions, enabling the network to better utilize the correlations between features in both spatial and channel aspects. In maize pest detection, SCConv helps the model better integrate multi-level features, thus enhancing the model's adaptability to different scales and complex backgrounds, and reducing the model's miss rate.

The SCConv model, as shown in Figure 3, consists of two units: the Spatial Reconstruction Unit (SRU) and the Channel Reconstruction Unit (CRU), arranged sequentially. The input features first pass through the SRU, resulting in spatially refined features, and then through the CRU, which refines the features at the channel level to produce the output. The SCConv module effectively utilizes the spatial and channel redundancy among features and can be seamlessly integrated into any CNN architecture to reduce redundancy among intermediate feature mappings and enhance the CNN's feature representation capabilities.

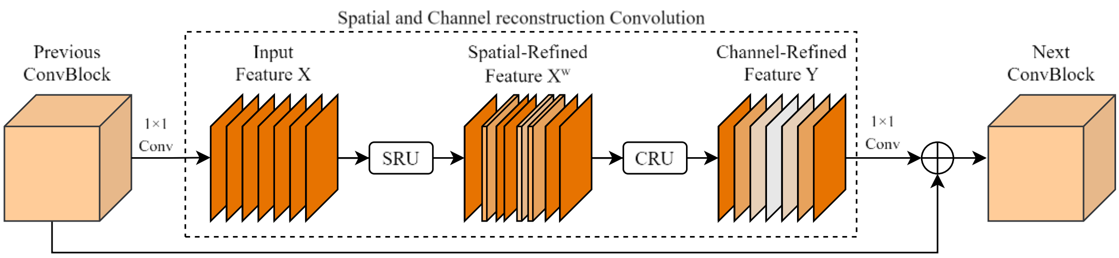

Fig. 3 - Spatial and Channel reconstruction convolution

(1) SRU - Spatial Reconstruction Unit

To utilize the spatial redundancy of features, a Spatial Reconstruction Unit (SRU) is introduced, as depicted in Figure 4, utilizing separation and reconstruction operations.

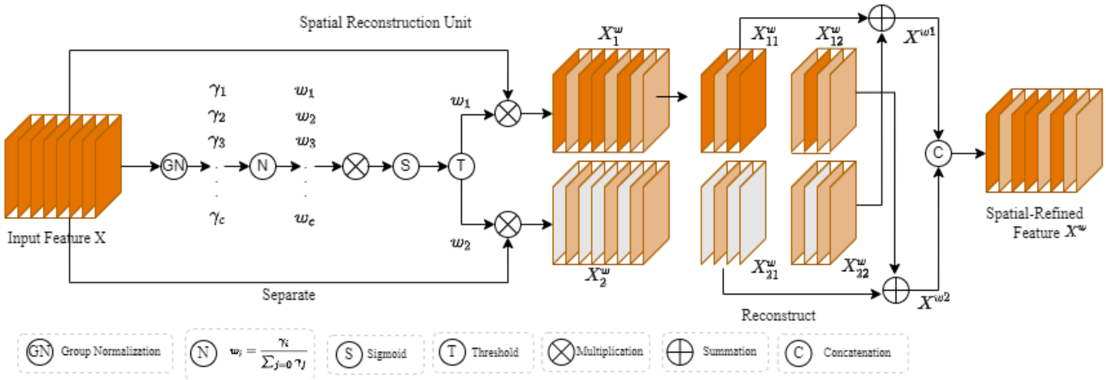

Fig. 4 - The architecture of Spatial Reconstruction Unit

The purpose of the separation operation is to distinguish feature maps with high spatial information content from those with less information. The scaling factor in Group Normalization (GN) is used to assess the information content in different feature maps, as shown in equations (1) to (3):

$$X_{out} = GN(X) = \gamma \frac{X-\mu}{\sqrt{\sigma^2+\varepsilon}} \tag{1}$$

$$W_\gamma = \{w_i\} = \frac{\gamma_i}{\sum_{j=1}^{C}\gamma_j}, i, j = 1, 2, ..., C \tag{2}$$

$$W = Gate\left(Sigmoid\left(W_\gamma\left(GN(X)\right)\right)\right) \tag{3}$$

where:

μ and σ are the mean and standard deviation of X, γ and β are trainable variables, and ε is a very small positive constant added for division stability. A larger γ represents greater variability between pixels, indicating richer spatial information.

The reconstruction operation adds features rich in information to those with less information, generating more informative features, thereby saving space. Using a cross-reconstruction operation, the weighted different information features are fully integrated to strengthen the information flow between them. Then, the cross-reconstructed features X^{W_1} and X^{W_2} are concatenated to produce a finely mapped spatial feature X^W, as indicated in equation (4):

$$\begin{cases} X_1^w = W_1 \otimes X, \\ X_2^w = W_2 \otimes X, \\ X_{11}^w \oplus X_{22}^w = X^{w_1}, \\ X_{21}^w \oplus X_{12}^w = X^{w_2}, \\ X^{w_1} \cup X^{w_2} = X^w. \end{cases} \tag{4}$$

where:

$\otimes$ represents element-wise multiplication, $\oplus$ is the summation of elements, and $\cup$ is the union operation.

Applying SRU to the intermediate input feature X not only separates information-rich features from those with less information but also reconstructs them, enhancing representative features and suppressing redundant features in the spatial dimension.

(2) CRU - Channel Reconstruction Unit

The Channel Reconstruction Unit (CRU) utilizes a splitting transformation and fusion strategy to reduce redundancy in the channel dimension as well as computational costs and storage. The CRU structure is illustrated in Figure 5.

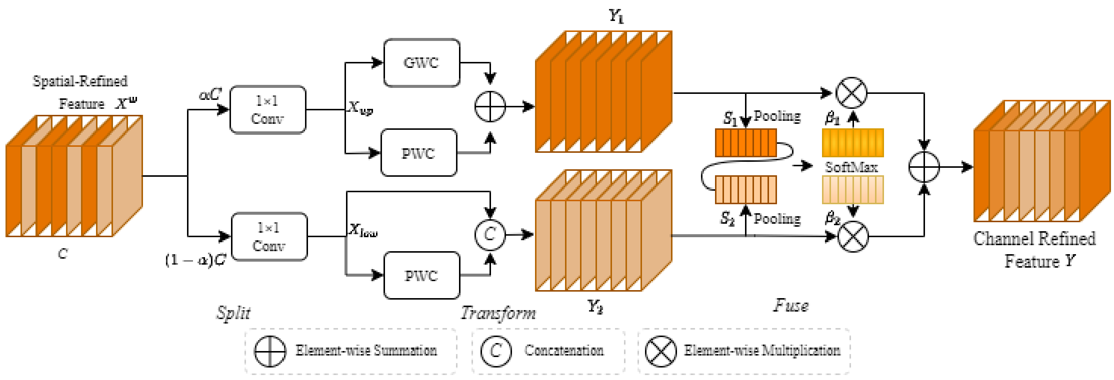

Fig. 5 - The architecture of Channel Reconstruction Unit

The splitting operation divides the input refined spatial features X^w into two parts, one with a channel count of αC and the other with a channel count of $(1-\alpha)C$, and then compresses the channel count of both groups of features using a 1x1 convolution, resulting in X_{up} and X_{low}.

The transformation operation takes the input X_{up} as the input for "rich feature extraction," undergoing GWC and PWC separately, and then the outputs are added to get the output Y_1. The input X_{low} acts as a supplement for "rich feature extraction," undergoing PWC, and the result is combined with the original input to form Y_2.

The fusion operation uses a simplified SKNet method to adaptively merge Y_1 and Y_2. First, global average pooling is used to combine global spatial information and channel statistics, obtaining pooled S_1 and S_2. Then, Softmax is applied to S_1 and S_2 to obtain feature weight vectors β_1 and β_2, and finally, these feature weight vectors are used to produce the output $Y = \beta_1 Y_1 + \beta_2 Y_2$, Y as the refined channel features.

SCConv is a plug-and-play architectural unit that can be directly used to replace standard convolutions in convolutional neural networks. Incorporating SCConv into models can improve performance by reducing redundant features of maize pests and significantly lower the complexity and computational cost of the YOLOv8n model for maize pest detection, yielding better results in detecting small target maize pests.

Design of Maize Pest Detection Model Based on Improved YOLOv8n

Building on the structure of the YOLOv8n network, DAttention is added to the C2F module before the SPPF in the backbone network, and SCConv is added to the feature fusion network in the Head, following the C2F module. The improved YOLOv8n algorithm is illustrated in Figure 6.

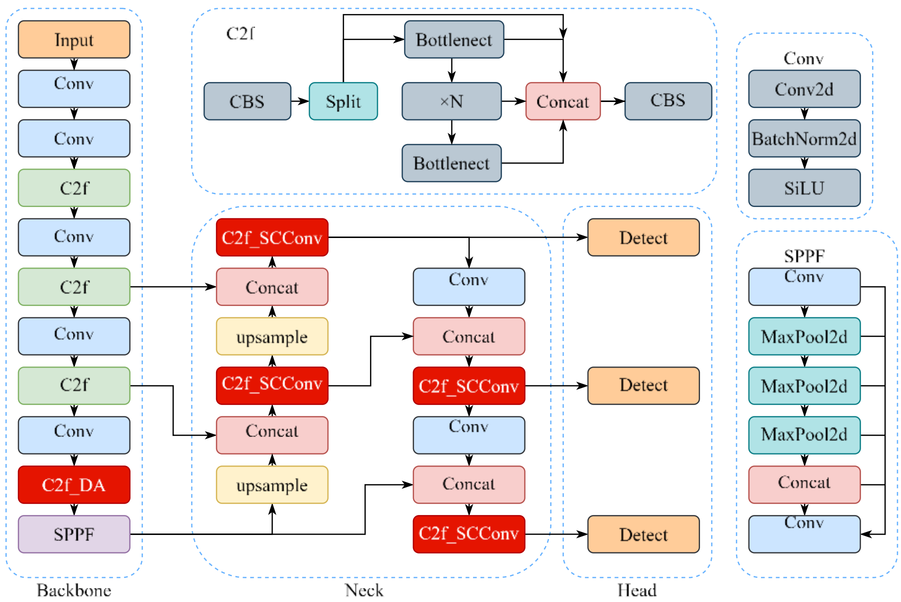

Fig. 6 - Improved YOLOv8n network structure

The introduction of the DAttention module enhances the feature extraction capability for maize pests. Initially, the YOLOv8n model extracts feature maps from the original maize pest images. Subsequently, the DAttention module is added to the C2F to enhance the extraction of features at smaller scales. This integration combines feature channels and spatial dimensions to form an attention mechanism. The focused range information from this attention mechanism is then multiplied with the input feature maps, allowing for adaptive feature refinement and thereby strengthening the feature extraction capability for maize pests.

The inclusion of SCConv enables simultaneous processing of both spatial (shape, structure) and channel (color, depth) information of images, making SCConv more precise and efficient in analyzing maize pest images. Embedded into the feature fusion network, SCConv increases detection precision for small targets, thereby enhancing the overall detection performance of the model.

Construction of the Experimental Dataset

The experimental dataset consists of 2,756 images of maize pests, including five types: aphids, mole crickets, wireworms, red spiders, and corn borers. The images of these five pest categories are shown in Figure 7. The dataset is divided into training, testing, and validation sets with a ratio of 7:1:2.

a) b) c) d) e)

Fig. 7 - Images of corn pests
a) mole cricket; b) wireworm; c) corn borer; d) red spider; e) aphids

Configuration of the Experimental Environment

The experimental environment for this study is a Windows 10 system, with an Intel i7-11800H CPU, 16GB of RAM, and an RTX4060 (8GB) GPU. All models are implemented using PyTorch. Initially, frozen training is employed, where certain convolutional layer weights are kept unchanged to retain the features learned during the pre-training phase. This approach better adapts the model to the new dataset and helps prevent overfitting. Since the initial parameters of the original model are already optimal, the model is trained with the default initial network parameters to ensure stable training. The parameters are as follows: a confidence level of 0.5, 100 training epochs, a learning rate of 1×10^{-3} for the first 50 epochs with a batch size of 16, and a learning rate of 1×10^{-4} for the last 50 epochs with a batch size of 8. During the frozen training phase, fine-tuning is performed with a smaller batch size to avoid overfitting, which requires less GPU memory. In the unfrozen phase, normal training with larger GPU memory usage is performed, hence the batch size is set to 8 for the latter 50 epochs. The input image size for the model is 640×640. The basic configuration of the local computer is shown in Table 1.

Table 1

Basic configuration of the local computer

Computer Configuration	Specific parameters/version
operating system	Windows 10
CPU	Intel i7-11800HQ
RAM	12GB
GPU	NVIDIA RTX-4060(8G)
Python	3.9
PyTorch	1.9.0
CUDA	11.3

Model Performance Evaluation Metrics

To validate the detection performance of images, the following model evaluation metrics are provided:

(1) Precision (P): Precision measures the accuracy of the model and is defined as the proportion of true positive predictions out of all positive predictions made by the model.

The formula is as follows:

$$P = \frac{TP}{TP + FP} \tag{5}$$

(2) Recall (R): Recall measures the comprehensiveness of the model and is defined as the proportion of true positives that were correctly identified by the model out of all actual positives. The formula is as follows:

$$R = \frac{TP}{TP + FN} \tag{6}$$

(3) Average Precision (AP): Average Precision assesses the detection performance of the model by calculating the precision for each sample. The formula is provided below:

$$AP = \int_0^1 P(R)dR \tag{7}$$

(4) Mean Average Precision (mAP): Mean Average Precision is a comprehensive evaluation metric that considers the precision of the model across different categories and is the average of the AP values across all categories. The formula is as follows:

$$mAP = \frac{1}{C}\sum_{i=1}^{C} AP_i \tag{8}$$

where:

AP_i represents the AP for an individual category, and C represents the total number of categories.

RESULTS

Analysis of the Experimental Process

The dataset images were input into the improved YOLOv8n model for training, with results shown in Figure 8. The loss curve displays the training errors over the dataset during training. For bounding box, category, and confidence losses, a sharp decline was observed in the first 10 iterations, indicating high learning efficiency. Subsequently, the loss curve gradually stabilized. Conversely, for accuracy and recall, the model initially showed good learning efficiency with rapid increases in values, achieving good average precision and recall rates within dozens of iterations, and then stabilizing in later iterations.

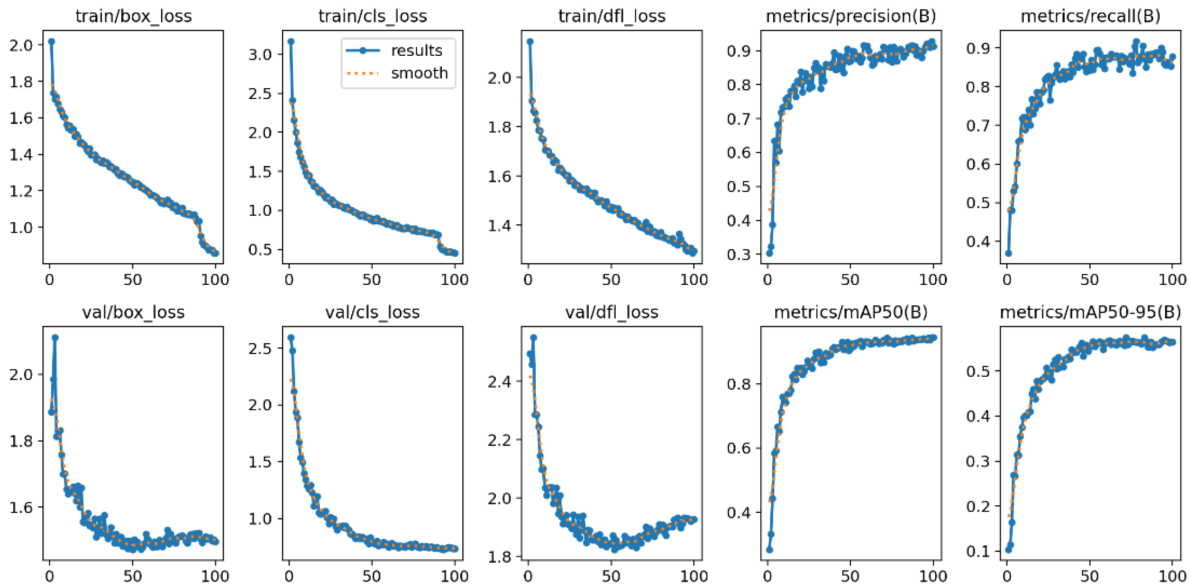

Fig. 8 - Loss function training results

Analysis of Experimental Results

To validate the performance of the improved YOLOv8n model in detecting five types of maize pests, 276 images were randomly selected for testing and evaluation with the improved algorithm. The experimental results are shown in Figure 9.

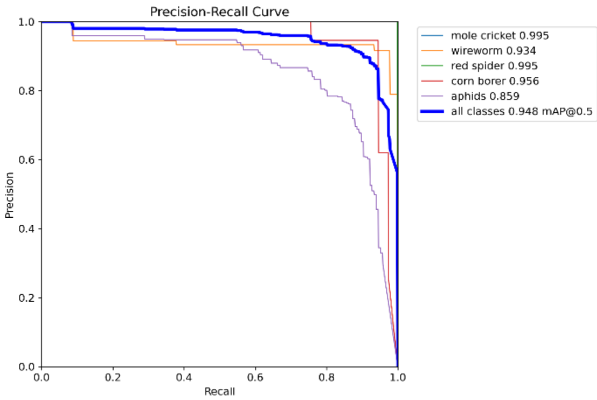

Fig. 9 - The mAP value of improve YOLOv8n

After incorporating the DAttention and SCCnov modules, the YOLOv8n algorithm achieved an mAP of 94.8%. Among these, the AP values for mole cricket and red spider were the highest at 99.5%, while the lowest AP value for aphids was 85.9%. The AP values for other maize pests ranged from high to low with corn borer at 95.6%, and wireworm at 93.4%. Overall, the improved YOLOv8n algorithm demonstrated good detection performance for maize pests. The significant difference in AP values between aphids and other pests is mainly due to their smaller size and higher color overlap with the maize background, resulting in relatively poorer detection performance. However, there was a 2.9% improvement compared to the original YOLOv8n algorithm.

To further demonstrate the detection effectiveness of the improved YOLOv8n algorithm on maize pests, a visual analysis was conducted on the five types of maize pests. The detection results are displayed in Figure 10, with detection boxes for mole cricket, aphids, red spider, corn borer, and wireworm colored in red, yellow, orange, peach, and pink, respectively. Predicted categories and confidence levels are displayed at the top left corner of each box.

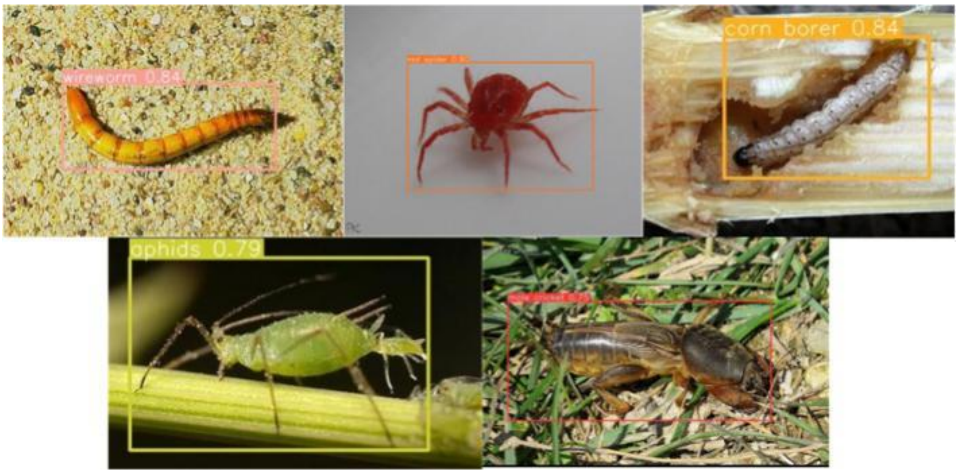

Fig. 10 - Various pest detection results chart

As shown in Figure 10, despite complex backgrounds, small sizes of various maize pests, and some occlusions, the improved YOLOv8n algorithm accurately identified all five categories of maize pests. The highest confidence level was for the red spider at 90%, and the lowest was for the mole cricket at 75%. The detection boxes were also well-fitted, accurately locating and identifying each category of maize pests. Thus, the improved YOLOv8n algorithm provides accurate localization and identification of maize pests, demonstrating good detection performance on small and occluded targets across different categories.

MODEL COMPARATIVE ANALYSIS

To explore the performance enhancements brought by incorporating the DAttention attention mechanism and SCCnov module into the YOLOv8n model and to validate the effectiveness of each component, a comparative ablation study was conducted. The study analyzed the experimental data on precision, recall, mAP, and FPS during the training processes of YOLOv5, YOLOv8n, YOLOv8n with DAttention, YOLOv8n with SCConv, and YOLOv8n with both DAttention and SCConv. A comparison of the evaluation metrics for different models is shown in Table 2.

Table 2

Comparison table of evaluation index parameters of different models

Model	DAttention	SCConv	Precision/%	Recall/%	mAP/%	FPS
YOLOv5			94.5	78.3	86.4	128
YOLOv8n			89.3	98	92.7	181
YOLOv8n	√		87.2	98	93.8	185
YOLOv8n		√	89.6	97	93.2	169
YOLOv8n	√	√	93.2	93.5	94.8	171

As can be seen from Table 2, although the YOLOv5 model performs well in terms of precision, it has relatively lower recall, mAP, and FPS. YOLOv8n maintains high precision and recall, and also improves mAP and FPS, indicating that the YOLOv8n model has been optimized for accuracy and detection speed, making it more effective for detecting maize pest targets.

Introducing DAttention into YOLOv8n resulted in a slight decrease in precision from 89.3% to 87.2%, but an increase in mAP from 92.7% to 93.8%. This suggests that DAttention helps the YOLOv8n model better balance precision and recall overall, thus enhancing the mAP value. The addition of DAttention also reduces the computational load, leading to an increase in FPS.

Adding SCConv to YOLOv8n increased the precision to 89.6%, with a slight decrease in recall to 97%, and an improvement in mAP to 93.2%, but a decrease in FPS to 169. This shows that while SCConv improves precision, it has a minimal impact on recall but does reduce processing speed.

When both DAttention and SCConv are introduced into YOLOv8n, precision increased from 89.3% to 93.2%, and mAP reached the highest value of 94.8%. Meanwhile, FPS reached 171. The improved YOLOv8n algorithm sacrificed some detection speed, but achieved superior performance in detection accuracy.

(a) YOLOv8n; (b) YOLOv8n-DAttention; (c) YOLOv8n-SCConv; (d) YOLOv8n-DAttentionSCConv

Fig. 11 - Different Model Detection Results

To validate the improvements of different models on the YOLOv8n algorithm, the detection effects of maize pests were analyzed subjectively through visual inspection. The detection results of the different models are shown in Figure 11.

From Figure 11, it is evident that different YOLOv8n models can accurately identify maize pests, with detection accuracies exceeding 70%. Among these, the original YOLOv8n model had the lowest recognition accuracy, while the YOLOv8n model with both DAttention and SCConv achieved the highest detection accuracy, with the bounding boxes more closely fitting the targets. However, the original YOLOv8n model experienced false detections when identifying corn borers, with one target being identified with two prediction boxes. The improved YOLOv8n model, with better integration of features from different scales of maize pests, more accurately recognized maize pests and enhanced detection performance.

Small datasets can limit the model from acquiring sufficient regularities and features, and changing the initial conditions or the presence of random disturbances can lead to model instability. Large datasets provide more information and training samples, which helps to learn sample features and improve model generalization, and also reduces the phenomenon of overfitting of the model to the dataset.

CONCLUSIONS

In response to the challenges posed by maize being susceptible to pest and disease damage, low yield, complex detection backgrounds, and the difficulty in identifying small-sized maize pests, this paper presents a multi-target detection method for maize pests based on the improved YOLOv8n. Initially, the DAttention module was introduced into the backbone network of YOLOv8n to enhance the model's feature extraction capabilities. Subsequently, the SCConv convolution was incorporated into the Head module to improve the model's feature fusion capabilities and to obtain richer feature information about maize pests. Ultimately, the proposed method was embedded into the YOLOv8n algorithm, achieving accurate detection of multiple categories of maize pests. Experimental results demonstrate that the improved YOLOv8n algorithm reached an mAP of 94.8%, an increase of 2.1% over the original YOLOv8n, although the FPS decreased from 181 to 171. The improved algorithm for multi-target detection of corn pests improves the accuracy, but the detection speed relatively decreases. Overall, the algorithm can meet the requirements of multi-target detection of corn pests.

In future research, the multi-class pest dataset can be increased, a larger dataset can be established for training, and the trained model can be applied to the identification and detection of various field pests. As the preliminary data labeling consumes a lot of time, online fast labeling can be established to reduce the preliminary preparation time. In terms of modeling, the model is made more lightweight by reducing the computational complexity and memory occupation of the model, and the lightweight complex model can be deployed on mobile detection equipment to achieve real-time on-site detection, which is convenient for agricultural technicians to use to prevent pests in time.

ACKNOWLEDGEMENT

This research, titled " Multi-Target Detection Method for Maize Pests Based on Improved YOLOv8", was funded by Basic Scientific Research Expenses Project of Universities in Heilongjiang (2023-KYYWF-0571); Heilongjiang Province Excellent Young Teachers Basic Research Support Program (YQJH2023218); Key Laboratory Open Topics of Field Agricultural Equipment Engineering Technology in Heilongjiang(TJNY202302); Science and Technology Plan Innovation Incentive Program of Jiamusi (NY2023JL0006); Heilongjiang Province Department of Education Science and Technology Innovation Team Construction Plan Project (2021-KYYWF-0639).

REFERENCES

[1] Al-Qizwini, M., Barjasteh, I., Al-Qassab, H., and Radha, H. (2017). Deep learning algorithm for autonomous driving using Google net. In *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE.

[2] Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.

[3] Ge, Z., Liu, S., Wang, F., Li, Z., and Sun, J. (2021). Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*.

[4] Hao, S., Ling, W., and Luquan, S. (2020). An improved SSD network model for rice pest identification. *Journal of Zhengzhou University (Science Edition)*, 52:49–54.

[5] Hui, Z., Biao, H., and Hongjun, W. (2023). Research on pest identification algorithms in complex agricultural environments based on improved yolov7. *Journal of Agricultural Machinery*, 54:246–254.

[6] Jun, H., Shuai, L., Jinming, S., Yue, L., Jingwei, W., and Peng, J. (2018). Facial expression recognition based on VGGNet convolutional neural network. *In 2018 Chinese Automation Congress (CAC)*. IEEE.

[7] Li, J., Wen, Y., and He, L. (2023). SCConv: Spatial and channel reconstruction convolution for feature redundancy. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.
[8] Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). Feature pyramid networks for object detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.
[9] Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J. (2018). Path aggregation network for instance segmentation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE.
[10] Maican, E., Iosif, A., and Maican, S. (2023). Precision corn pest detection: Two-step transfer learning for beetles (coleoptera) with MobileNet-SSD. *Agriculture*, 13(12):2287.
[11] Qi, C., Liudi, Y., and Qiuling, W. (2021). Monitoring and analysis of lepidopteran pests during the flowering and grain-filling stages of summer maize in luohe city. *Shandong Agricultural Sciences*, 53:105–110.
[12] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.
[13] Redmon, J. and Farhadi, A. (2017). Yolo9000: Better, faster, stronger. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.
[14] Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
[15] Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149.
[16] Ronghua, Z., Xue, B., and Jiangchuan, F. (2024). Pest target detection algorithm in complex scenes: YOLOv8-extend. *Smart Agriculture* (Bilingual), pages 1–14.
[17] Targ, S., Almeida, D., and Lyman, K. (2016). Resnet in Resnet: Generalizing residual architectures. *arXiv preprint arXiv:1603.08029*.
[18] Tian, H., Wang, T., Liu, Y., Qiao, X., and Li, Y. (2020). Computer vision technology in agricultural automation —a review. *Information Processing in Agriculture*, 7(1):1–19.
[19] Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.
[20] Xiaoyu, T., Qiuchi, X., and Xiaoning, H. (2023). Agricultural pest detection based on improved YOLOv5. *Journal of South China Normal University (Natural Science Edition)*, 55:42–49.
[21] Xinlu, J., Tian'en, C., and Cong, W. (2023). A review of deep learning algorithms for agricultural pest detection. *Computer Engineering and Applications*, 59:30–44.
[22] Xintao, D., Shen, W., and Qing, Z. (2023). Research on detection methods for major pests of summer maize based on improved YOLOv4. *Shandong Agricultural Sciences*, 55:167–173.
[23] Xu, W., Li, W., Wang, L., and Pompelli, M. F. (2023). Enhancing corn pest and disease recognition through deep learning: A comprehensive analysis. *Agronomy*, 13(9):2242.
[24] Yuan, Z.-W. and Zhang, J. (2016). Feature extraction and image retrieval based on AlexNet. In Falco, C. M. and Jiang, X., editors, *Eighth International Conference on Digital Image Processing (ICDIP 2016). SPIE*.
[25] Zhang, C., Hu, Z., Xu, L., and Zhao, Y. (2023). A YOLOv7 incorporating the Adan optimizer based corn pests identification method. *Frontiers in Plant Science*, 14.
[26] Zhang, W., Sun, Y., Huang, H., Pei, H., Sheng, J., and Yang, P. (2022). Pest region detection in complex backgrounds via contextual information and multi-scale mixed attention mechanism. *Agriculture*, 12(8):1104.
[27] Zhu, R., Hao, F., and Ma, D. (2023). Research on polygon pest-infected leaf region detection based on YOLOv8. *Agriculture*, 13(12):2253.
[28] Zhu, X., Su, W., Lu, L., Li, B., Wang, X., and Dai, J. (2020). Deformable DETR: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*.
[29] Zhu, Y. and Newsam, S. (2017). DenseNet for dense flow. In *2017 IEEE International Conference on Image Processing (ICIP). IEEE*.
[30] Zongwang, L., Shuaixin, Q., and Fuyan, S. (2023). Lightweight grain storage pest detection method based on improved YOLOv5s. *Journal of the Chinese Cereals and Oils Association*, 38:221–228.