SAFF-YOLO-BASED LIGHTWEIGHT DETECTION METHOD FOR THE DIAMONDBACK MOTH

基于 SAFF-YOLO 的白菜小菜蛾轻量化检测方法

Miao WU [1,2,)], Hang SHI [1,2,)], Changxi LIU [1,2,)], Hui ZHANG [1,2,)], Yufei LI [1,2,)], Derui BAO [1,2,)], Jun HU *[1,2,)],
[1)]College of Engineering, Heilongjiang Bayi Agricultural University, Daqing / China;
[2)]Heilongjiang Province Conservation Tillage Engineering Technology Research Center, Daqing / China;
Corresponding author: Jun HU; E-mail: +86 13836962331; E-mail: gcxykj@126.com
DOI: https://doi.org/10.35633/inmateh-76-13

Keywords: *YOLO11; Diamondback moth; Lightweight; Real-time detection*

ABSTRACT

The diamondback moth (Plutella xylostella) is a destructive pest that severely compromises Chinese cabbage production. Infestations caused by this pest significantly reduce both yield and quality, making efficient and accurate detection crucial for cultivation management. To address the challenges of detecting small targets and extracting phenotypic features in complex environments, this study proposes SAFF-YOLO—a YOLO11-based pest detection algorithm specifically designed for diamondback moths in Chinese cabbage fields. First, the loss function was refined to enhance the model's learning capacity for pest samples, optimizing it for precision-driven bounding box regression. Second, Alterable Kernel Convolution (AKConv) was incorporated into the backbone network, strengthening feature extraction capabilities while reducing model parameters. Third, Single-Head Self-Attention (SHSA) was integrated into the C2PSA (Channel and Position Spatial Attention) module, enhancing the backbone network's feature processing efficacy. Fourth, the neck network employed Frequency-aware Feature Fusion (FreqFusion) as the upsampling operator, specifically designed for precise localization of densely distributed targets. Finally, the Feature Auxiliary Fusion Single-Stage Head (FASFFHead) detection module was implemented to boost multi-scale target detection adaptability. Experimental results demonstrate that SAFF-YOLO achieved detection metrics of 90.7% precision, 89.4% recall, and 92.4% mAP50 for diamondback moths in Chinese cabbage, representing improvements of 7.4%, 8.0%, and 8.4% respectively over YOLO11. With only 7.3 million parameters and computational cost of 12.8 GFLOPs, this corresponds to 60.1% and 40.7% reductions compared to the baseline model. These results confirm an optimal balance between model lightweighting and high detection accuracy. Under complex field conditions characterized by small and densely distributed targets, severe background interference, and intense illumination, SAFF-YOLO consistently demonstrates robust detection capabilities, effectively reducing both false negative and false positive rates while maintaining high operational robustness. This research provides a practical solution for real-time diamondback moth detection in field-grown Chinese cabbage.

摘要

小菜蛾是严重危害白菜生产的害虫，其导致的虫害会使白菜产量、质量严重下降，因此高效、准确地检测小菜蛾对白菜栽培至关重要。针对复杂环境下小菜蛾检测存在目标小、表型特征提取困难等问题，本研究提出了基于 YOLO11 的白菜小菜蛾害虫检测算法 SAFF-YOLO。首先，改进损失函数来增强模型对害虫样本的学习能力，使其更适合边界框回归的准确性需求；引入可变核卷积（Alterable Kernel Convolution，AKConv）作为主干网络，增强了特征提取能力，减少了模型参数的数量；将单头自注意力（Single-Head Self-Attention，SHSA）集成至 C2PSA（Channel and Position Spatial Attention）模块中，提高了骨干网络的特征处理能力；颈部网络使用频率感知特征融合（Frequency-aware Feature Fusion，FreqFusion）作为上采样算子，旨在更好的对密集目标识别定位；最后通过 FASFFHead（Feature Auxiliary Fusion Single-Stage Head）检测头增强模型对不同尺度目标的检测能力。试验结果表明，SAFF-YOLO 对白菜小菜蛾的检测准确率、召回率、平均精度均值（mean average precision，mAP50）达到 90.7%、89.4%和 92.4%，对比 YOLO11 各提高了 7.4%、8.0%和 8.4%，且参数量为 7.3M，每秒浮点运算次数（Giga Floating-point Operations Per Second，GFLOPs）为 12.8，相较于基准模型分别降低 60.1%和 40.7%，实现了模型轻量化和较高检测精度的平衡。在小菜蛾小且密集、背景干扰严重、光照强烈等复杂环境下，SAFF-YOLO 均能较好地识别出目标个体，有效地降低漏检率和误检率，具有较好的鲁棒性。本研究可为田间白菜小菜蛾实时检测提供有效技术支持。

INTRODUCTION

Globally cultivated Chinese cabbage (Brassica rapa subsp. pekinensis) serves as a nutritionally essential vegetable crop. However, expanding cultivation and climate change have escalated phytopathological threats (*Ritonga et al., 2024; Li et al., 2016; Zhang et al., 2024; Shi et al., 2025*). The diamondback moth (*Plutella xylostella*), a devastating pest for Chinese cabbage, imposes substantial economic losses on farmers and challenges sustainable agricultural production systems annually (*Ahmed et al., 2022; Shehzad et al., 2023; Li et al., 2016*). Adult diamondback moths exhibit phloem-feeding behavior and rapid reproduction rates, impairing normal plant development and causing yield reduction in Chinese cabbage (*Rahman et al., 2019*). Larval chewing damage during growth stages induces extensive defoliation and leaf perforation, with severe infestations leading to complete leaf skeletonization (*Hussain et al., 2020; Hu et al., 1997*). Morphologically distinct between life stages, adults are flight-capable with predominantly gray-brown to dark brown coloration, while larvae exhibit green or pale yellow pigmentation that enables adaptive concealment within host plants. This phenotypic divergence creates significant detection challenges across developmental phases of diamondback moth (*Chen et al., 2011*). Consequently, developing rapid and accurate detection methods for diamondback moth in Chinese cabbage, coupled with effective monitoring of pest population dynamics, to implement targeted control measures, enables comprehensive pest management across critical growth stages - seedling, rosette, and heading phases - thereby safeguarding crops from significant yield losses. This constitutes an urgent agricultural priority demanding immediate resolution.

The rapid advancement of artificial intelligence has positioned machine learning and deep learning-based object detection algorithms as pivotal technologies in plant disease and pest identification research (*Chakrabarty et al., 2024*). At present, representative two-stage object detection methods such as Faster RCNN (*Ali et al., 2023; Hou et al., 2023; Wang et al., 2017*) and representative single-stage object detection methods such as SSD (*Zhai et al., 2020; Lyu et al., 2021*), YOLO are widely used for the detection and recognition of targets such as crop diseases and pests *(Wang et al., 2024; Dongfang et al., 2024*). To address the challenge of early identification of pests and diseases such as coffee leaf rust and miner pests, *Fragoso et al., (2025)*, proposed a real-time detection solution based on the YOLO series models (versions 8 to 11). The models were trained using the BRACOL dataset. Results demonstrated that YOLOv8s exhibited the optimal inference speed, with its qualitative predictive performance being significantly superior to other versions. While maintaining high accuracy, it meets the requirements for real-time field monitoring, thereby delivering robust technological support for the sustainable management of coffee cultivation. Slim *et al., (2023)*, proposed an intelligent pest detection system based on the YOLOv5 deep learning model by employing transfer learning and data augmentation techniques, the system addressed the challenge of insufficient training data. A companion mobile application was developed to assist farmers in real-time pest identification, localization, and quantification. This system significantly reduced farm inspection costs while also providing efficient, data-driven decision-making support for pest management. Liang et al., (*2024*) proposed an optimized YOLOv8n architecture for corn pest detection, incorporating Deformable Attention (DAttention) and Spatial and Channel Reconstruction Convolution (SCConv) modules. Trained on a dedicated corn pest dataset, this approach maintains high detection accuracy at 71 frames per second (FPS), enabling rapid and precise field-level infestation monitoring.

In field agriculture, micro-dense pests constitute the predominant infestation type. To address small-target detection challenges, optimizing object detection network architectures has emerged as the primary technical approach *(Wen et al., 2022*). *Teixeira et al.,* (*2023*), used the YOLOv5 deep learning model to achieve efficient recognition on the Pest24 dataset in response to the challenge of automatic detection of agricultural pests. Their experiments revealed that the relative scale of insects is a critical factor influencing detection accuracy. This approach provides a novel solution for pest detection in scenarios characterized by dense small objects, while also highlighting that enhancing the detection capability for small-sized insects represents a promising future research direction. Addressing the challenge of citrus disease detection, *Dananjayan et al., (2022),* developed the meticulously annotated CCL'20 dataset, and systematically evaluated the performance of seven state-of-the-art CNN detectors. Experimental results demonstrated that Scaled YOLOv4 P7 achieved the fastest inference speed for early-stage disease prediction, while CenterNet2 with Res2Net-101 DCN-BiFPN significantly outperformed others in detection accuracy for early-stage diseases, leveraging multi-scale feature fusion and attention mechanisms to excel particularly in identifying small-target lesion areas. This framework provides an efficient technical solution for real-time citrus disease diagnosis, effectively balancing the dual demands of inference speed and detection accuracy.

Tian et al., (*2023*), proposed MD-YOLO (Multi-scale Dense YOLO) featuring fused feature extraction/aggregation pathways with DenseNet blocks and adaptive attention modules. Deployed on sticky insect boards, this IoT-embedded system detects three lepidopteran pest species with validated field performance, demonstrating practical applicability.

While object detection algorithms have demonstrated considerable success, significant challenges persist in small-target detection research. When detecting minute pests, inherent complexities include multi-scale targets, dense clustering, and frequent occlusion scenarios. Although algorithmic advances have mitigated feature degradation caused by edge information loss and background interference, such accuracy improvements typically incur increased computational complexity that exceeds the capabilities of edge deployment platforms (*Liu et al., 2020; Lippi et al., 2021; Song et al., 2023*). To address these constraints, SAFF-YOLO is proposed—a lightweight architecture for diamondback moth detection in Chinese cabbage—which maintains high detection accuracy while substantially reducing computational footprint for efficient edge device implementation.

MATERIALS AND METHODS

Data Acquisition and Processing

The primary image acquisition for our diamondback moth dataset was conducted in Anda City, Suihua Municipality, Heilongjiang Province. To ensure data reliability and enhance dataset diversity for robust model generalization, 1,772 raw images were compiled in lossless PNG format, supplemented with carefully selected web-sourced imagery to augment phenotypic variation and ecological representativeness.

To ensure model fidelity, data acquisition employed multiple device platforms including iPhone 15 Pro and Honor 30 Pro, capturing images at dual resolutions (1280×720 and 4096×3072 pixels) thereby enhancing the recognition system's robustness and generalization capabilities across diverse imaging conditions and hardware configurations.

The initial dataset comprises raw images categorized into diamondback moth larvae and adults, supplemented through comprehensive data augmentation techniques including geometric transformations (translation, rotation), photometric adjustments (saturation/exposure modulation), random occlusion, and Gaussian noise injection. As depicted in Figure 1, these augmentations simulate diverse field conditions to enhance dataset variability while improving model sensitivity to phenotypic variations, ensuring sustained high recognition accuracy in complex real-world environments.

Fig. 1 - Data Enhancement

A final dataset of 4,794 images was selected to ensure data quality and diversity, which was then divided into training, validation, and test sets at an 8:1:1 ratio. To ensure annotation accuracy, this study employed the open-source labeling software LabelImg to manually annotate diamondback moth specimens in images, ultimately establishing the Plutella xylostella Dataset.

SAFF-YOLO-based lightweight detection method for the diamondback moth

YOLO11 is an advanced single-stage object detection model whose architecture comprises three core components: a backbone network, a neck network, and a detection head. Building upon its predecessors, YOLO11 incorporates significant optimizations—including a more powerful backbone network to enhance feature extraction efficiency and an innovative neck design featuring an enhanced EfficientDet-inspired FPN architecture to improve multi-scale object detection capabilities. The detection head employs a refined design that achieves precise prediction of object categories and locations, enabling efficient and accurate identification and localization of diverse targets even in complex, dynamic scenarios. The overall architectural layout of the YOLO11 model is illustrated in Figure 2.

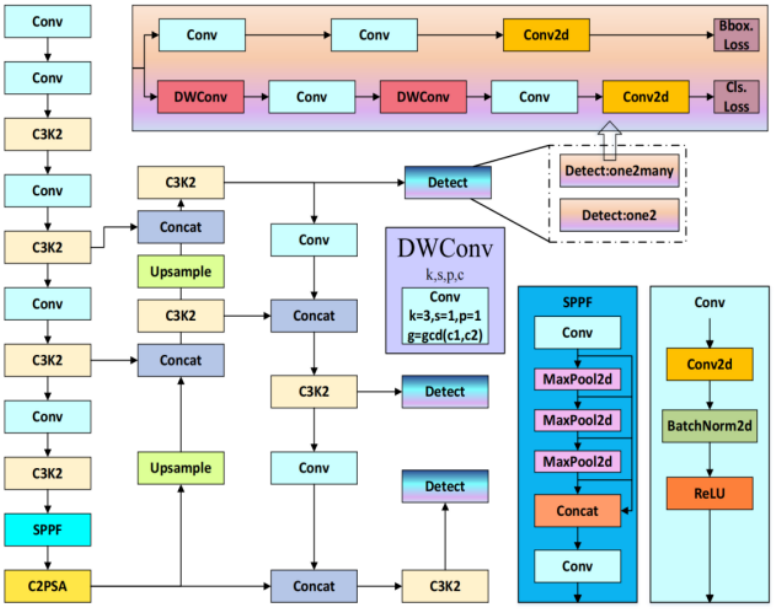

Fig. 2 - YOLO11 network structure

Despite its capability to accurately detect and identify targets, support multi-class object detection, and perform real-time tracking, YOLO11 remains challenged in achieving precise detection of early-stage small-scale pest infestations in field conditions. Specifically for diamondback moth detection, it struggles to extract phenotypic characteristics and accurately identify small targets. To address these limitations, this study proposes SAFF-YOLO (an enhanced YOLO11-based algorithm) for detecting diamondback moth infestations in Chinese cabbage fields, which significantly improves performance in agricultural pest detection tasks. The architecture of the proposed model is illustrated in Figure 3.

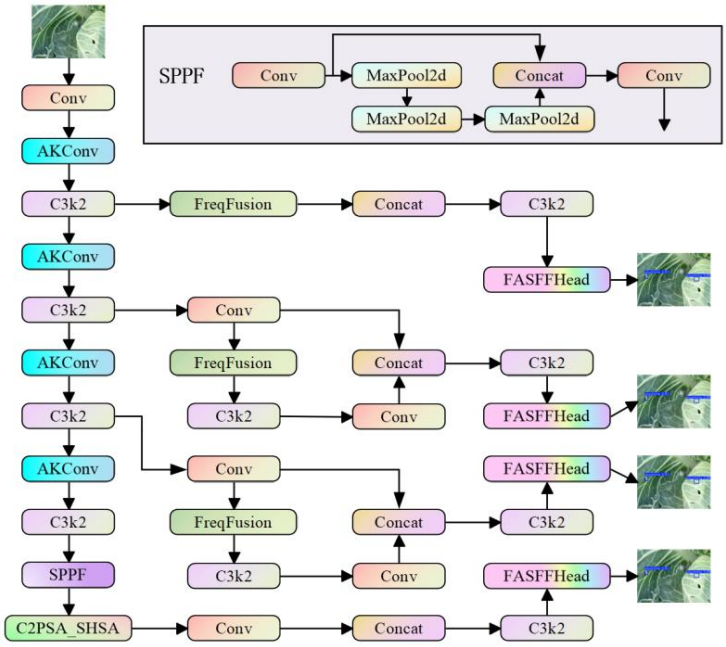

Fig. 3 - Structure of the SAFF-YOLO network

The original loss function was replaced with the Unified-IoU (UIoU) loss function (*Luo et al., 2024*), which dynamically adjusts the model's focus on prediction boxes of varying quality to optimize bounding box regression accuracy in object detection, thereby enabling more precise pest localization. The AKConv lightweight convolutional module was introduced to replace partial network structures *(Zhang et a., 2023*), enhancing feature extraction capability while reducing model parameters. Furthermore, SHSA and C2PSA modules were integrated into a unified SHSA_C2PSA module (*Yun et al., 2024*), which augments the backbone network's feature processing capacity, improves small-target detection performance, and reduces computational redundancy and memory access costs.

The neck network employs FreqFusion as its upsampling operator (*Chen et al., 2024*), enhancing localization in dense object scenarios while reducing computational complexity and improving processing speed. Within the detection head, the adaptive spatial feature fusion (ASFF) method was integrated with the P2 detection layer, evolving into a novel FASFFHead module (*Liu et al., 2019*). This integration mitigates feature loss during cross-scale fusion and implements secondary feature extraction for small targets, thereby enhancing model recognition accuracy.

Improvement of Loss Function

In the context of diamondback moth detection, the YOLO11 model faces challenges due to target density, minute scale, occlusion, and partial body visibility, leading to suboptimal recognition performance. To address this limitation, the original Complete Intersection over Union (CIoU) loss was replaced with the Unified-IoU (UIoU) loss function. This novel approach dynamically reallocates model focus from low-quality to high-quality prediction boxes through adaptive weight assignment, enhancing detection performance on both high-precision and densely clustered datasets while maintaining training efficiency. The mechanism effectively captures spatial relationships and overlapping region information among targets. The UIoU computation is formalized in Equation (1).

$$L_{UIoU} = 1 - IoU + \frac{\rho^2(b,b^{gt})}{c^2} + \alpha v \tag{1}$$

where ρ denotes the Euclidean distance between two points, b represents the center coordinate of the prediction box, b^{gt} indicates the ground truth box center, c is the diagonal length of the smallest enclosing rectangle, α serves as a scaling coefficient for balance, and v incorporates the aspect ratios of both prediction and target boxes into the sigmoid function. This formulation fulfills a dual role: (1) thresholding values to prevent excessive oscillation, and (2) quantifying aspect ratio consistency between boxes. The formal definitions of α and v are given in Equations (2) and (3), respectively.

$$\alpha = \frac{v}{(1-IoU)+v} \tag{2}$$

$$v = \left(\frac{1}{1+e^{-\frac{\omega 1}{h1}}} - \frac{1}{1+e^{-\frac{\omega}{h}}}\right)^2 \tag{3}$$

In the formula, $\omega 1$ and ω represent the lengths of the target box and the predicted box, while h_1 and h represent the widths of the target box and the predicted box.

The derivative calculation of UIoU is shown in equations (4) and (5).

$$\frac{\alpha v}{\alpha \omega} = 2\left(\frac{1}{1+e^{-\frac{\omega^{gt}}{h^{gt}}}} - \frac{1}{1+e^{-\frac{\omega}{h}}}\right)\left[\frac{1}{1+e^{-\frac{\omega}{h}}}\left(1 - \frac{1}{1+e^{-\frac{\omega}{h}}}\right)\right]/h \tag{4}$$

$$\frac{\alpha v}{\alpha h} = 2\left(\frac{1}{1+e^{-\frac{\omega^{gt}}{h^{gt}}}} - \frac{1}{1+e^{-\frac{\omega}{h}}}\right)\left[\frac{1}{1+e^{-\frac{\omega}{h}}}\left(1 - \frac{1}{1+e^{-\frac{\omega}{h}}}\right)\right] \times \frac{\omega}{h^2} \tag{5}$$

UIoU achieves adaptive weight allocation for prediction boxes of varying quality through strategic scaling of both prediction and ground truth boxes. This approach eliminates redundant bounding box computations while maintaining geometric integrity. After acquiring the bounding box dimensions (height, width) and center coordinates, UIoU applies scale factors to proportionally expand or contract these dimensions. This controlled scaling dynamically modulates the model's attention across prediction box qualities—where box contraction intensifies focus on high-quality predictions, thereby enhancing precision detection performance for well-defined targets.

AKConv

Convolutional kernels have achieved remarkable success in deep learning, yet they exhibit two inherent limitations. First, the confinement to local receptive fields restricts their capacity to capture global contextual information, while their fixed sampling patterns further limit adaptability. Second, conventional square-shaped kernels with fixed dimensions cause parameter counts to increase quadratically with kernel size. These rigid sampling geometries and kernel shapes struggle to adapt to the diverse target morphologies and scales across datasets and spatial locations (*Zhang et al., 2023*).

To overcome these limitations, AKConv (Adaptive Kernel Convolution) is introduced, which employs a flexible convolutional mechanism permitting kernels with arbitrary parameter counts. By dynamically adjusting kernel shapes to accommodate diverse image characteristics, this approach enhances model adaptability and computational efficiency. Consequently, AKConv not only improves model performance but also reduces parameter quantities. The architectural configuration of the AKConv module is illustrated in Figure 4.

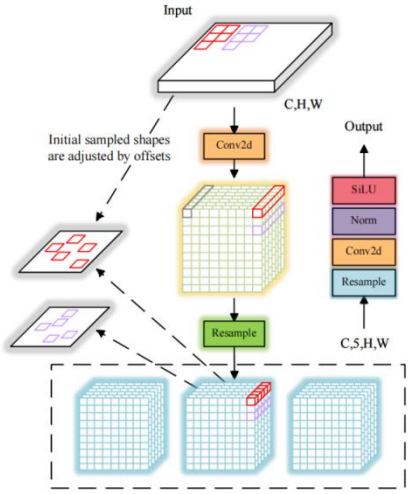

Fig. 4 - AKConv structure diagram

AKConv introduces a novel coordinate generation algorithm that dynamically adapts sampling shapes to varying images and targets. This algorithm generates initial sampling coordinates for kernels of arbitrary sizes and geometries (Fig.5), significantly enhancing flexibility in detecting multi-scale targets. To accommodate target variations, AKConv adjusts sampling positions of non-rectangular kernels through learned offsets, thereby improving feature extraction accuracy.

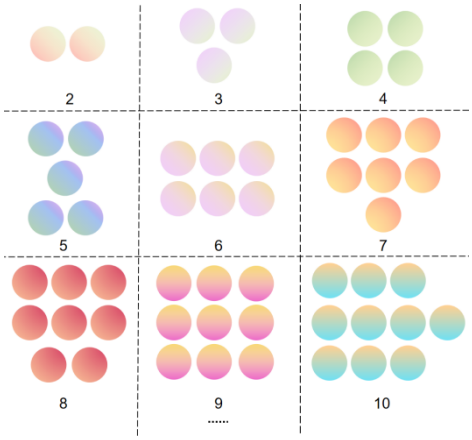

Fig. 5 - Adaptive initial sampling shape

Unlike traditional convolutional kernels that employ regular sampling grids, AKConv targets deformable kernels with irregular geometries. This innovation necessitated the development of an arbitrary-size convolution algorithm that generates initial sampling coordinates P_n for convolutional kernels (*Wu et al., 2024*). The algorithm first generates a regular sampling grid, subsequently constructs an irregular grid for residual sampling points, and finally concatenates both components into a unified sampling structure.

Within this framework, the top-left coordinate (0,0) serves as the sampling origin. Given initial coordinates P_n and kernel parameters ω for irregular convolution, the convolutional operation at position P_0 is formally defined by Equation (6).

$$ConvP_0 = \sum \omega \times (P_n + P_0) \quad (6)$$

AKConv resolves the fundamental mismatch between irregular sampling coordinates and fixed-size convolution operations through its algorithmic innovations.

Traditional convolution suffers from quadratic growth in both parameter count and computational load with increasing kernel size, leading to inefficiency in resource-constrained environments. In contrast, AKConv's unique design reduces model parameters and computational overhead, enabling dynamic complexity adjustment according to task requirements and hardware capabilities. Furthermore, AKConv adapts to spatial feature variations through offset-adjusted kernel positioning, effectively handling non-rigid deformations, occlusions, and complex backgrounds. This capability significantly enhances detection robustness for subsequent diamondback moth identification. The schematic representation of this adaptation process is illustrated in Fig. 6.

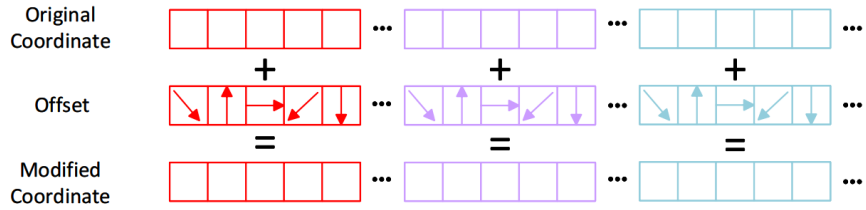

Fig. 6 - Offset adjustment sample shape

Single-Head Self-Attention

Accurate extraction of spatial location information is critical in object detection tasks. In pest-containing images where natural backgrounds dominate, convolutional kernels process non-target regions, introducing substantial redundant information that compromises pest recognition accuracy. The Single-Head Self-Attention (SHSA) mechanism addresses this by applying attention to a subset of input channels (CP=rC), reducing computational redundancy while integrating global and local features to enhance efficiency and precision. As illustrated in Fig. 7, SHSA operates by: (1) applying a single-head attention layer to a channel subset (CP=rC), where r denotes the reduction ratio) for spatial feature aggregation, while (2) preserving original information in remaining channels.

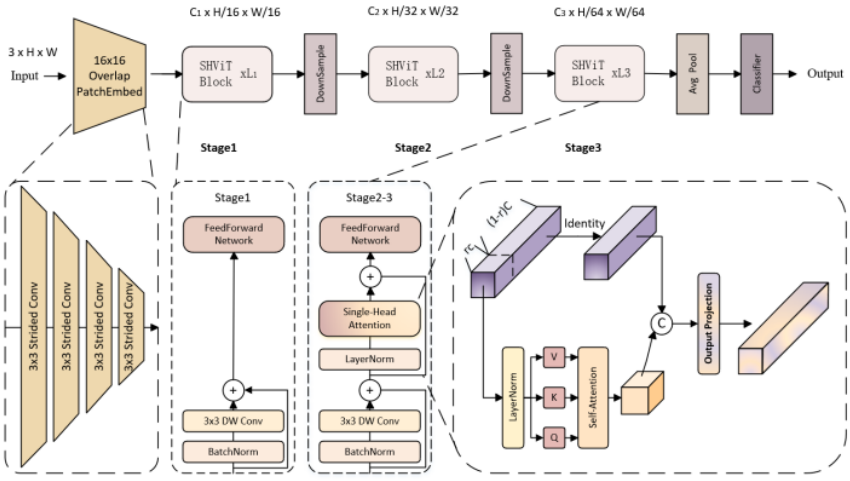

Fig. 7 - Structure of SHSA attention mechanism

The calculation formulas for its operation are shown in equations (7) to (10).

$$SHSA(X) = Concat\left(\widetilde{\mathrm{X}}_{att},\ X_{res}\right)W^{Q} \quad (7)$$

$$\widetilde{\mathrm{X}}_{att} = Attention\left(X_{att}W^{Q},\ X_{att}W^{K},\ X_{att}W^{V}\right) \tag{8}$$

$$Attention\left(Q,\ K,\ V\right) = Softmax\left(QK^{T}/\sqrt{d_{qk}}\right)V \tag{9}$$

$$X_{att},\ X_{res} = Split\left(X,\ \left|C_{p},\ C - C_{p}\right|\right) \tag{10}$$

where W^{Q}, W^{K}, and W^{V} denote projection weight matrices, dqk represents the dimensionality of queries and keys, and Concat signifies the concatenation operation. To maintain memory access consistency, the initial C_p channels serve as representative proxies for the complete feature map. Moreover, SHSA's final projection operates across all channels—not solely the initial C_p subset—ensuring efficient propagation of attention features to residual channels.

Frequency-aware Feature Fusion

Feature extraction of diamondback moths is challenged by small target regions, uneven density distribution, and low image resolution, resulting in insufficient valid information. However, dense image prediction requires high-precision category information and spatial boundaries. Conventional feature fusion methods underperform in maintaining category feature consistency and preserving boundaries, often causing significant intra-class feature variations and boundary ambiguity.

FreqFusion is a frequency-aware feature fusion framework comprising three core modules: an Adaptive Low-Pass Filter Generator (ALPF), an Offset Generator, and an Adaptive High-Pass Filter Generator (AHPF). The ALPF predicts spatially adaptive low-pass filters to mitigate intra-class inconsistency; the Offset Generator predicts feature offsets to refine internal representations and boundary characteristics; while the AHPF extracts high-frequency details for precise boundary delineation. As illustrated in Fig.8, these modules operate synergistically to resolve intra-class inconsistencies and boundary ambiguities in dense image prediction. Through frequency-aware feature refinement, this integrated mechanism enhances model performance.

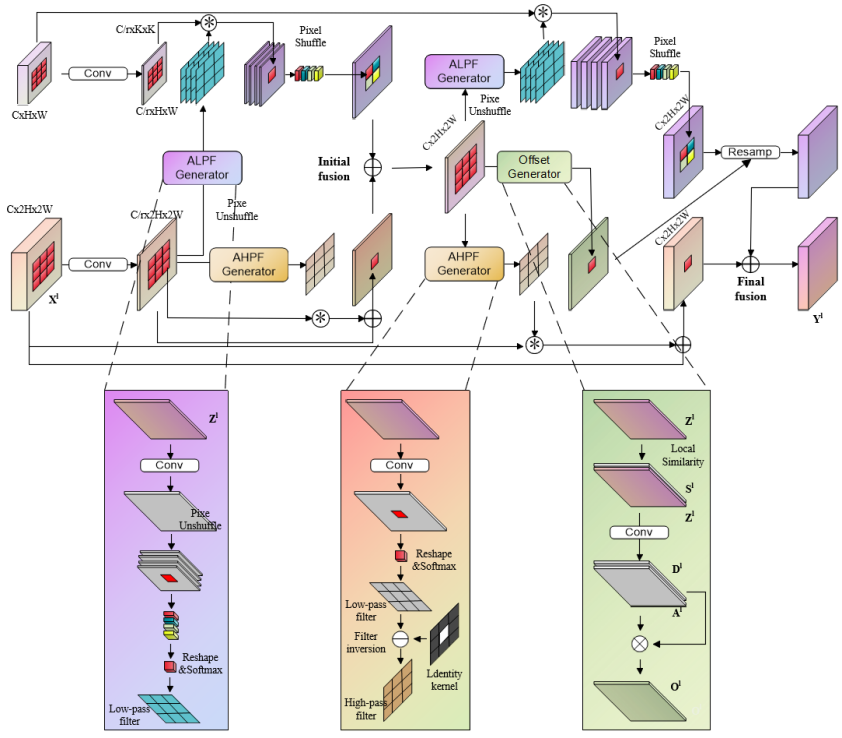

Fig. 8 - FreqFusion structure diagram

The generation calculation formula for FreqFusion is shown in Equations (11) and (12).

$$Y_{i,j}^{l} = \tilde{Y}_{i+u,j+v}^{\ l+1} + \tilde{X}_{i,j}^{\ l} \tag{11}$$

$$\tilde{Y}^{\,l+1} = \mathcal{F}^{UP}\left(\mathcal{F}^{LP}(Y^{l+1})\right),\ \tilde{X}^{l} = \mathcal{F}^{HP}(X^{l}) + X^{l} \tag{12}$$

In this formulation, $\mathcal{F}^{UP}$ denotes the low-pass filter predicted by the ALPF generator, while (u, v) represents the offset values predicted by the Offset Generator for feature coordinates at (i, j), and $\mathcal{F}^{UP}$ signifies the high-pass filter generated by AHPF. These components collectively resolve class inconsistency and boundary displacement by: (1) adaptively smoothing high-level features using spatially adaptive low-pass filtering, (2) replacing inconsistent features through resampling of adjacent class-consistent features, and (3) enhancing high-frequency boundary details in low-level features.

Feature Auxiliary Fusion Single-Stage Head

In object detection, conventional YOLO series achieve efficient detection through multi-scale feature fusion. While the YOLO11 detection head demonstrates advantages, it exhibits limitations in complex scenes or small object detection. The FASFFHead addresses this by: (1) introducing auxiliary feature layers and a feature selection module, (2) fusing multi-level features to enhance network sensitivity towards multi-scale and complex objects, and (3) improving extraction and representation of multi-scale features for superior discriminative feature capture. As illustrated in Fig. 9, the FASFFHead primarily consists of two key components: a Shallow-level Feature Fusion module (SFB) and a High-level Feature Extraction module (HFE), operating through the following mechanism.

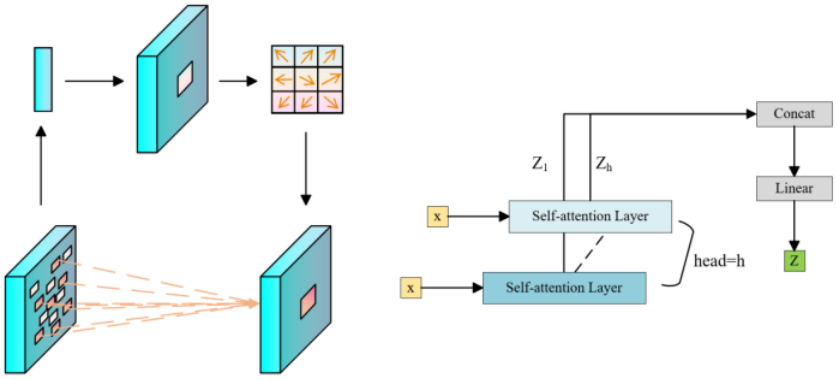

Fig. 9 - FASFFHead structure diagram

Within the SFB module, shallow-level and high-level features are integrated through a cascaded residual network and feature pyramid structure. This fusion preserves detailed spatial information from shallow features while incorporating global semantic context from high-level features, thereby enriching feature maps and enhancing representational capacity. The HFE module subsequently processes these fused features through depthwise convolutional layers to extract deep contextual information, while employing a Spatial Pyramid Pooling (SPP) layer to capture multi-scale representations. This dual mechanism significantly improves detection capability for objects across varying scales.

By synergistically combining SFB and HFE modules, the FASFFHead effectively fuses auxiliary features with high-level representations from the backbone network. This integration substantially boosts detection performance, simultaneously enhancing feature expressiveness while improving model robustness and generalization capability for practical object detection applications.

RESULTS AND ANALYSIS

Test environment and evaluation index

To ensure experimental validity, detailed hardware and software configurations are specified in Table 1.

Table 1

Software and Hardware Environment Configuration Table

Configuration Parameter	Configuration Item
CPU	16 vCPU Intel(R) Xeon(R) Platinum 8352V CPU @ 2.10GHz
RAM	120GB
GPU	NVIDIA GeForce RTX 4090 24GB
Operating system	Ubuntu 18.04
CUDA	12.4
Compiled language	Python 3.9
Deep learning framework	PyTorch3.8.0
Epochs	200
Batch size	32

The evaluation employs three metrics: $mAP@50$, precision (P), and recall (R). Precision quantifies the proportion of true positives among all positive predictions, while recall measures the proportion of actual positives correctly identified. Their computational formulations are given by Equation (13) and Equation (14), respectively.

$$P = \frac{T_P}{T_P + F_P} \tag{13}$$

$$R = \frac{T_P}{T_P + F_N} \tag{14}$$

where F_N denotes the number of false negatives (missed detections), F_P represents false positives (detections with IoU below the threshold), and T_P indicates true positives (detections with IoU ≥ threshold).

For mAP computation, the average precision (A) corresponds to the area under the Precision-Recall (P-R) curve, calculated as specified in Equation (15).

$$A = \int_2^1 P(R)dR \tag{15}$$

Performance comparison test of different models

To evaluate the pest recognition and detection performance of the SAFF-YOLO model, comparative experiments were conducted against prevalent object detection models: Faster R-CNN, SSD, YOLOv8, YOLOv9, YOLOv10, and YOLO11. Identical experimental setups (hardware/software configurations) were maintained across all models to ensure methodological rigor and validity. Model performance was comprehensively assessed using five metrics: precision, recall, mean average precision (mAP), floating-point operations (FLOPs), and parameter count. Comparative results are presented in Table 2.

Table 2

Training effect of different models

Model	Precision(%)	Recall(%)	mAP0.5(%)	GFLOPs	Model Size(MB)
Faster-RCNN	82.1	80.3	83.7	124.2	71.6
SSD	80.1	77.9	80.7	61.1	39.2
YOLOv5	81.3	76.8	81.8	5.9	4.4
YOLOv8	82.0	79.9	84.6	6.9	5.4
YOLOv10	81.7	78.4	82.3	6.6	5.2
YOLO11	83.3	81.4	84.0	21.6	18.3
SAFF-YOLO	90.7	89.4	92.4	12.8	7.3

According to Table 2, the SAFF-YOLO model demonstrates mAP@50 improvements of 8.7%, 11.7%, 10.6%, 7.8%, 10.1%, and 8.4% over Faster R-CNN, SSD, YOLOv5, YOLOv8, YOLOv10, and YOLO11, respectively. These gains indicate superior target localization accuracy, enhanced recognition capability for pests in complex backgrounds and dense populations, and higher confidence scores with improved reliability in detection outcomes.

Ablation Experiments

Targeting diamondback moth infestation detection, this study progressively enhanced the YOLO11 architecture. Each modification underwent statistical analysis to validate efficacy in pest recognition, with quantitative results detailed in Table 3.

Table 3

SAFF-YOLO model ablation experiments

Model	Precision(%)	Recall(%)	mAP0.5(%)	GFLOPs	Model Size(MB)
YOLO11	83.3	81.4	84.0	21.6	18.3
YOLO11-U	85.0	83.5	85.8	21.6	18.3
YOLO11-UA	83.1	82.3	85.1	5.5	4.4
YOLO11-UAB	86.2	83.7	86.7	5.4	4.3
YOLO11-UABC	88.6	87.3	89.6	6.1	4.3
SAFF-YOLO	90.7	89.4	92.4	12.8	7.3

Note: U represents replacing the UIoU loss function; A represents the AKConv improvement; B represents the SHSA improvement; C represents the FreqFusion improvement

Stage 1: UIoU replacement improved precision, recall, and mAP50 of baseline YOLO11 by 1.7%, 2.1%, and 1.8% respectively. Stage 2: AKConv modification reduced parameter count by 76% and computational load by 16.1 GFLOPs versus baseline. Stage 3: Integrating SHSA and FreqFusion modules increased mean average precision by 1.7-2.9 percentage points without computational overhead, enhancing all metrics.

Final integration: Incorporating the FASFFHead module, SAFF-YOLO demonstrated significant improvements over YOLOv11 across all evaluated metrics: a 40.7% reduction in model size, an 8.8 GFLOPs decrease in computational cost, and respective gains of 7.4% in precision, 8.0% in recall, and 8.4% in mAP50. This demonstrates SAFF-YOLO's superior accuracy and robustness when detecting micro-scale pests in dense infestations with scale variations and homogeneous backgrounds.

Model visualization analysis and comparison

Comparative detection results for diamondback moths using YOLOv5, YOLOv8, YOLO11, and SAFF-YOLO are presented in Fig. 10, validating SAFF-YOLO's superior performance in this study. Visual evidence demonstrates SAFF-YOLO's enhanced robustness in complex environments with significant interference, effectively reducing false positives and misclassifications prevalent in baseline YOLO variants. Under challenging conditions featuring small target pests and high identification difficulty, comparator models exhibit frequent false detections, duplicate identifications, and low-confidence predictions. SAFF-YOLO significantly outperforms these models through improved environmental adaptability and superior discriminative feature representation capabilities.

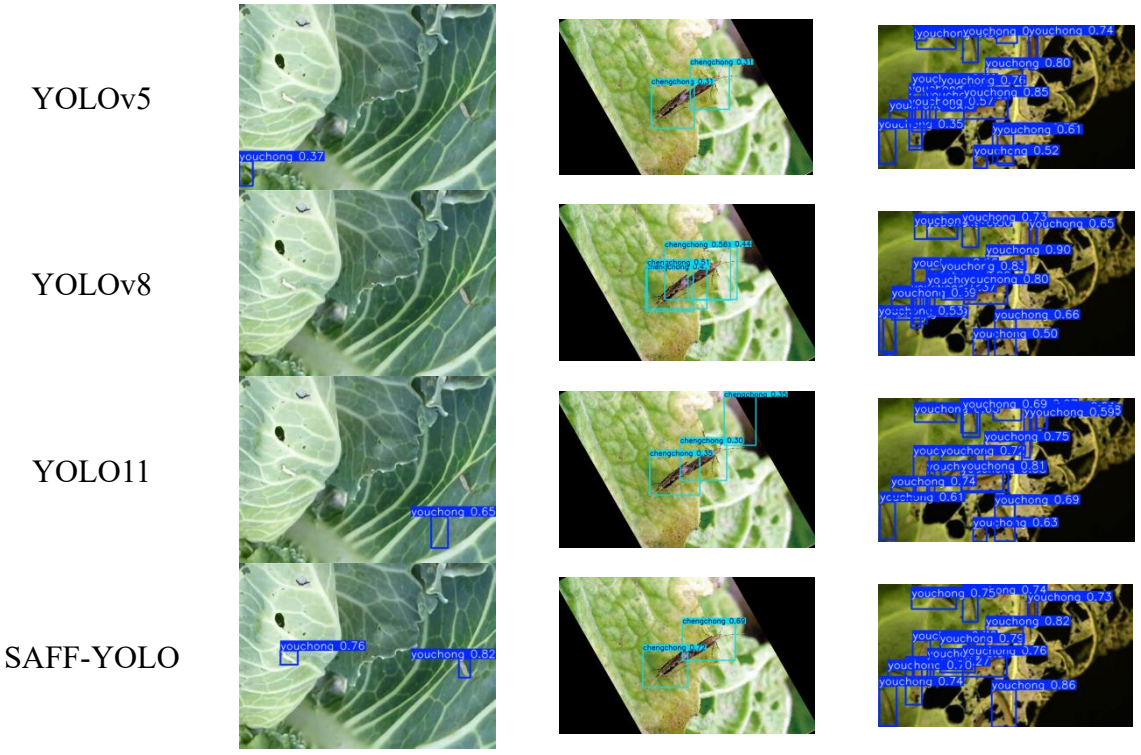

Fig. 10 - Target detection results of different models for the little Chinese cabbage diamondback moth

SAFF-YOLO: Diamondback Moth Pest Detection System

This study designed and implemented a visualization detection system for diamondback moths using the PyQt framework, as illustrated in Fig.11. PyQt integrates Python's concise syntax with Qt's robust functionality, offering comprehensive GUI components, cross-platform compatibility, and efficient signal-slot mechanisms. This architecture ensures consistent and stable user interfaces across diverse operating systems.

The application implements two core functionalities:

1) Pest image import module: Supports four input modalities: single image import, batch directory import, video file processing, and real-time camera-based detection.

2) Detection visualization interface: Presents detection metadata including inference duration, target count, pest classification, confidence scores, and bounding box coordinates. Post-detection, the system generates a comprehensive report listing target serial numbers, file paths, categories, confidence values, and positional coordinates.

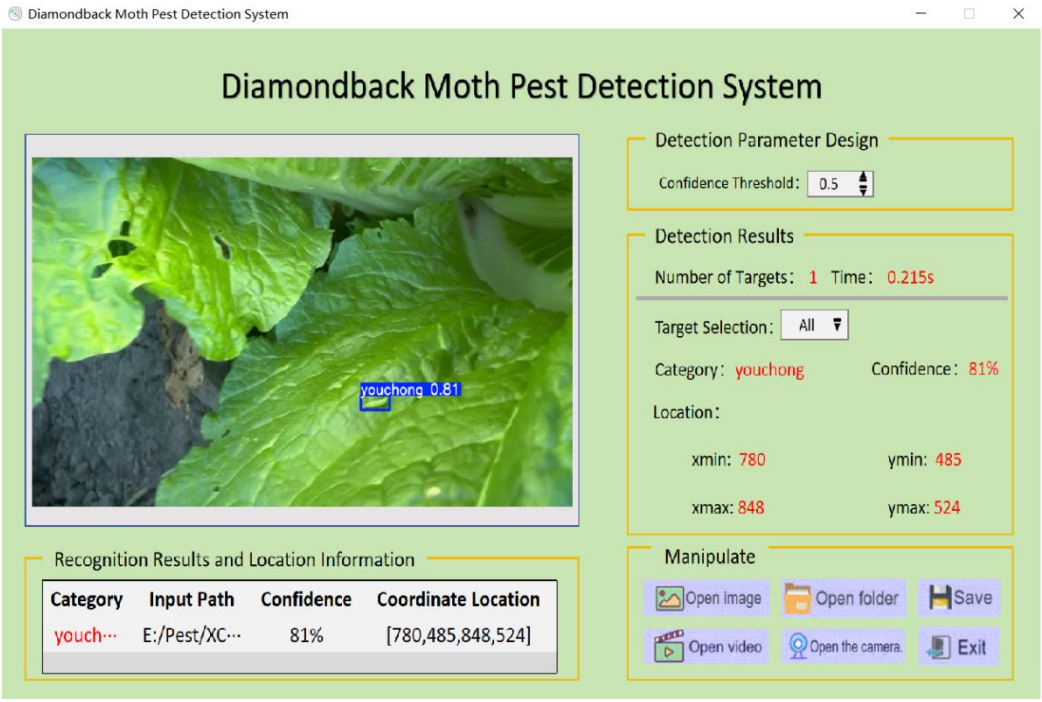

Fig. 11 - Diamondback Moth Pest Detection System

CONCLUSIONS

This study addresses key challenges in Chinese cabbage pest detection—including scale variation, target density, and phenotypic feature extraction difficulties—by proposing SAFF-YOLO: an efficient algorithm for diamondback moth detection on Chinese cabbage plants. The architecture modifies the YOLO11 framework through: (1) replacing the baseline loss function with UIoU to enhance discriminative learning of moth characteristics; (2) integrating AKConv lightweight convolutions and embedding SHSA attention within the C2PSA module to strengthen feature extraction while reducing parameters; (3) implementing FreqFusion as the neck network's upsampling operator for dynamic computation allocation and precise spatial localization; and (4) incorporating the FASFFHead for multi-scale detection capability enhancement. Experimental results demonstrate SAFF-YOLO's 11 MB model size reduction and 8.8 GFLOPs computation decrease versus YOLO11, while achieving 7.4% precision, 8.0% recall, and 8.4% mAP improvements—confirming superior efficiency and accuracy in cabbage moth detection. This approach enables precise localization of infestations for timely intervention in cabbage cultivation, offering valuable references for pest management in related agricultural systems.

ACKNOWLEDGEMENTS

This research was funded by Heilongjiang Province Natural Science Foundation Joint Guidance Project (No. LH2023E106), Heilongjiang Province "Double First-Class" Discipline Collaborative Innovation Achievement Project (No. LJGXCG2023-045), China University Industry-University-Research Innovation Foundation Funded Project (No.2023RY059) and Heilongjiang Province Key Research and Development Program Major Project (No.2023ZX01A06).

REFERENCES

[1] Ahmed, M. A., Cao, H. H., Jaleel, W., Amir, M. B., Ali, M. Y., Smagghe, G., Liu, T. X., (2022). Oviposition preference and two-sex life table of Plutella xylostella and its association with defensive enzymes in three Brassicaceae crops. *Crop Protection*, Vol. 151, pp. 105816, England.

[2] Ali, F., Qayyum, H., Iqbal, M. J., (2023). Faster-PestNet: A Lightweight deep learning framework for crop pest detection and classification. *IEEE Access*, Vol. 11, pp. 104016-104027, United States.

[3] Chakrabarty, S., Shashank, P.R., Deb, C.K., Haque, M.A., Thakur, P., Kamil, D., Marwaha, S., Dhillon, M.K., (2024). Deep learning-based accurate detection of insects and damage in cruciferous crops using YOLOv5. *Smart Agricultural Technology*, Vol. 9, pp. 100663, Netherlands.

[4] Chen, L., Fu, Y., Gu, L., Yan, C., Harada, T., Huang, G., (2024). Frequency-aware feature fusion for dense image prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 10763-10780, United States.

[5] Chen, Y. X., Tian, H. J., Wei, H., Zhan, Z., Huang, Y., (2011). Morphological Identification of Sexes in Larvae, Pupae, and Adults of the Diamondback Moth (*Plutella xylostella*) (小菜蛾幼虫、蛹和成虫的雌雄形态识别). *Fujian Journal of Agricultural Sciences*, Vol. 26, pp. 4, Fujian/China.

[6] Dananjayan, S., Tang, Y., Zhuang, J., Hou, C. and Luo, S., (2022). Assessment of state-of-the-art deep learning based citrus disease detection techniques using annotated optical leaf images. *Computers and Electronics in Agriculture*, Vol. 193, pp. 106658, England.

[7] Dongfang, Q. I. U., (2024). Research on locust target detection algorithm based on YOLO V7-MOBILENETV3-CA. *INMATEH-Agricultural Engineering*, Vol. 74, pp. 283-292, Romania.

[8] Fragoso, J., Silva, C., Paixão, T., Alvarez, A. B., Júnior, O. C., Florez, R., Palomino-Quispe, F., Savian, L. G., Trazzi, P. A., (2025). Coffee-Leaf Diseases and Pests Detection Based on YOLO Models. *Applied Sciences*, Vol. 15(9), pp. 5040, Romania.

[9] Hou, J., Yang, C., He, Y., Hou, B., (2023). Detecting diseases in apple tree leaves using FPN–ISResNet–Faster RCNN. *European Journal of Remote Sensing*, Vol. 56, pp. 2186955, Italy.

[10] Hu, G. Y., Mitchell, E. R., Okine, J. S., (1997). Diamondback moth (Lepidoptera: Plutellidae) in cabbage: Influence of initial immigration sites on population distribution, density and larval parasitism. *Journal of Entomological Science*, Vol. 32, pp. 56-71, United States.

[11] Hussain, M., Gao, J., Bano, S., Wang, L., Lin, Y., Arthurs, S., Qasim, M., Mao, R., (2020). Diamondback moth larvae trigger host plant volatiles that lure its adult females for oviposition. *Insects*, Vol.11, pp.725, Switzerland.

[12] Li, Z., Feng, X., Liu, S. S., You, M., Furlong, M. J., (2016). Biology, ecology, and management of the diamondback moth in China. *Annual review of entomology*, Vol. 61, pp. 277-296, United States.

[13] Li, Z., Zalucki, M. P., Yonow, T., Kriticos, D. J., Bao, H., Chen, H., Hu, Z., Feng, X., Furlong, M. J., (2016). Population dynamics and management of diamondback moth(Plutella xylostella) in China:the relative contributions of climate, natural enemies and cropping patterns. *Bulletin of Entomological Research*, Vol. 106, pp. 197-214, England.

[14] Liang, Q., Zhao, Z., Sun, J., Jiang, T., Guo, N., Yu, H., Ge, Y., (2024). Multi-target detection method for maize pests based on improved YOLOv8. *INMATEH-Agricultural Engineering*, Vol. 73, pp. 227-238, Romania. DOI: https://doi.org/10.35633/inmateh-73-19

[15] Lippi, M., Bonucci, N., Carpio, R. F., Contarini, M., Speranza, S., Gasparri, A., (2021). A yolo-based pest detection system for precision agriculture. In 2021 29th Mediterranean Conference on Control and Automation (MED), pp. 342-347, Puglia/Italy

[16] Liu, J., Wang, X., (2020). Tomato diseases and pests detection based on improved Yolo V3 convolutional neural network. *Frontiers in plant science*, Vol. 11, pp. 898, Switzerland.

[17] Liu, S., Huang, D., Wang, Y., (2019). Learning spatial fusion for single-shot object detection. arxiv 2019. *arxiv preprint arxiv*:1911.09516.

[18] Luo, X., Cai, Z., Shao, B., Wang, Y., (2024). Unified-IoU: For High-Quality Object Detection. *arxiv preprint arxiv:* 2408.06636, 2024.

[19] Lyu, Z., Jin, H., Zhen, T., Sun, F., Xu, H., (2021). Small object recognition algorithm of grain pests based on SSD feature fusion. *IEEE Access*, Vol. 9, pp. 43202-43213, United States.

[20] Rahman, M. M., Zalucki, M. P., Furlong, M.J., (2019). Diamondback moth egg susceptibility to rainfall: effects of host plant and oviposition behavior. *Entomologia Experimentalis et Applicata*, Vol. 167, pp. 701-712, England.

[21] Ritonga, F. N., Gong, Z., Zhang, Y., Wang, F., Gao, J., Li, C., Li, J., (2024). Exploiting Brassica rapa L. subsp. pekinensis Genome Research. *Plants*, Vol. 13, pp. 2823, Switzerland.

[22] Shehzad, M., Bodlah, I., Siddiqui, J. A., Bodlah, M. A., Fareen, A. G. E., Islam, W, (2023). Recent insights into pesticide resistance mechanisms in Plutella xylostella and possible management strategies. *Environmental Science and Pollution Research*, Vol. 30, pp. 95296-95311, Germany.

[23] Shi, H., Liu, C., Wu, M., Zhang, H., Song, H., Sun, H., Li, Y., Hu, J., (2025). Real-time detection of Chinese cabbage seedlings in the field based on YOLO11-CGB. *Frontiers in Plant Science*, Vol. 16, pp. 1558378, Switzerland.

[24] Slim, S.O., Abdelnaby, I.A., Moustafa, M.S., Zahran, M.B., Dahi, H.F. and Yones, M.S., (2023). Smart insect monitoring based on YOLOV5 case study: Mediterranean fruit fly Ceratitis capitata and Peach fruit fly Bactrocera zonata. *The Egyptian Journal of Remote Sensing and Space Sciences*, Vol. 26, pp. 881-891, Egypt.

[25] Song, L., Liu, M., Liu, S., Wang, H., Luo, J., (2023). Pest species identification algorithm based on improved YOLOv4 network. *Signal, Image and Video Processing*, Vol. 17, pp. 3127-3134, England.

[26] Tian, Y., Wang, S., Li, E., Yang, G., Liang, Z., Tan, M., (2023). MD-YOLO: Multi-scale Dense YOLO for small target pest detection. *Computers and Electronics in Agriculture*, Vol. 213, pp. 108233, England.

[27] Teixeira, A.C., Morais, R., Sousa, J.J., Peres, E. and Cunha, A., (2023). Using deep learning for automatic detection of insects in traps. *Procedia Computer Science*, Vol. 219, pp.153-160, Netherlands.

[28] Wang, R. F., Su, W. H., (2024). The Application of Deep Learning in the Whole Potato Production Chain: A Comprehensive Review. *Agriculture*, Vol. 14, pp. 1225, Switzerland.

[29] Wang, X., Shrivastava, A., Gupta, A., (2017). A-fast-RCNN: Hard positive generation via adversary for object detection. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2606-2615, HI/USA

[30] Wen, C., Chen, H., Ma, Z., Zhang, T., Yang, C., ... Chen, H., (2022). Pest-YOLO: A model for large-scale multi-class dense and tiny pest detection and counting. *Frontiers in Plant Science*, Vol. 13, pp. 973985, Switzerland.

[31] Wu, D., Fang, C., Zheng, X., Liu, J., Wang, S., Huang, X., (2024). AMW-YOLOv8n: Road Scene Object Detection Based on an Improved YOLOv8. *Electronics*, Vol. 13, pp. 4121-4121, Switzerland.

[32] Yun, S., Ro, Y., (2024). Shvit: Single-head vision transformer with memory efficient macro design. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5756-5767, Seattle/USA

[33] Zhai, S., Shang, D., Wang, S., Dong, S., (2020). DF-SSD: An improved SSD object detection algorithm based on DenseNet and feature fusion. *IEEE Access*, Vol. 8, pp. 24344-24357, United States.

[34] Zhang, H., Hu, J., Shi, H., Liu, C., Wu M., (2024). Precision Target Spraying System Integrated with Remote Deep Learning Recognition Model for Cabbage Plant Centers (融合远端深度学习识别模型的白菜株心精准对靶喷雾系统). *Smart Agriculture,* Vol. 6, pp. 85-95, Heilongjiang/China.

[35] Zhang, X., Song, Y., Song, T., Yang, D., Ye, Y., Zhou, J., Zhang, L., (2023). AKConv:convolutional kernel with arbitrary sampled shapes and arbitrary number of parameters. *arXiv* :2311.11587.

[36] Zhang, S., Liu, Z., Chen, Y., Jin, Y., Bai, G., (2023). Selective kernel convolution deep residual network based on channel-spatial attention mechanism and feature fusion for mechanical fault diagnosis. *ISA transactions*, Vol. 133, pp. 369-383, United States.