

Article

HFU-YOLO: A Method for Detecting Underground Pests in Complex Soil Backgrounds

Zhen Wang ^{1,2,†}, Zhonghua Miao ^{1,†} , Zhichong Wang ² , Mengdi Xu ², Jun Shi ^{1,2}, Changyuan Zhai ²  and Si Li ^{2,*}

¹ School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China; ghz01524@163.com (Z.W.); zhzmiao@shu.edu.cn (Z.M.); string2002@163.com (J.S.)

² Intelligent Equipment Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; wangzc@nercita.org.cn (Z.W.); mengdixu05@163.com (M.X.); zhaicy@nercita.org.cn (C.Z.)

* Correspondence: lis@nercita.org.cn; Tel.: +86-10-5150-3886

† These authors contributed equally to this work.

Abstract

Underground pests are difficult to monitor because of their hidden activity, and traditional manual surveys are inefficient. Computer vision provides a promising approach for efficient detection; however, pest boundaries and textures in complex soil backgrounds can be obscured or confounded by soil particles, residual roots, shadows, and other factors, thereby reducing detection accuracy. In this study, an image dataset of grubs, cutworm larvae, and cutworm pupae was constructed, and a lightweight detection model, HFU-YOLO, was proposed using YOLO11n as the baseline. A Haar Wavelet Downsampling (HWD) module was introduced to reduce the loss of shallow edge and texture information. A Lightweight Frequency-Adaptive Dilated Convolution (LFADC) module was designed to enhance target feature representation and reduce soil background interference. A UGS-Lite auxiliary loss function was adopted to improve the localization stability of weak-boundary targets. The results showed that HFU-YOLO achieved Precision, Recall, mAP@0.5, and mAP@0.5:0.95 values of 77.6%, 78.3%, 83.5%, and 44.1% on the test set, respectively, representing improvements of 4.2, 3.2, 6.1, and 6.1 percentage points over YOLO11n. With 2.59 M parameters and 6.2 GFLOPs, HFU-YOLO provides a reference for intelligent recognition of underground pests in complex soil backgrounds.

Keywords: underground pest detection; complex soil backgrounds; soil image analysis; object detection; weak-boundary localization



Academic Editor: Roberto Alves Braga Júnior

Received: 12 June 2026

Revised: 9 July 2026

Accepted: 14 July 2026

Published: 17 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Underground pests are an important biotic stress affecting crop growth and yield formation [1,2]. Grubs, cutworm larvae, and other underground pests mostly inhabit soil or near-surface soil layers, where they feed on crop roots, basal stems, and tender underground tissues, causing plant wilting, growth inhibition, missing seedlings, broken ridges, and, in severe cases, plant death and yield loss. Compared with aboveground pests, underground pests have hidden activity, are difficult to observe directly, and show more uneven field distributions. Traditional pest surveys usually rely on soil sampling, soil excavation, trapping surveys, and manual empirical judgment [3]. These procedures are time-consuming and labor-intensive, making it difficult to continuously obtain pest occurrence information. Therefore, research on intelligent recognition of underground

pests is of great significance for improving pest survey efficiency, determining appropriate control timing, and reducing production losses.

In recent years, the development of image sensors, automatic monitoring equipment, and deep learning has provided new technical pathways for automatic pest recognition [4–6]. Image-based pest recognition and automatic monitoring systems can reduce manual workload and improve the timeliness of pest surveys [7–9]. In agricultural visual recognition and detection tasks, deep learning can automatically extract target features from images and has been widely used in pest classification, object detection, and pest image analysis. Among these methods, the YOLO series adopts a single-stage detection framework and has the advantages of fast inference, end-to-end optimization, and convenient deployment [10–12]. In recent years, it has been widely applied to agricultural pest and disease detection scenarios involving citrus, maize, tea plants, fruit trees, and other crops [13–15]. Existing studies on visual pest detection have mainly focused on aboveground visible scenes, such as leaves, fruits, trapping boards, and crop canopies, with the detected targets mostly being exposed or semi-exposed pests. In contrast, image detection of underground pests under complex soil backgrounds remains insufficient.

More broadly, recent studies on intelligent agricultural vision have extended beyond conventional pest detection to pest and disease classification, domain adaptation, lightweight model design, and field-oriented crop-protection applications. For example, resource-efficient cotton disease and pest classification models have provided useful references for lightweight feature extraction and model deployment in agricultural scenarios [16]. Joint adversarial transfer learning has been used to reduce the influence of domain differences in pest image recognition [17]. For field pest monitoring, JassidNet was proposed as a quantization-aware lightweight phenotyping framework for high-throughput cotton jassid detection and counting. By combining lightweight detection, standardized preprocessing, pseudo-label iteration, and quantization-aware model design, this method demonstrated the application value of lightweight pest-recognition models in complex leaf backgrounds and resistance screening [18]. In addition, deep learning has been integrated with field operation equipment, such as mechanical weeding systems, to improve the perception and decision-making ability of crop-protection machinery [19]. These studies indicate that agricultural pest-recognition methods are developing toward lightweight modeling, cross-domain adaptability, complex-environment perception, and practical field deployment, providing useful references for underground pest detection under complex soil backgrounds.

However, underground pest detection differs from aboveground pest detection, crop disease recognition, and weed recognition tasks. In the above studies, the targets are mainly located on leaves, fruits, crop canopies, or field surfaces, and their visual features are usually more exposed and structurally complete. In contrast, underground pests are located in soil or near-surface soil layers, where their visible features are easily affected by soil particles, roots, cracks, shadows, moisture conditions, and partial soil coverage. Therefore, underground pest detection is closely related to broader agricultural vision and intelligent crop-protection research in terms of image-based perception and lightweight model design, but it presents distinct challenges in imaging medium, background interference, target-detail preservation, and weak-boundary localization. These differences further motivate the following analysis of the main detection difficulties of underground pests in complex soil backgrounds.

The difficulty of underground pest detection is mainly reflected in three aspects: target scale, background interference, and boundary localization. Underground pests are usually small targets, and their body edges and local textures are easily lost after repeated sampling in the backbone network, resulting in insufficient small-target representation [20,21]. Exist-

ing studies on small pest detection have shown that excessive feature compression, loss of shallow fine-grained features, and ineffective multi-scale feature fusion are major factors limiting improvements in small pest detection accuracy [22–24]. Meanwhile, soil particles, cracks, residual roots, and shadows introduce substantial background interference, and pest contours are similar to soil textures. Targets such as cutworm larvae, whose body color is close to that of soil, are especially prone to confusion with the background [25–27]. Under soil-covered or semi-covered conditions, the visible target area is limited, the boundary is blurred, the morphology is incomplete, and the uncertainty of bounding-box regression increases markedly. Studies on agricultural pest detection have shown that small targets, incomplete targets, complex backgrounds, and occlusion can easily lead to missed detections, false detections, and unstable localization [28–30]. Although pupae do not cause feeding damage directly, detecting them is still valuable because it helps infer pest developmental stages and provides a reference for adult emergence monitoring and control timing.

For small-target detection in complex backgrounds, existing studies have improved detection models through multi-scale feature fusion, attention mechanisms, lightweight detection heads [31], and loss function optimization [32], achieving certain progress in agricultural pest detection tasks [33–35]. However, it remains difficult to simultaneously preserve target details, suppress soil-background interference, and stabilize weak-boundary localization in underground pest images. Haar Wavelet Downsampling (HWD) can preserve low-frequency structures and high-frequency details during downsampling through wavelet decomposition, providing a reference for reducing the loss of edge and texture information in small targets [36]. Frequency-Adaptive Dilated Convolution (FADC) can regulate dilated convolution feature responses from the frequency dimension, offering a useful idea for suppressing complex background interference [37]. Distributed bounding-box regression methods can represent boundary position uncertainty using probability distributions, providing a basis for localization optimization of weak-boundary and occluded targets [38].

To address these problems, this study proposed HFU-YOLO, an underground pest detection method for complex soil backgrounds, using YOLO11n as the baseline. The method improved the model from three aspects: shallow detail preservation, complex background feature enhancement, and weak-boundary localization constraint. Specifically, an HWD module was introduced to reduce the loss of edge and texture information during downsampling. Based on the frequency-adaptive idea of FADC, an LFADC module was designed to enhance pest-related feature representation and reduce soil background interference. In addition, a UGS-Lite auxiliary loss was adopted to constrain the localization distribution of positive samples, thereby improving the localization stability of weak-boundary targets. This study aimed to improve underground pest image detection and provide a reference for constructing lightweight recognition models for such targets.

2. Materials and Methods

2.1. Dataset Construction

2.1.1. Image Acquisition and Scene Construction

To investigate visual detection of underground pests under complex soil backgrounds and different occlusion conditions, an indoor soil-box imaging platform with dimensions of 50 cm × 30 cm × 40 cm was constructed. Three soil media were placed in the box to control soil particle structure, surface reflection, and pore characteristics: substrate soil without perlite, substrate soil with perlite, and field soil. Images were acquired using a JC600-Z (manufactured by Foshan Shunde Wende Ye Trading Firm, Foshan, China) industrial endoscope with a resolution of 1920 × 1080, and a supplementary lighting module was integrated at the front end of the device. During image acquisition, the imaging distance

and viewing angle were adjusted within a small range according to the target position and soil occlusion condition. The endoscope position was slightly varied among different scenes, while the built-in supplementary light and exposure setting were kept consistent. Grubs, cutworm larvae, and cutworm pupae were selected as the experimental targets (Figure 1).

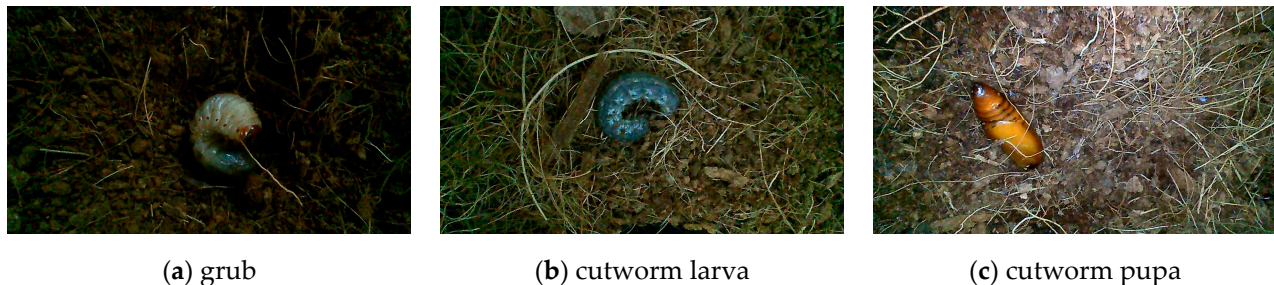


Figure 1. Representative images of three types of underground pests.

A total of 49 imaging scenarios were designed. By regulating the degree of soil coverage, exposed body parts, target posture, soil moisture, root clusters, dark pore regions, soil block edges, and other factors, typical imaging conditions were artificially constructed, including partial soil-cover occlusion, blurred pest boundaries, and pest-soil texture confusion. After image acquisition was completed for each scenario, the scene was reset by turning and re-laying the soil, adjusting the target posture, replacing local interference structures, and changing moisture conditions. After acquisition, images with severe defocus, abnormal exposure, uncertain target identification, or near-duplicate content were removed, and the remaining valid samples were retained for annotation.

2.1.2. Dataset Construction and Partitioning

All samples were manually annotated using X-AnyLabeling (version 3.3.3), and the annotations were saved in the YOLO object detection format. The dataset consisted of three target classes, namely grub, cutworm larva, and cutworm pupa, together with pure-background negative samples. For these negative samples, no pest bounding boxes were annotated, and empty label files were used in the YOLO format. They were included in the training, validation, and test sets together with positive images, so that the model could learn background regions with pest-like soil structures and reduce false-positive detections. According to the degree of background clutter, the images were divided into simple-background and complex-background samples. In simple-background samples, the pest targets were clear and there was little interference from soil debris. In complex-background samples (Figure 2), multiple interference factors were incorporated, including different soil media, soil moisture levels, and perlite reflections.

A total of 4795 valid images were finally selected, including 1372 simple-background samples and 3423 complex-background samples. The dataset was divided at a ratio of 8:1:1 at the imaging-scenario level, yielding 3859 training images, 481 validation images, and 455 test images. Images from the same constructed imaging scenario were assigned to only one subset to avoid near-duplicate or highly similar images appearing simultaneously in the training, validation, and test sets. Each imaging scenario was defined according to the pest individual, soil medium, moisture condition, occlusion level, and local background interference. The same partitioning scheme was used for baseline network selection, ablation experiments, and model comparison experiments. The numbers of images, annotated instances, negative images, and class-specific instances in each subset are shown in Table 1.

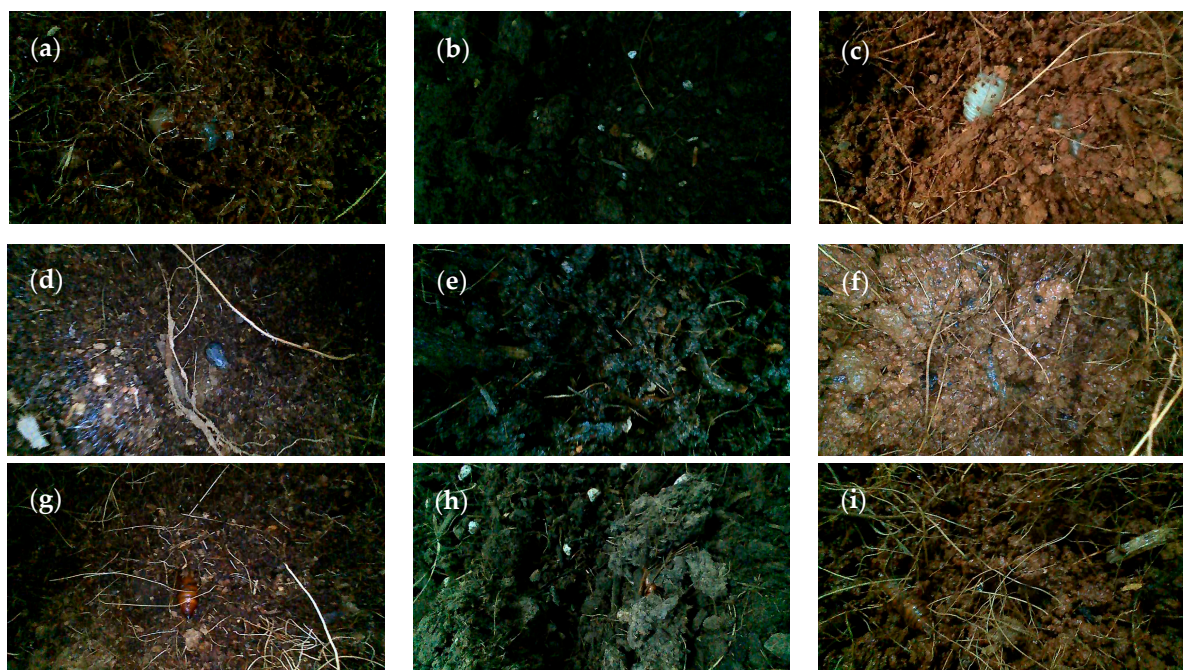


Figure 2. Sample images of three types of underground pests under different background conditions. Note: (a–c) grubs; (d–f) cutworm larvae; (g–i) cutworm pupae.

Table 1. Partition statistics of the underground pest object detection dataset.

Subset	Images	Instances	Negative Images	Grubs	Cutworm Larvae	Cutworm Pupae
Training	3859	5445	427	1653	2013	1779
Validation	481	525	53	152	236	137
Test	455	510	55	171	193	146
Total	4795	6480	535	1976	2442	2062

In addition to the original training, validation, and test sets, an independent cross-condition test set was constructed to evaluate the robustness of the trained models under unseen background conditions. This additional test set contained 275 newly acquired images, including 19 pest-free negative background images and 294 annotated pest instances. The images were acquired under an unseen soil-background condition with peanut-shell residues as additional non-target interference, and the soil background was different from the soil media used in the original dataset. This test set was not used for model training, validation, hyperparameter selection, or model structure design, and was used only for the final cross-condition evaluation.

2.2. Construction of the Underground Pest Recognition Model

YOLO11n was used as the baseline network for the underground pest detection task. To address the problems of small target scale, blurred boundaries, weak texture responses, and partial occlusion in complex soil backgrounds, a lightweight detection model, HFU-YOLO, was constructed. The three components in HFU-YOLO have different sources and roles. HWD was directly adopted from the existing Haar wavelet downsampling method [36] and used to replace shallow Conv-based downsampling layers, rather than being proposed as a new downsampling operation. LFADC was developed as a lightweight task-specific modification of the FADC design [37], retaining the frequency-selection idea while simplifying the original frequency-adaptive dilated convolution structure. UGS-Lite was newly introduced in this study as an auxiliary entropy regularization term based on the

localization-distribution idea of DFL/GFL [38]. It is used only during training to improve the localization stability of weak-boundary targets.

2.2.1. HWD: An Edge- and Texture-Preserving Downsampling Module for Underground Pests

In the shallow backbone network of YOLO11n, conventional Conv-based downsampling can easily weaken the edge and local texture responses of small-scale targets. Following the HWD method [36], this study replaced two shallow Conv downsampling structures with Haar Wavelet Downsampling, and the module structure is shown in Figure 3. For an input feature map x , HWD first performs the Haar wavelet transform to obtain four sub-band features:

$$\{LL, LH, HL, HH\} = \text{HWT}(x), \quad (1)$$

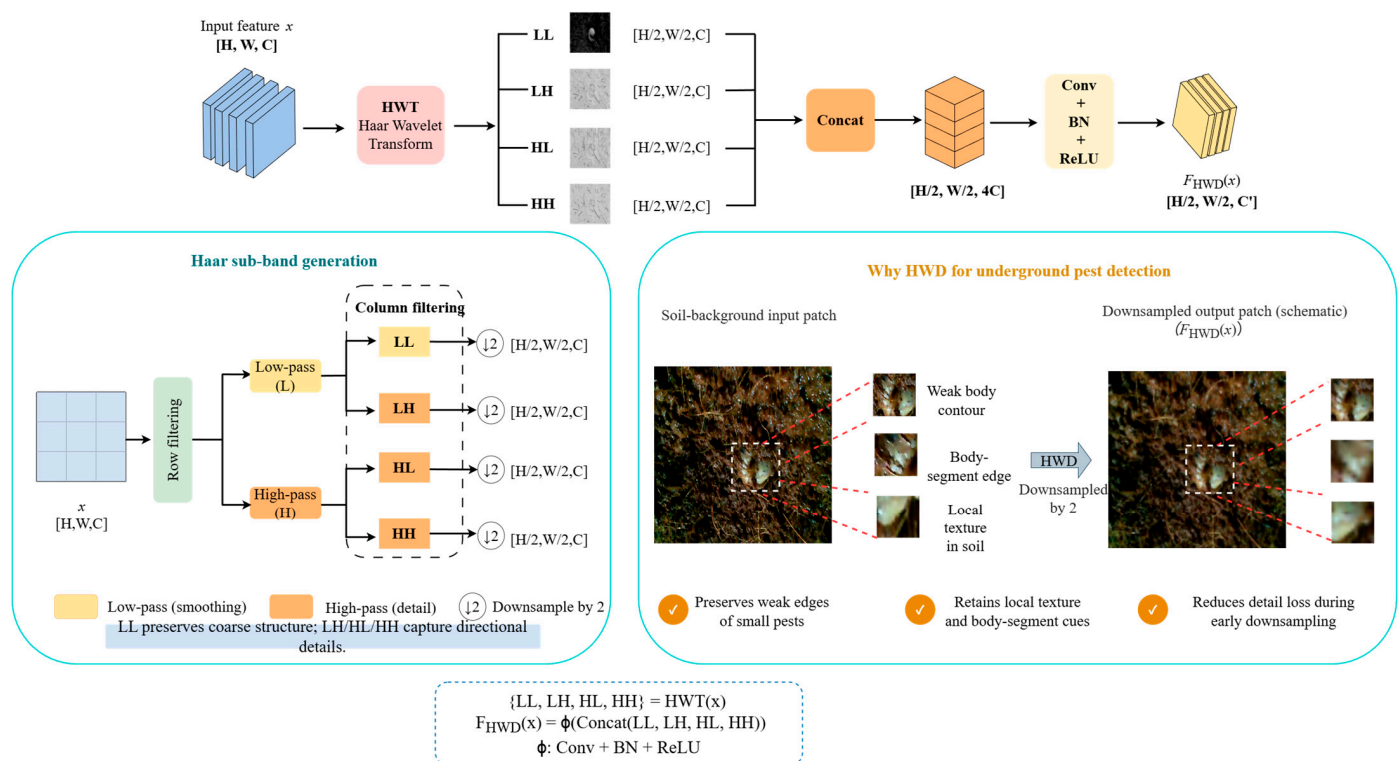


Figure 3. Schematic structure of the HWD module.

The LL sub-band preserves the main target structure and low-frequency contours, whereas the LH, HL, and HH sub-bands represent high-frequency edge and texture responses in different directions. This enables HWD to maintain both the overall pest morphology and local discriminative details.

Subsequently, HWD concatenates the four sub-bands along the channel dimension and fuses them through convolution, batch normalization, and an activation function:

$$F_{\text{HWD}}(x) = \phi(\text{Concat}(LL, LH, HL, HH)), \quad (2)$$

where $\text{Concat}(\cdot)$ denotes the channel-wise concatenation operation, and $\phi(\cdot)$ denotes the feature fusion process composed of convolution, batch normalization, and an activation function.

The final output feature is given by:

$$F_{\text{HWD}}(x) \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times C'}, \quad (3)$$

where C' denotes the number of output channels after fusion.

This process preserves contour, edge, and texture information related to underground pests while reducing the spatial resolution. Therefore, HWD helps alleviate missed detections and localization offsets of weak-edge and small-scale underground pests during shallow downsampling.

2.2.2. LFADC: A Feature Enhancement Module for Complex Soil Backgrounds

In complex soil backgrounds, soil particles, cracks, residual roots, dark pore regions, perlite reflections, shadows, and other factors can easily be confused with pest features. Inspired by FADC [37], this study designed LFADC as a lightweight feature enhancement module for underground pest detection. The module consists of three stages: frequency selection, multi-dilation-rate feature fusion, and lightweight channel calibration. The structure is shown in Figure 4.

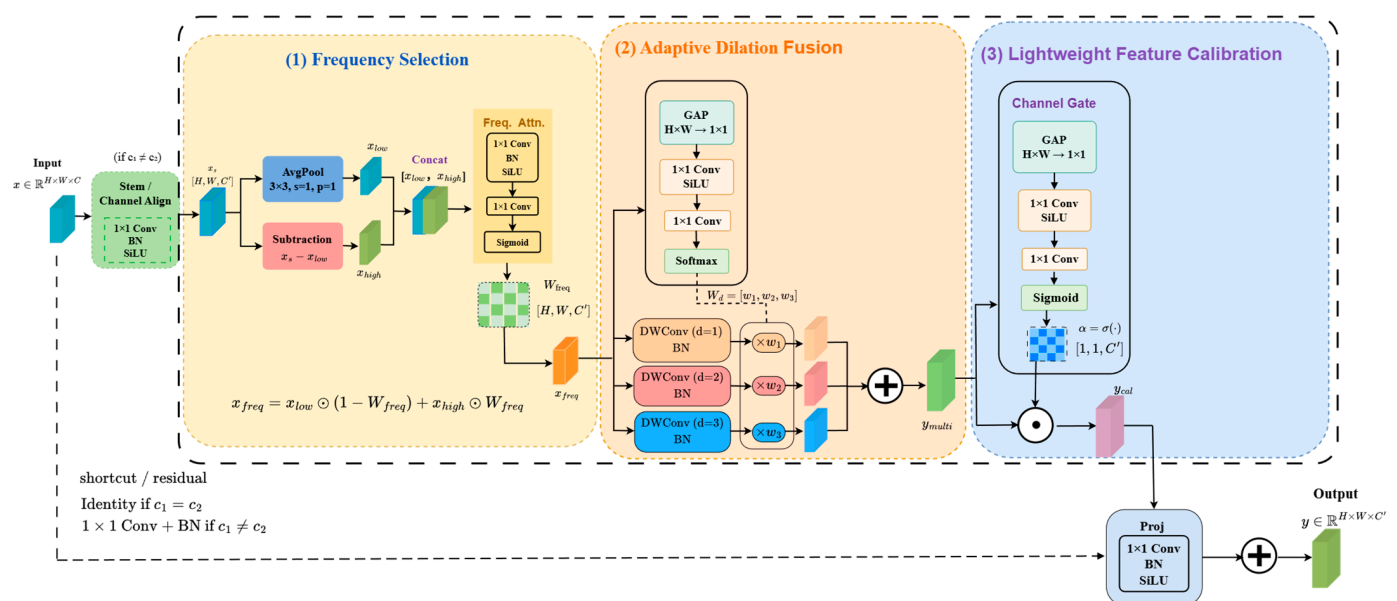


Figure 4. Schematic structure of the LFADC module.

Let the input feature be $x \in \mathbb{R}^{H \times W \times C}$. After stem convolution, the intermediate feature is denoted as x_s . To distinguish structural information of underground pests from background detail responses, LFADC first obtains a low-frequency structural component through average pooling and then obtains a high-frequency detail component through residual subtraction:

$$x_{low} = \text{AvgPool}_{3 \times 3}(x_s), \quad (4)$$

$$x_{high} = x_s - x_{low}, \quad (5)$$

where x_{low} describes the overall contour of underground pests and local background trends, whereas x_{high} reflects target edges, local textures, and fine-grained variations.

To avoid excessive amplification of invalid high-frequency responses caused by soil particles, cracks, reflective spots, and other factors, LFADC generates high-frequency weights through a frequency attention branch and adaptively fuses low-frequency structural information with high-frequency detail information:

$$W_{freq} = \sigma(f_{freq}(\text{Concat}(x_{low}, x_{high}))), \quad (6)$$

$$x_{freq} = x_{low} \odot (1 - W_{freq}) + x_{high} \odot W_{freq}, \quad (7)$$

where $f_{\text{freq}}(\cdot)$ denotes the frequency attention branch, $\sigma(\cdot)$ denotes the Sigmoid function, and $\odot$ denotes element-wise multiplication.

This process enables the model to adjust the fusion ratio between structural information and detail information according to the frequency characteristics of local regions.

In the multi-dilation-rate feature fusion stage, LFADC sets depthwise separable dilated convolution branches with dilation rates of 1, 2, and 3, and predicts the branch weights through global average pooling and an MLP:

$$W_d = \text{Softmax}\left(\text{MLP}\left(\text{GAP}\left(x_{\text{freq}}\right)\right)\right), \quad (8)$$

$$y_{\text{multi}} = \sum_{d=1}^3 w_d \cdot \text{BN}\left(\text{Conv}_{\text{DW}}^d\left(x_{\text{freq}}\right)\right), \quad (9)$$

where $\text{GAP}(\cdot)$ denotes global average pooling, and W_d denotes the normalized weight of the branch with a different dilation rate. $\text{Conv}_{\text{DW}}^d(\cdot)$ denotes depthwise separable convolution with a dilation rate of d , and $\text{BN}(\cdot)$ denotes batch normalization.

This design enables the model to simultaneously exploit local texture details and broader contextual information, thereby improving its adaptability to underground pest targets with different scales, postures, and occlusion states.

In the lightweight channel calibration stage, LFADC uses global contextual information to generate channel weights, recalibrates the multi-scale fused features, and obtains the output feature through projection convolution and a residual connection:

$$\alpha = \sigma(\text{MLP}(\text{GAP}(y_{\text{multi}}))), \quad (10)$$

$$y = \text{Proj}(y_{\text{cal}}) + x_{\text{res}}, \quad (11)$$

where α denotes the channel attention weight, $\text{Proj}(\cdot)$ denotes the 1×1 projection convolution, and x_{res} denotes the residual branch feature.

Through the above design, LFADC achieves frequency information selection, multi-scale contextual fusion, and channel feature calibration with low computational cost, thereby enhancing target-related features of underground pests and reducing texture interference from complex soil backgrounds.

2.2.3. UGS-Lite: A Loss Function for Weak-Boundary Localization Distribution Constraint

Under soil-covered, semi-covered, or partially exposed conditions, the body boundaries of underground pests are easily confused with soil particles, residual roots, shadows, and dark pore regions, leading to unstable bounding-box regression. Existing GFL/DFL methods [38] represent bounding-box positions as discrete probability distributions. Based on DFL, this study introduced UGS-Lite to impose a lightweight constraint on the uncertainty of the boundary localization distribution for positive samples. For the s -th boundary direction, the localization probability distribution after Softmax normalization is obtained as follows:

$$p_{s,k} = \frac{\exp(z_{s,k})}{\sum_{j=0}^{K-1} \exp(z_{s,j})}, k = 0, 1, \dots, K-1, \quad (12)$$

where $p_{s,k}$ denotes the predicted probability of the s -th boundary direction in the k -th discrete bin.

The more blurred the boundary is, the flatter the localization probability distribution becomes, and the higher the uncertainty of boundary prediction. To constrain this uncertainty, information entropy was introduced to measure the dispersion degree of the localization distribution:

$$H(p_s) = -\sum_{k=0}^{K-1} p_{s,k} \log(p_{s,k}), \quad (13)$$

A larger entropy value indicates a more dispersed boundary localization distribution, whereas a smaller entropy value indicates a more concentrated localization response. In this study, entropy was calculated only for the boundary distributions of positive samples involved in regression, and its mean value was used as the UGS-Lite auxiliary constraint term:

$$L_{um} = \frac{1}{4N} \sum_{i \in \Omega} \sum_{s=1}^4 H(p_{i,s}), \quad (14)$$

where $p_{i,s}$ denotes the discrete localization probability distribution of the i -th positive sample in the s -th boundary direction.

This loss term encourages the boundary response to become moderately more concentrated by minimizing the average entropy, while avoiding a forced overly sharp distribution through a small auxiliary weight. Finally, UGS-Lite was added to the original DFL loss with a small weight:

$$L'_{dfl} = L_{dfl} + \lambda L_{um} \quad (15)$$

where L_{dfl} denotes the original localization distribution loss, L_{um} denotes the UGS-Lite auxiliary entropy constraint loss, and λ denotes the auxiliary loss weight.

During implementation, L_{um} was calculated only from the regression distributions of positive samples involved in bounding-box regression and was then added to the original DFL branch with a small weight. UGS-Lite participates in training only as a lightweight regularization term. It does not change the detection head structure or inference process, nor does it increase the computational burden during inference.

2.2.4. Overall Structure of the HFU-YOLO Model

Based on the above modules, HFU-YOLO was constructed within the YOLO11n framework, and its overall structure is shown in Figure 5. The model retains the basic backbone, neck, and three-scale detection head of YOLO11n, while introducing local improvements according to the feature hierarchy of the baseline network. HWD was inserted into two shallow downsampling stages because early downsampling directly affects the retention of pest contours and body-segment texture information. LFADC was inserted into the middle backbone layers because these layers contain both local texture responses and preliminary semantic information, which are needed to distinguish pest bodies from soil particles, roots, cracks, and shadows. In LFADC, candidate dilation rates of 1, 2, and 3 were selected to cover local texture, adjacent boundary information, and short-range contextual cues, respectively. Larger dilation rates were not used because they may introduce sparse sampling effects and irrelevant soil-background responses while increasing computational cost. UGS-Lite is used only as an auxiliary loss during training and does not change the detection head or inference process. Therefore, the improvement of HFU-YOLO is intended to match the imaging characteristics of underground pests rather than simply increase the number of architectural components.

2.3. Model Training

Model training and validation were implemented based on the Ultralytics framework. All models used the same dataset partitioning, input size, and training strategy to ensure the comparability of the results. The input image size was 640×640 , the number of training epochs was 300, the batch size was 32, and the early-stopping patience was set to 50. AdamW was used as the optimizer, with an initial learning rate of 0.001, a final learning rate factor of 0.01, and a weight decay coefficient of 0.0005. Model training was initialized using COCO pretrained weights, and the random seed was set to 42 for the main comparison experiments.

During training, data augmentation strategies, including color jittering, random rotation, random translation, random scaling, horizontal flipping, and Mosaic augmentation, were used to improve the adaptability of the model to variations in underground pest posture, local viewing angle, illumination, and soil-background disturbance. The augmentation parameters were set as follows: hsv_h = 0.015, hsv_s = 0.7, hsv_v = 0.4, rotation angle = 90°, translation ratio = 0.1, scaling ratio = 0.5, horizontal flipping probability = 0.5, and Mosaic probability = 1.0. Shear, perspective transformation, vertical flipping, MixUp, and Copy-Paste augmentation were not used. The same augmentation settings were used for all compared models.

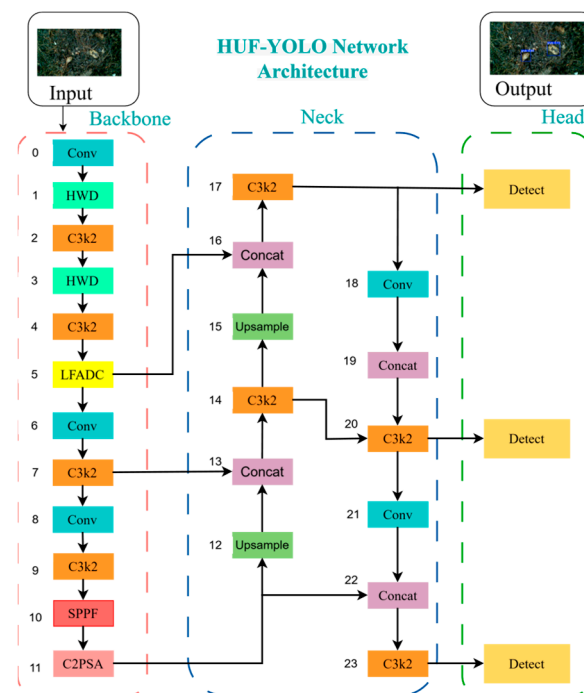


Figure 5. Overall network structure of the HUF-YOLO model.

2.4. Experimental Environment and Evaluation Metrics

The experimental environment was configured as follows: The operating system was Windows 11, the CPU was an Intel Core i5-12600KF, and the GPU was an NVIDIA GeForce RTX 5060 Ti with 16 GB of memory. The programming language was Python 3.11.14, and the deep learning framework was PyTorch 2.9.1 with CUDA 12.8. In addition, the object detection models in this study were constructed based on the Ultralytics 8.3.163 library.

To evaluate the detection performance of the improved model, Precision (P), Recall (R), and mean Average Precision (mAP) were used as the main accuracy metrics. The number of parameters (Params), floating-point operations (GFLOPs), inference time, and frames per second (FPS) were used to evaluate model complexity and detection speed. P, R, and mAP were calculated as follows:

$$P = \frac{C_{TP}}{C_{TP} + C_{FP}}, \quad (16)$$

$$R = \frac{C_{TP}}{C_{TP} + C_{FN}}, \quad (17)$$

$$AP = \int_0^1 P(R) dR, \quad (18)$$

$$mAP_{0.5} = \frac{1}{N} \sum_{i=1}^N AP_i(IoU = 0.5), \quad (19)$$

$$mAP_{0.5:0.95} = \frac{1}{10} \sum_{t \in \{0.50, 0.55, \dots, 0.95\}} mAP_t \quad (20)$$

where C_{TP} denotes the number of samples correctly predicted as positive by the model, C_{FP} denotes the number of samples incorrectly predicted as positive, and C_{FN} denotes the number of samples incorrectly predicted as negative. AP denotes average precision, AP_i denotes the average precision of the i -th class, and N denotes the number of detected classes, with $N = 3$. $mAP_{0.5}$ denotes the mean average precision at an intersection over union (IoU) threshold of 0.5. In addition, $mAP_{0.5:0.95}$ was introduced as a stricter supplementary metric for comprehensive evaluation.

For evaluating model lightweightness and detection speed, Params, GFLOPs, the average inference time per image (Time), and frames per second (FPS) were used for comprehensive assessment. All models were evaluated on the same test set under the same hardware and software conditions, with an input size of 640×640 , a batch size of 1, and FP32 precision. The inference speed was measured using the Ultralytics validation procedure, and the reported Time was averaged over all images in the test set. Time included image preprocessing, model inference, non-maximum suppression, and postprocessing. FPS was calculated as $1000/\text{Time}$.

3. Results

3.1. Detection Result Analysis

PR curves and normalized confusion matrices were used to analyze class-wise detection performance, as shown in Figure 6. Compared with YOLO11n, HFU-YOLO achieved higher AP values for all three pest categories, with the overall $mAP@0.5$ increasing from 77.4% to 83.5%. The normalized confusion matrices showed that HFU-YOLO reduced the confusion between cutworm larvae and pupae, indicating improved class-wise discrimination under complex soil backgrounds.

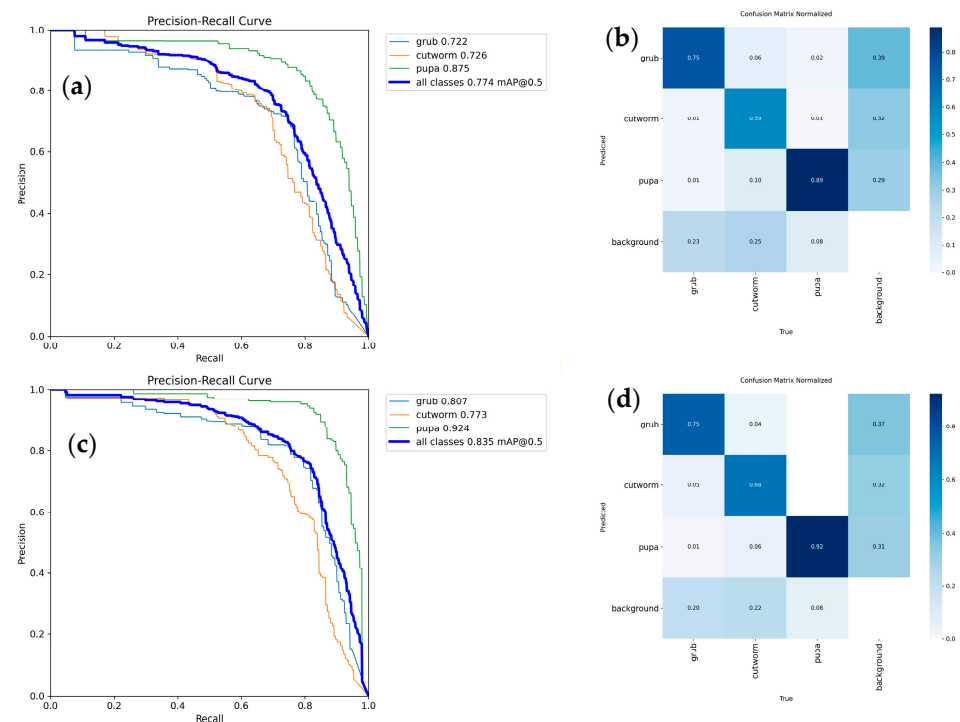


Figure 6. PR curves and normalized confusion matrices of YOLO11n and HFU-YOLO. Note: (a) PR curve of YOLO11n; (b) normalized confusion matrix of YOLO11n; (c) PR curve of HFU-YOLO; (d) normalized confusion matrix of HFU-YOLO.

Three typical samples from the test set were selected for visual comparison, corresponding to multi-target missed detection, weak-boundary target missed detection, and class misclassification scenarios, respectively (Figure 7). The results showed that, in the grub sample, YOLO11n detected only one target, whereas HFU-YOLO detected two targets. In the cutworm larva sample, YOLO11n missed the target, whereas HFU-YOLO correctly detected it. In the pupa sample, YOLO11n misclassified the target as cutworm larva, whereas HFU-YOLO correctly identified it as pupa. These results indicate that HFU-YOLO provides more stable detection performance for samples with complex backgrounds, weak boundaries, and easily confused classes.

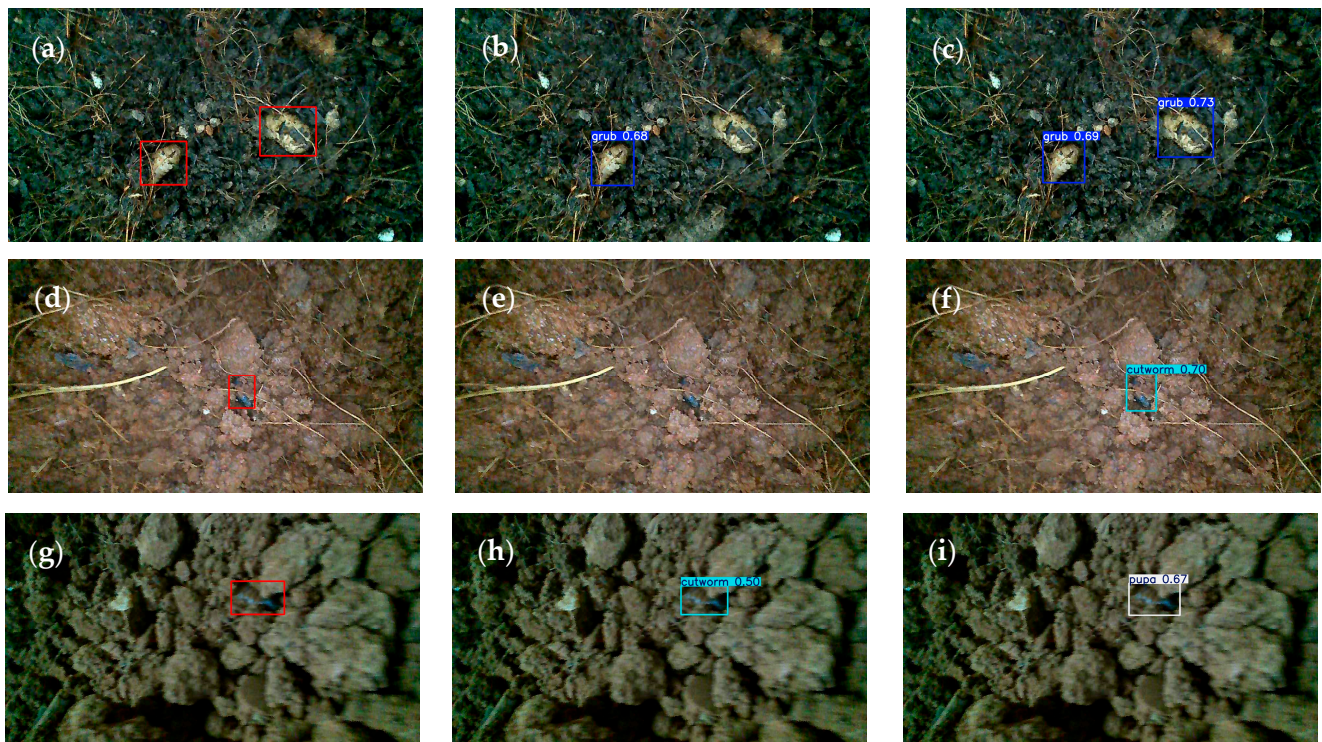


Figure 7. Comparative examples of detection results obtained by YOLO11n and HFU-YOLO in typical scenarios. Note: (a,d,g) manual annotation results; (b,e,h) YOLO11n detection results; (c,f,i) HFU-YOLO detection results. (a–c) grub samples; (d–f) cutworm larvae samples; (g–i) pupa samples.

Among the three pest categories, cutworm larva remained relatively difficult to detect because its body color and texture were close to moist soil. In the failure cases, missed detections mainly occurred when the visible body region was small or partially covered by soil, while localization errors were often observed around blurred boundaries. Negative background images were also used to examine false-positive detections in pest-free soil backgrounds. The false-positive cases mainly occurred in regions with pest-like textures, such as soil block edges, residual roots, dark pore regions, and shell-like debris. These observations indicate that the use of negative samples and background-interference suppression modules helped reduce invalid responses to pest-like background structures in complex soil backgrounds.

3.2. Comparative Experiments with Different Network Models

To evaluate the effectiveness of HFU-YOLO, YOLOv5n, YOLOv8n, YOLOv10n, YOLO11n, YOLO11s, YOLOv12n, and DGS-Tiny were selected for comparison. All models used the same dataset partitioning, input size, and training parameters, and were evaluated on the test set. The results are shown in Table 2.

Table 2. Comparative experimental results of different network models.

Model	P (%)	R (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Param (M)	GFLOPs	Time (ms)	FPS
YOLOv5n	77.8	73.1	79.5	41.9	2.50	7.1	7.60	131.66
YOLOv8n	79.8	73.8	79.2	40.0	3.01	8.1	11.21	89.20
YOLOv10n	77.7	71.4	78.3	42.4	2.27	6.5	7.62	131.17
YOLO11n	73.4	75.1	77.4	38.0	2.58	6.3	8.75	114.26
YOLO11s	76.0	73.2	80.2	41.2	9.41	21.3	26.20	38.17
YOLO12n	78.3	72.8	78.8	42.0	2.56	6.3	8.90	112.36
DGS-Tiny	72.5	75.7	78.8	36.3	4.30	9.8	3.10	322.58
HFU-YOLO	77.6	78.3	83.5	44.1	2.59	6.2	9.35	106.96

HFU-YOLO achieved mAP@0.5 and mAP@0.5:0.95 values of 83.5% and 44.1%, respectively, which were both 6.1 percentage points higher than those of YOLO11n. Compared with YOLOv5n, YOLOv8n, YOLOv10n, YOLO11s, and YOLOv12n, HFU-YOLO maintained advantages in both mAP metrics. The comparison with YOLOv12n further indicates that the proposed task-specific design remains effective among recent lightweight YOLO-based detectors. HFU-YOLO had 2.59 M parameters, 6.2 GFLOPs, and an average inference time of 9.35 ms per image, indicating that it achieved a good balance among detection accuracy, model size, and inference efficiency.

3.3. Cross-Condition Evaluation Under an Unseen Soil Background

To further examine whether the performance improvement was maintained under unseen background conditions and to provide an additional evaluation beyond the original 455-image test set, YOLO11n and HFU-YOLO were evaluated on the independent cross-condition test set. The results are shown in Table 3. On this additional test set, HFU-YOLO achieved Precision, Recall, mAP@0.5, and mAP@0.5:0.95 values of 83.6%, 76.2%, 82.4%, and 46.3%, respectively. Compared with YOLO11n, these values increased by 9.2, 4.6, 3.6, and 0.6 percentage points, respectively. These results indicate that HFU-YOLO maintained better overall detection performance than the baseline under unseen soil-background interference, showing its preliminary robustness in the independent cross-condition evaluation.

Table 3. Detection results on the independent cross-condition test set.

Model	P (%)	R (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
YOLO11n	74.4	71.6	78.8	45.7
HFU-YOLO	83.6	76.2	82.4	46.3

3.4. Ablation Experiments

To evaluate the contributions of HWD, LFADC, and UGS-Lite, ablation experiments were conducted using YOLO11n as the baseline, and the results are shown in Table 4. When HWD, LFADC, and UGS-Lite were introduced separately, mAP@0.5:0.95 increased to 41.9%, 40.2%, and 40.7%, respectively. Among them, HWD produced a larger improvement in mAP@0.5:0.95 when used alone, mainly because wavelet-based downsampling helped retain a shallow edge and texture information. However, the effects of the three modules were not completely independent, and their combined contribution should be interpreted together with the results of different module combinations.

From the combined results, LFADC + UGS-Lite achieved mAP@0.5 and mAP@0.5:0.95 values of 82.4% and 43.9%, respectively, which were close to those of the complete HFU-YOLO model. After HWD was further introduced, mAP@0.5 increased to 83.5%, while mAP@0.5:0.95 increased slightly to 44.1%. This indicates that the additional improvement brought by HWD in the complete model was relatively limited. Nevertheless, HWD was

retained because it contributed to lightweight downsampling and helped the complete model maintain slightly lower GFLOPs than YOLO11n, while also preserving shallow edge and texture information during downsampling. Therefore, HWD mainly plays a complementary role in lightweight downsampling and low-level feature preservation, while LFADC and UGS-Lite contribute more directly to feature enhancement and localization stability. Overall, HFU-YOLO achieved P, R, mAP@0.5, and mAP@0.5:0.95 values of 77.6%, 78.3%, 83.5%, and 44.1%, respectively, representing improvements of 4.2, 3.2, 6.1, and 6.1 percentage points over YOLO11n.

Table 4. Ablation experiment results.

YOLO11n	HWD	LFADC	UGS-Lite	P (%)	R (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Param (M)	GFLOPs
✓				73.4	75.1	77.4	38.0	2.58	6.3
✓	✓			81.5	68.8	79.9	41.9	2.56	5.9
✓		✓		76.3	73.7	79.2	40.2	2.61	6.6
✓			✓	81.0	73.5	79.8	40.7	2.58	6.3
✓	✓	✓		73.0	79.2	80.9	41.0	2.59	6.2
✓	✓		✓	80.9	73.8	81.2	41.1	2.56	5.9
✓		✓	✓	76.3	79.9	82.4	43.9	2.61	6.6
✓	✓	✓	✓	77.6	78.3	83.5	44.1	2.59	6.2

To further examine the influence of training randomness, repeated experiments were conducted for YOLO11n and HFU-YOLO using three fixed random seeds, namely 0, 42, and 3407. The results are shown in Table 5.

Table 5. Repeated training results under different random seeds.

Model	Seed	P (%)	R (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
YOLO11n	42	73.4	75.1	77.4	38.0
YOLO11n	0	74.3	73.8	77.8	39.6
YOLO11n	3407	75.9	72.2	77.9	39.8
YOLO11n	Mean ± SD	74.5 ± 1.3	73.7 ± 1.5	77.7 ± 0.3	39.1 ± 1.0
HFU-YOLO	42	77.6	78.3	83.5	44.1
HFU-YOLO	0	79.6	73.5	81.2	41.3
HFU-YOLO	3407	79.8	75.7	81.3	41.5
HFU-YOLO	Mean ± SD	79.0 ± 1.2	75.8 ± 2.4	82.0 ± 1.3	42.3 ± 1.6

The repeated experiments showed that HFU-YOLO achieved higher mean mAP@0.5 and mAP@0.5:0.95 values than YOLO11n, with improvements of 4.3 and 3.2 percentage points, respectively. Although performance fluctuations were observed across different random seeds, the overall improvement of HFU-YOLO over the baseline was maintained.

3.5. Weight Sensitivity Analysis of UGS-Lite

To analyze the effect of the UGS-Lite auxiliary loss weight on detection performance, sensitivity experiments were conducted by setting λ to 0, 0.05, 0.10, 0.15, and 0.20 while keeping the HWD and LFADC structures unchanged. The results are shown in Table 6.

Table 6. Effect of the UGS-Lite auxiliary loss weight λ on detection performance.

λ	P (%)	R (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
0	73.0	79.2	80.9	41.0
0.05	78.4	78.3	83.7	42.8
0.10	77.6	78.3	83.5	44.1
0.15	80.9	75.1	81.9	41.4
0.20	81.7	73.4	82.1	43.2

When $\lambda = 0$, mAP@0.5 and mAP@0.5:0.95 were 80.9% and 41.0%, respectively. When $\lambda = 0.05$, mAP@0.5 increased to 83.7%. When $\lambda = 0.10$, mAP@0.5:0.95 reached the highest value of 44.1%. As λ was further increased, Precision improved, but Recall decreased, indicating that an excessively strong entropy constraint may make the localization distribution too concentrated and increase the risk of missed detections for ambiguous or partially occluded pests. No convergence failure was observed within the tested weight range. Considering the balance between localization accuracy, Recall, and the need to reduce missed detections in practical pest monitoring, λ was finally set to 0.10 in this study.

3.6. Detection Performance Analysis for Different Classes

To analyze the detection capability of HFU-YOLO for different classes, the detection performance of YOLO11n and HFU-YOLO on the grub, cutworm larva, and cutworm pupa classes was evaluated, and the results are shown in Table 7. Compared with YOLO11n, HFU-YOLO achieved higher AP@0.5, AP@0.75, and AP@0.5:0.95 values for all three target classes. At AP@0.75, HFU-YOLO improved by 10.3, 9.6, and 5.4 percentage points for grub, cutworm larva, and cutworm pupa, respectively, indicating improved localization performance under a stricter IoU requirement. Among the three classes, cutworm larva showed lower Recall and AP@0.5 values, suggesting that missed detections remained more likely for this class under complex soil backgrounds.

Table 7. Comparison of detection performance between YOLO11n and HFU-YOLO for different underground pest classes.

Model	Class	P (%)	R (%)	AP@0.5 (%)	AP@0.75 (%)	AP@0.5:0.95 (%)
YOLO11n	grub	70.5	74.9	72.2	22.5	32.1
YOLO11n	cutworm larva	78.6	62.7	72.6	30.6	37.9
YOLO11n	cutworm pupa	71.0	87.7	87.5	42.8	44.0
HFU-YOLO	grub	77.9	76.6	80.7	32.8	41.1
HFU-YOLO	cutworm larva	78.1	66.6	77.3	40.2	41.3
HFU-YOLO	cutworm pupa	76.9	91.8	92.4	48.2	49.7

3.7. Feature Response Visualization Analysis

To analyze the feature representation capability of HFU-YOLO in complex soil backgrounds, representative samples of the grub, cutworm larva, and cutworm pupa classes were selected, and the response distributions of YOLO11n and HFU-YOLO at the P4 and P5 feature layers were visualized, as shown in Figure 8. The Sample row shows the original images and locally enlarged target regions. Rectangular boxes indicate the target locations, and elliptical boxes are used to assist observation of the corresponding heatmap regions.

Under background interference from soil particles, roots, debris, and other factors, part of the response of YOLO11n was distributed in non-target regions. In contrast, the high-response regions of HFU-YOLO at the P4 and P5 feature layers were more concentrated on the target body and adjacent regions. For the grub and pupa samples, HFU-YOLO showed more compact target responses at the P5 layer. For the cutworm sample, whose color and texture were similar to those of the background, the response regions remained relatively concentrated. These heatmap results suggest that HFU-YOLO enhances pest-related feature responses and reduces invalid background responses, which is consistent with the improvements in localization accuracy, class-wise detection performance, and the false-positive and false-negative analysis.

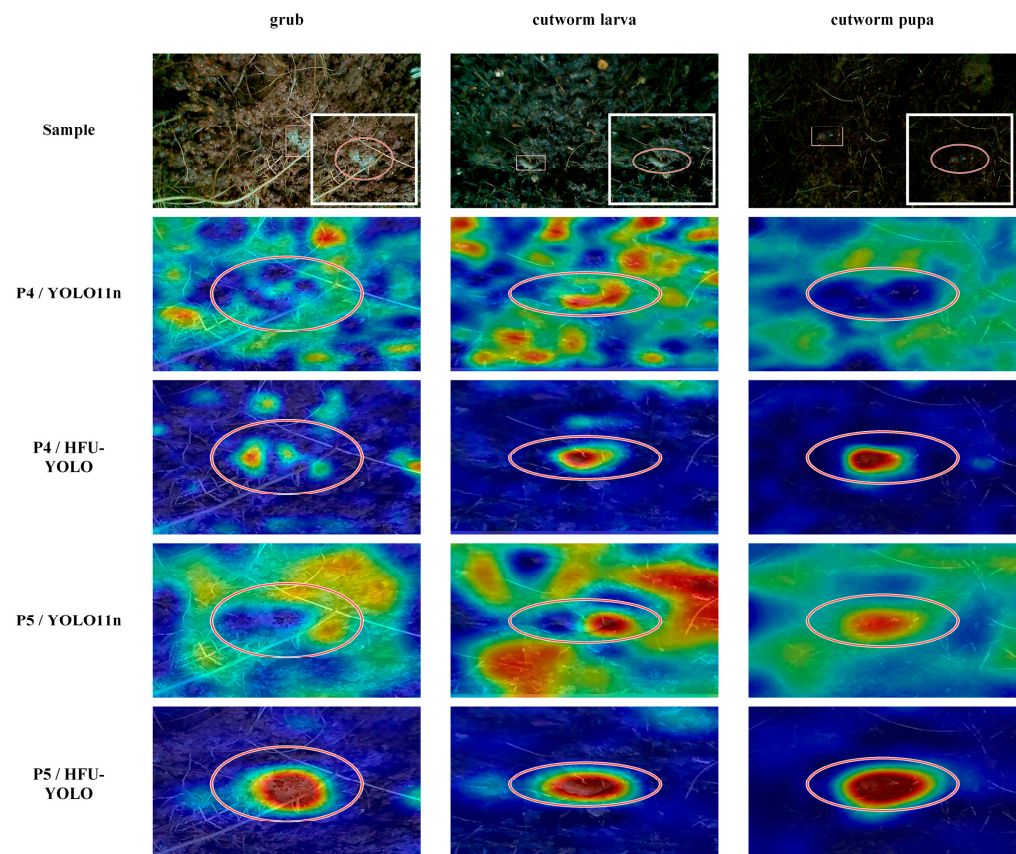


Figure 8. Feature response heatmaps of YOLO11n and HFU-YOLO for different underground pest samples. Note: Warmer colors indicate stronger feature responses.

4. Discussion

4.1. Overall Performance and Task Adaptability

In recent years, studies in this field have mainly improved YOLO detection frameworks from three aspects: complex backgrounds, small-target recognition, and lightweight deployment. Yuan et al. [23] developed I-YOLOv10-SC for small pests and complex backgrounds in tea plantations by using Space-to-Depth convolution, attention mechanisms, and loss function optimization, enabling better small-target localization capability in external validation and complex scenarios. Cen et al. [29] constructed YOLO-LCE based on the Pest24 dataset and improved mAP@0.5 and mAP@0.5:0.95 while reducing the number of parameters and GFLOPs through lightweight convolutional structures, Ghost modules, and the WIoUv3 loss function. Unlike the above studies on aboveground visible pests or public datasets, this study focuses on underground pest images with partial exposure, weak boundaries, and texture confusion in soil media. Rana et al. [39] also emphasized that agricultural detector selection should consider not only mAP and inference speed, but also robustness under background complexity, degraded imaging conditions, and false-classification errors. Therefore, this study further evaluated HFU-YOLO using stricter localization metrics, PR curves, confusion matrices, negative-background false-positive analysis, repeated-seed variability, and cross-condition performance, rather than relying only on mAP@0.5.

The results in Table 2 show that HFU-YOLO achieved mAP@0.5 and mAP@0.5:0.95 values of 83.5% and 44.1%, respectively, both 6.1 percentage points higher than those of the baseline YOLO11n. Compared with YOLOv5n, YOLOv8n, YOLOv10n, and YOLO11s, HFU-YOLO achieved higher detection accuracy while maintaining lower numbers of parameters and GFLOPs. Zhang et al. [35] reported that LTF-YOLO achieved higher mAP

on IP102 and MCP15 despite having significantly fewer parameters and GFLOPs than YOLO11s, and indicated that its performance gains mainly came from effective structural design rather than simple model scaling. These findings suggest that task-specific feature preservation, background suppression, and localization optimization are important for achieving a favorable accuracy–complexity balance. Nevertheless, the current results should be interpreted within the scope of the constructed dataset, which was mainly collected under indoor soil-box conditions and included three underground pest categories.

The additional cross-condition evaluation further supports the robustness analysis of HFU-YOLO. Although this test set was not collected under natural farmland conditions, it introduced an unseen soil-background condition with peanut-shell residues that was not included in the original dataset. The higher overall Precision, Recall, and mAP@0.5 of HFU-YOLO on this test set suggest that the proposed feature-preservation and background-suppression design may improve robustness under certain unseen background disturbances. Nevertheless, cross-condition images may contain non-target fragments with pest-like texture or shape characteristics, which can affect fine localization and model stability. For practical pest-monitoring decisions, Recall is particularly important because missed detections may delay control actions and allow pest damage to continue. The Recall of 76.2% on the cross-condition test set may support preliminary screening under controlled image-acquisition settings, but higher Recall may be required for independent field early-warning and pest-control decisions. The acceptable Recall level should be determined according to pest density, crop growth stage, economic control threshold, control cost, and tolerance for false alarms.

DGS-Tiny achieved the highest FPS, although it had more parameters and GFLOPs than HFU-YOLO. This difference indicates that Params and GFLOPs reflect theoretical model complexity, whereas the measured FPS is also affected by operator implementation, memory access, parallel execution efficiency, and postprocessing cost. All models in this study were evaluated using the same test set, input size, batch size, precision mode, hardware, and software environment, and the reported Time included preprocessing, model inference, non-maximum suppression, and postprocessing. Therefore, the FPS comparison was conducted under the same evaluation protocol. However, DGS-Tiny showed a lower mAP@0.5:0.95, suggesting that higher FPS did not necessarily correspond to better localization performance for underground pest targets with weak boundaries and complex soil backgrounds.

4.2. Module Contributions and Mechanisms

The ablation experiment results show that HWD, LFADC, and UGS-Lite contribute to model performance improvement in different ways. Introducing the HWD module alone increased mAP@0.5:0.95 by 3.9 percentage points, indicating that detail preservation during shallow downsampling is critical for underground pest detection. Underground pest targets are usually small, and their edges and body-segment textures are easily weakened during continuous downsampling. HWD preserves low-frequency structures and high-frequency details through wavelet decomposition, which helps reduce the loss of shallow edge and local texture information, thereby improving the feature representation of small targets and weak-boundary targets.

The contribution of LFADC is mainly reflected in reducing interference from complex backgrounds. Soil particles, cracks, residual roots, and shadows can generate substantial local noise similar to pest textures, making it difficult for conventional convolution to distinguish target textures from redundant background textures. LFADC adaptively fuses contextual information at different scales through frequency selection and multi-dilation-rate depthwise convolution branches, thereby enabling more stable extraction of

pest-related features. In the ablation results, the combination of LFADC and UGS-Lite achieved mAP@0.5 and mAP@0.5:0.95 values of 82.4% and 43.9%, respectively, approaching the performance of the complete model. This indicates that feature enhancement and localization constraint are key contributors to the performance improvement.

The parameter sensitivity experiment of UGS-Lite further shows that the localization distribution constraint needs to be controlled within an appropriate range. When $\lambda = 0.10$, the model achieved the best overall localization performance. As the weight continued to increase, Recall decreased, mainly because excessive regularization could interfere with the original regression logic of DFL. Therefore, $\lambda = 0.10$ was selected as the final setting in this study, allowing UGS-Lite to act only as a lightweight regularization constraint while avoiding interference with the basic regression learning of the model.

4.3. Detection Differences Among Classes and Visualization Analysis

The detection results for different classes show that the accuracy gains of HFU-YOLO for grubs and pupae were more pronounced than those for cutworm larvae. This difference may be related to the inherent morphology and living habits of the pests. Grubs have ring-shaped segment protrusions and show clear grayscale differences from soil, whereas pupae have hard outer shells and regular shapes. These two classes therefore have high feature distinguishability. In contrast, cutworm larvae have smooth body surfaces without distinctive textures, and their long-term soil-dwelling habits make their body color close to that of moist soil. Soil-cover occlusion further weakens their boundary features, making this class difficult to improve through network optimization alone.

The feature heatmaps further support these class-specific patterns. Compared with YOLO11n, HFU-YOLO showed activation regions at the P4 and P5 feature layers that were more concentrated on the pest body and adjacent boundary regions, while invalid feature responses from the soil background were reduced. This observation supports the feature optimization effects of HWD and LFADC. However, feature dispersion still existed in some cutworm larvae samples, indicating that algorithmic structure optimization has performance limitations in scenarios where pest features are highly similar to the soil background. In future work, difficult samples should be expanded and targeted data augmentation should be used to address this limitation.

4.4. Research Limitations and Future Work

This study still has three objective limitations. First, most experimental data were collected in an indoor soil-box environment, where soil composition and debris types were controllable. Although an independent cross-condition test set was added under an unseen soil-background condition, it was still acquired under controlled experimental conditions rather than natural farmland conditions. Therefore, the current results should be interpreted as evidence of model robustness under controlled and cross-condition image-acquisition settings, rather than as sufficient validation for practical field deployment. This differs from natural farmland conditions with variable residues, stones, roots, illumination, moisture, soil depth, compaction, organic matter, and imaging distances, which may result in a domain shift. Second, the experiment included only three representative underground pest classes, while other common underground pests, such as mole crickets and wireworms, as well as visually similar non-pest soil organisms, were not included in the dataset. Therefore, the current model has only been validated for these three categories, and its broad-spectrum recognition capability remains to be further verified. Third, the current model only performs detection on single static images and has not yet been adapted to continuous real-time video monitoring in field scenarios. In practical field inspection, image acquisition may be implemented by an endoscope inserted into the soil layer or by a push-rod micro-camera

device for local underground imaging. However, practical deployment still requires further consideration of camera movement, variable imaging distance, motion blur, severe soil coverage, and repeated counting of the same pest in continuous video detection, because these factors may reduce boundary clarity, increase background interference, and affect the stability of pest localization and counting.

Future work will narrow the gap between laboratory testing and field application. First, in situ field pest images will be collected from farmland with different soil types, such as sandy soil, clay loam, and loam, under different soil moisture conditions, occlusion levels, illumination conditions, soil depths, soil compaction levels, organic-matter contents, crop-residue interference, and imaging distances, and field-scale validation under natural agricultural conditions will be conducted. This will expand the field dataset and reduce laboratory-to-field domain differences. Second, additional underground pest samples, such as mole crickets and wireworms, as well as visually similar non-pest soil organisms, will be incorporated to extend the applicable target range of the model. Third, video-stream-based algorithm optimization will be carried out to develop a portable monitoring prototype integrating image acquisition, real-time detection, repeated-target suppression, and pest-density statistics. Compared with traditional soil sampling, this image-based workflow is expected to support rapid screening of soil sample images by automatically locating suspected pest targets for manual confirmation. However, the reductions in labor cost, visual searching time, and sampling time were not quantitatively measured in this study and should be evaluated in future field trials.

5. Conclusions

To address the detection challenges of blurred boundaries, partial occlusion, and strong background interference for underground pests in complex soil backgrounds, this study constructed an underground pest dataset containing grubs, cutworm larvae, and cutworm pupae, and proposed HFU-YOLO, a lightweight detection model based on YOLO11n. The model uses HWD to preserve edge and texture details during shallow downsampling, LFADC to suppress soil background interference, and UGS-Lite to constrain the localization distribution of weak-boundary targets. These three components adapt the model to the imaging characteristics of underground pests from the perspectives of shallow detail preservation, background interference suppression, and weak-boundary localization.

The results showed that HFU-YOLO achieved Precision, Recall, mAP@0.5, and mAP@0.5:0.95 values of 77.6%, 78.3%, 83.5%, and 44.1% on the test set, respectively, representing improvements of 4.2, 3.2, 6.1, and 6.1 percentage points over the baseline YOLO11n. While improving detection accuracy, the model maintained low complexity, with 2.59 M parameters and 6.2 GFLOPs, and simultaneously improved target detection and boundary localization accuracy under a lightweight design. These results indicate that structural improvements tailored to underground pest images can be more effective for weak-target detection in complex soil backgrounds than directly using the original lightweight baseline model.

HFU-YOLO improved the detection performance of the three evaluated underground pest categories under controlled soil-box and independent cross-condition image-acquisition settings. These results suggest that the proposed method has potential as a methodological reference for underground pest image recognition, soil sample image analysis, and pest survey assistance. Future research will expand in situ field image datasets by including additional underground pest categories and more diverse imaging conditions, such as different soil types, moisture conditions, illumination conditions, soil depths, crop-residue interference, and imaging distances, and conduct field-scale cross-scenario

validation to further evaluate the generalization ability and stability of the model before deployment in practical underground pest-monitoring scenarios.

Author Contributions: Conceptualization, Z.W. (Zhen Wang), Z.M. and S.L.; methodology, Z.W. (Zhen Wang), Z.M. and S.L.; software, Z.W. (Zhen Wang) and J.S.; validation, Z.W. (Zhen Wang), Z.W. (Zhichong Wang) and M.X.; investigation, Z.W. (Zhen Wang), Z.W. (Zhichong Wang) and M.X.; resources, Z.M., Z.W. (Zhichong Wang), C.Z. and S.L.; data curation, Z.W. (Zhen Wang), Z.W. (Zhichong Wang) and M.X.; writing—original draft preparation, Z.W. (Zhen Wang); writing—review and editing, Z.W. (Zhen Wang), M.X., J.S. and S.L.; visualization, Z.W. (Zhen Wang) and Z.W. (Zhichong Wang); supervision, Z.M. and S.L.; project administration, Z.M., C.Z. and S.L.; funding acquisition, Z.M., C.Z. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: (1) The National Key Research and Development Program of China (Funding number: 2024YFD2001100); (2) the Natural Science Foundation of China General Program (Funding number: 32572886).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the corresponding author on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Danu, N.; Paschapur, A.U.; Stanley, J.; Singh, A.K.; Bisht, I.; Gupta, J.P.; Mishra, K.K. Distribution and diversity of white Grub beetles across agro-ecological zones in Uttarakhand, India. *Sci. Rep.* **2025**, *15*, 33777. [[CrossRef](#)] [[PubMed](#)]
2. Slone, E.; Green, J.; Kaur, N.; Walenta, D.L.; Anderson, N.P.; Cruse, C.; Dorman, S.J. Forecasting the seasonal phenology of *Agrotis ipsilon* in Oregon grass seed and vegetable agroecosystems. *Front. Insect Sci.* **2024**, *4*, 1505524. [[CrossRef](#)] [[PubMed](#)]
3. Furlan, L.; Bourlot, G.; Turchi, A.; Snichelotto, V.; Fusaro, M.; Bona, S. The Association Between Soil Sampling and Bait Traps in Wireworm Monitoring: A Methodological and Statistical Approach. *Insects* **2026**, *17*, 419. [[CrossRef](#)] [[PubMed](#)]
4. Preti, M.; Verheggen, F.; Angeli, S. Insect pest monitoring with camera-equipped traps: Strengths and limitations. *J. Pest Sci.* **2020**, *94*, 203–217. [[CrossRef](#)]
5. Li, W.; Zheng, T.; Yang, Z.; Li, M.; Sun, C.; Yang, X. Classification and detection of insects from field images using deep learning for smart pest management: A systematic review. *Ecol. Inform.* **2021**, *66*, 101460. [[CrossRef](#)]
6. Teixeira, A.; Ribeiro, J.; Morais, R.; Sousa, J.; Cunha, A. A Systematic Review on Automatic Insect Detection Using Deep Learning. *Agriculture* **2023**, *13*, 713. [[CrossRef](#)]
7. Mu, J.; Ma, B.; Wang, Y.; Ren, Z.; Liu, S.; Wang, J. Review of Crop Disease and Pest Detection Algorithms Based on Deep Learning. *Trans. Chin. Soc. Agric. Mach.* **2023**, *54*, 301–313.
8. Luo, J.; Zhu, H.; Tang, R.; He, S. Intelligent dual-modal hyperspectral imaging system with deep learning for in-field soil fauna identification. *Comput. Electron. Agric.* **2025**, *239*, 111040. [[CrossRef](#)]
9. Wu, X.; Zhan, C.; Lai, Y.-K.; Cheng, M.-M.; Yang, J. IP102: A Large-Scale Benchmark Dataset for Insect Pest Recognition. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 8779–8788. [[CrossRef](#)]
10. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [[CrossRef](#)]
11. 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. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475. [[CrossRef](#)]
12. Wang, C.-Y.; Yeh, I.-H.; Mark Liao, H.-Y. Yolov9: Learning what you want to learn using programmable gradient information. In Proceedings of the European Conference on Computer Vision, Milan, Italy, 29 September 2024; pp. 1–21. [[CrossRef](#)]
13. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. Yolov10: Real-time end-to-end object detection. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 107984–108011. [[CrossRef](#)]
14. Hao, S.; Gao, E.; Ji, Z.; Ganchev, I. BCS_YOLO: Research on Corn Leaf Disease and Pest Detection Based on YOLOv11n. *Appl. Sci.* **2025**, *15*, 8231. [[CrossRef](#)]
15. Wen, C.; Chen, H.; Ma, Z.; Zhang, T.; Yang, C.; Su, H.; Chen, H. Pest-YOLO: A model for large-scale multi-class dense and tiny pest detection and counting. *Front. Plant Sci.* **2022**, *13*, 973985. [[CrossRef](#)] [[PubMed](#)]

16. Wang, Z.; Zhang, H.-W.; Dai, Y.-Q.; Cui, K.; Wang, H.; Chee, P.W.; Wang, R.-F. Resource-Efficient Cotton Network: A Lightweight Deep Learning Framework for Cotton Disease and Pest Classification. *Plants* **2025**, *14*, 2082. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Sun, H.; Wang, R.-F.; Yu, R.; Sun, Y. Pest image recognition algorithm based on joint adversarial transfer learning. *Measurement* **2026**, *278*, 121620. [\[CrossRef\]](#)
18. Wang, R.-F.; Cui, K.; Schar dong, I.B.; Bauer, M.C.; Somala, R.V.; Xu, M.; West, D.; Jones, D.C.; Taylor, S.V.; Roberts, P.M.; et al. JassidNet as a quantization-aware lightweight phenotyping framework for high-throughput cotton jassid (*Amrasca biguttula*) detection and counting toward objective resistance screening. *Ind. Crops Prod.* **2026**, *249*, 123716. [\[CrossRef\]](#)
19. Wang, R.-F.; Zhao, C.-T.; Tu, Y.-H.; Chen, Z.-Q.; Su, W.-H. A Field-Adaptive Mechanical Weeding System Coupling Oscillating Pneumatic Mechanism With Deep Learning for Intra-Row Weed Control in Lettuce. *J. Field Robot.* **2026**. [\[CrossRef\]](#)
20. Zheng, Y.; Zheng, W.; Du, X. A lightweight rice pest detection algorithm based on improved YOLOv8. *Sci. Rep.* **2024**, *14*, 29888. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Wu, X.; Liang, J.; Yang, Y.; Li, Z.; Jia, X.; Pu, H.; Zhu, P. SAW-YOLO: A Multi-Scale YOLO for Small Target Citrus Pests Detection. *Agronomy* **2024**, *14*, 1571. [\[CrossRef\]](#)
22. Zhang, X.a.; Qiao, H.; Xi, L.; Ma, C.; Xu, X.; Hui, X.; Tang, Q.; Zhang, H. Detecting small objects for field crop pests using improved YOLOv8. *Trans. Chin. Soc. Agric. Eng.* **2025**, *41*, 225–233. [\[CrossRef\]](#)
23. Yuan, W.; Lan, L.; Xu, J.; Sun, T.; Wang, X.; Wang, Q.; Hu, J.; Wang, B. Smart Agricultural Pest Detection Using I-YOLOv10-SC: An Improved Object Detection Framework. *Agronomy* **2025**, *15*, 221. [\[CrossRef\]](#)
24. Dai, Q.; Liang, S.; Li, Z.; Lyu, S.; Xue, X.; Song, S.; Huang, Y.; Zhang, S.; Fu, J. YOLOv11-RDTNet: A Lightweight Model for Citrus Pest and Disease Identification Based on an Improved YOLOv11n. *Agronomy* **2025**, *15*, 1252. [\[CrossRef\]](#)
25. Huang, Y.; Liu, Z.; Zhao, H.; Tang, C.; Liu, B.; Li, Z.; Wan, F.; Qian, W.; Qiao, X. YOLO-YSTs: An Improved YOLOv10n-Based Method for Real-Time Field Pest Detection. *Agronomy* **2025**, *15*, 575. [\[CrossRef\]](#)
26. Lu, Y.; Liu, P.; Tan, C. MA-YOLO: A Pest Target Detection Algorithm with Multi-Scale Fusion and Attention Mechanism. *Agronomy* **2025**, *15*, 1549. [\[CrossRef\]](#)
27. Wang, X.; Zhang, S.; Zhang, T. Crop insect pest detection based on dilated multi-scale attention U-Net. *Plant Methods* **2024**, *20*, 34. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Liu, J.; Yu, E.; Li, Y.; Zhao, Y.; Mao, B. YOLO-DCPG: A lightweight architecture with dual-channel pooling gated attention for intensive small-target agricultural pest detection. *Front. Plant Sci.* **2025**, *16*, 1716703. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Cen, X.; Lu, S.; Qian, T. YOLO-LCE: A Lightweight YOLOv8 Model for Agricultural Pest Detection. *Agronomy* **2025**, *15*, 2022. [\[CrossRef\]](#)
30. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 12993–13000. [\[CrossRef\]](#)
31. Xu, C.; Wang, J.; Yang, W.; Yu, H.; Yu, L.; Xia, G.-S. Detecting tiny objects in aerial images: A normalized Wasserstein distance and a new benchmark. *ISPRS J. Photogramm. Remote Sens.* **2022**, *190*, 79–93. [\[CrossRef\]](#)
32. Yu, P.; Zong, B.; Geng, X.; Yan, H.; Liu, B.; Chen, C.; Liu, H.; Xu, X. DGS-Yolov7-Tiny: A lightweight pest and disease target detection model suitable for edge computing environments. *Sci. Rep.* **2025**, *15*, 29818. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Dong, Q.; Sun, L.; Han, T.; Cai, M.; Gao, C. PestLite: A novel YOLO-based deep learning technique for crop pest detection. *Agriculture* **2024**, *14*, 228. [\[CrossRef\]](#)
34. Song, J.; Cheng, K.; Chen, F.; Hua, X. RDW-YOLO: A deep learning framework for scalable agricultural pest monitoring and control. *Insects* **2025**, *16*, 545. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Zhang, C.; Shan, S.; Wang, X.; Shen, Y. LTF-YOLO: A lightweight framework for long-tailed agricultural pest detection in complex field environments. *Comput. Electron. Agric.* **2026**, *250*, 111901. [\[CrossRef\]](#)
36. Xu, G.; Liao, W.; Zhang, X.; Li, C.; He, X.; Wu, X. Haar wavelet downsampling: A simple but effective downsampling module for semantic segmentation. *Pattern Recognit.* **2023**, *143*, 109819. [\[CrossRef\]](#)
37. Chen, L.; Gu, L.; Zheng, D.; Fu, Y. Frequency-adaptive dilated convolution for semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 3414–3425. [\[CrossRef\]](#)
38. Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21002–21012. [\[CrossRef\]](#)
39. Rana, S.; Hensel, O.; Nasirahmadi, A. From Vineyard to Vision: Multi-Domain Analysis and Mitigation of Grape Cluster Detection Failures in Complex Viticultural Environments. *Results Eng.* **2025**, *29*, 108833. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.