



ARTICLE

## YOLOCSP-PEST for Crops Pest Localization and Classification

Farooq Ali<sup>1,\*</sup>, Huma Qayyum<sup>1</sup>, Kashif Saleem<sup>2</sup>, Iftikhar Ahmad<sup>3</sup> and Muhammad Javed Iqbal<sup>4</sup>

<sup>1</sup>Department of Software Engineering, University of Engineering and Technology, Taxila, 47050, Pakistan

<sup>2</sup>Department of Computer Science & Engineering, College of Applied Studies & Community Service, King Saud University, Riyadh, 11362, Saudi Arabia

<sup>3</sup>Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

<sup>4</sup>Department of Computer Science, University of Engineering and Technology, Taxila, 47050, Pakistan

\*Corresponding Author: Farooq Ali. Email: farooq.ali@uettaxila.edu.pk

Received: 08 November 2024; Accepted: 16 January 2025; Published: 17 February 2025

**ABSTRACT:** Preservation of the crops depends on early and accurate detection of pests on crops as they cause several diseases decreasing crop production and quality. Several deep-learning techniques have been applied to overcome the issue of pest detection on crops. We have developed the YOLOCSP-PEST model for Pest localization and classification. With the Cross Stage Partial Network (CSPNET) backbone, the proposed model is a modified version of You Only Look Once Version 7 (YOLOv7) that is intended primarily for pest localization and classification. Our proposed model gives exceptionally good results under conditions that are very challenging for any other comparable models especially conditions where we have issues with the luminance and the orientation of the images. It helps farmers working out on their crops in distant areas to determine any infestation quickly and accurately on their crops which helps in the quality and quantity of the production yield. The model has been trained and tested on 2 datasets namely the IP102 data set and a local crop data set on both of which it has shown exceptional results. It gave us a mean average precision (mAP) of 88.40% along with a precision of 85.55% and a recall of 84.25% on the IP102 dataset meanwhile giving a mAP of 97.18% on the local data set along with a recall of 94.88% and a precision of 97.50%. These findings demonstrate that the proposed model is very effective in detecting real-life scenarios and can help in the production of crops improving the yield quality and quantity at the same time.

**KEYWORDS:** Deep learning; classification of pests; YOLOCSP-PEST; pest detection

### 1 Introduction

Infestation by different kinds of pests has been an ongoing issue among the crops for a very long time and in the modern world it has become a huge issue leading, often, to food shortage in many areas. For individual farmers, it can be a very big problem and, is awful and very difficult to accurately detect pests from other friendly insights on their crops. Computer vision combined with deep learning has a deep founding impact on this situation as it helps farmers to detect all the kinds of pests in the crops accurately using localization and classification of predetermined classes of pests which helps in the quality and production of crops [1]. Our research summarizes technical details and methods across several processing phases to provide a comprehensive understanding of deep learning methodologies, which should aid future advancements in the field of smart pest monitoring. Recent developments in deep learning have revolutionized the detection of plant diseases and pests using digital image processing, exceeding traditional methods. The study focuses



on three main components: segmentation networks, detection networks, and classification networks [2], which compare deep learning to more conventional methods in this area of application. Although deep learning has made significant progress, problems still arise from a wide range of pests' environments and appearances, so there is a need for the development of more advanced methods for accurate pest detection [3]. Agricultural production and yield protection are significantly impacted by the new methodology's increased efficacy in pest identification. Smart farming has led to the development of new techniques that use the Convolutional Neural Network (CNN) to identify agricultural pests [4]. The limitations of identifying pests are outlined, encompassing the variety of pest species, the complexities of the environment, and the several phases of a pest's life cycle. Pest detection technologies have evolved significantly, leveraging advanced algorithms to address the complexities of agricultural pest management. CNN excels in feature extraction and spatial pattern recognition, enabling high detection accuracy but struggling with small or occluded pests and dataset quality dependencies. Long Short-Term Memory (LSTM) networks effectively analyze temporal patterns in pest behavior, though they face challenges with data scarcity and computational demands. Simpler models like Naive Bayes and K-Nearest Neighbors (KNN) offer efficiency and ease of implementation but are hindered by assumptions of feature independence, noise susceptibility, and scalability issues. Ensemble techniques like Random Forest provide robust feature handling and improved accuracy but may suffer from overfitting and reduced interpretability. Support Vector Machine (SVM) and Decision Trees offer high accuracy and interpretability, respectively, but face challenges with computational efficiency, large datasets, and data imbalances. These limitations highlight the need for hybrid approaches and enhanced data preprocessing to optimize pest detection systems for real-world applications. The introduction lists the difficulties in recognizing pests from images taken in the actual world, including their small size, complex settings, and variations in mass and position. It also highlights the difficulties posed by physical traits shared by several insect species and adverse environmental conditions that affect automated detection. The study proposes a deep learning model optimized for pest detection, YOLO-CSP-PEST, as a solution to these problems. This method improves the process of calculating image characteristics for accurate one-step localization and classification. Highlighted include the method's precision in recognizing and classifying multiclass pests, on challenging datasets like IP102, and its adaptability demonstrated on a local crop dataset, suggesting its performance in other agricultural applications. Our main goal is to improve crop monitoring and mitigation strategies. It also presents a robust approach for agricultural pest identification and shows how widely it may be applied in pest management systems. The following are the major contributions of our work:

- We have introduced a comprehensive analysis using the challenging IP102 dataset. By validating the efficiency of the proposed YOLO-CSP-PEST model in accurate localization and classification of pests in various conditions including noise, blurring, color, and light variations.
- Using a locally collected crop dataset, the YOLO-CSP-PEST performance demonstrates its adaptability to a range of environmental circumstances. This demonstrates that it can be applied in real-world agricultural scenarios.
- Our YOLO-CSP-PEST model simplifies the localization and classification of multiclass pests. It effectively manages a wide range of insect species throughout the life stages.
- The YOLO-CSP-PEST model solves the limitations of existing architectures by using a CSPNET backbone. This two-step locator significantly increases the accuracy of pest detection in complex agricultural conditions by using CSPNET for feature extraction.

The paper is divided into several sections. The existing works are explained in [Section 2](#), the proposed work is in [Section 3](#). The results are demonstrated in [Section 4](#) and the conclusion is demonstrated in [Section 5](#).

## 2 Related Work

Pest damage has posed serious problems to the agriculture industry in recent years, resulting in large financial losses. Numerous strategies, such as deep learning-based techniques like Pest-YOLO, have been put forth to address these issues [5]. These methods aim to improve accuracy and real-time detection capabilities through advancements such as efficient channel attention mechanisms and transformer encoders. The study reflects a trend towards enhancing feature extraction and incorporating global contextual information to achieve better performance in pest detection tasks.

The paper [6] highlights the pervasive problem of pest infestation causing substantial crop damage and economic losses worldwide. Automatic identification of invasive insects is crucial for swift recognition and removal of pests. The paper proposes ensembles of CNNs using diverse topologies and Adam optimization variants to enhance pest identification. The approach outperforms human expert classifications on insect datasets and achieves state-of-the-art results. CNNs, particularly employing Adam optimization, have become the dominant choice for pest classification. Different Adam variants and their effectiveness are discussed, emphasizing the significance of Adam in minimizing training loss. Future directions include active learning and further evaluation of Adam variants on additional CNN architectures like NasNet. Insects, comprising nearly 80% of Earth's animal species, often pose agricultural pest problems, leading to crop damage and failures [7]. Identifying these pests accurately remains a challenge. Leveraging CNN for pest image recognition offers a solution. This study focuses on optimizing small-sized MobileNetV2 models with dynamic learning rates. The proposed model outperforms larger baseline models, attaining an accuracy of 0.7132. Applying this efficient model on mobile devices like smartphones or drones aids in precise pest identification and controlled insecticide usage, enhancing agricultural practices. The study [8] introduces a cool way to deal with pesky insects that harm crops without going overboard on pesticides. They used fancy technology called deep Convolutional Neural Networks to spot 102 common crop pests better, even with a not-so-huge database of pests. They taught a MobileNet model to do this after lots of tweaking and testing different versions to find the best one. Then, they made a smartphone app using Flutter so that anyone can snap a pic of a bug or choose one from their photo gallery to identify it, whether they're online or offline. The significant issue of pests using excessive amounts of pesticides to damage crops and the environment was investigated in the study [9]. The authors used a comparison with existing approaches to emphasize the special qualities and effectiveness of their approach. They used a method that involved optimizing the YOLOv5 model's parameters to detect pests with remarkable recall and precision. They created a state-of-the-art quantum machine learning model and used it to accurately classify these pests.

In 2021, Liu et al. [10] conducted research on systematically addressing the challenges of plant ailments. They compiled a comprehensive crop illness dataset comprising 271 crop disease classes and 220,592 images to facilitate plant disease identification studies. In 2020, Chen et al. [11] also proposed integrating image recognition technologies with artificial intelligence into the landscape and the Internet of Things (IoT) for pest identification. They utilized YOLOv3 for picture segmentation to locate *Tessaratoma papillosa* and employed LSTM for analyzing environmental data from weather sensors to predict pest presence. The study [12] focuses on the growing issue of pests in agriculture and the need for better ways to detect and handle them. Traditional methods have their limitations, prompting a shift toward deep learning techniques like CNNs. This approach aims to strike a compromise between accuracy and efficiency by using pre-trained models and enhanced training approaches. This work provides a useful tool for detecting agricultural pests by proving its effectiveness on publicly available information. Still, other research indicates that more recent vision models, such as ViT, may have unrealized potential and provide even better outcomes. Making these intricate models workable for actual applications is the difficult part. LSTM networks offer strong potential for pest detection due to their ability to handle sequential data and capture longterm dependencies,

allowing for effective analysis of temporal patterns in pest behavior. However, LSTM models may struggle with limited data availability and require substantial computational resources for training and inference, posing challenges in real-time application and scalability for large-scale pest detection systems. In 2022, Hassan et al. [13] introduced a deep learning approach leveraging feedback connections and an innovative initial layer structure. They employed depth-wise separable convolution to effectively reduce the number of parameters. The model was trained and evaluated on three distinct plant disease datasets, achieving impressive performance accuracies of 99.39% on general plant diseases, 99.66% on rice illness, and 76.59% on cassava diseases. This approach, inspired by the Inception design, significantly decreased processing costs by a factor of 70 through variable reduction. In 2022, Wu et al. [14] also proposed a CNN framework for the Fine-Grained Categorization of Plant Leaf Diseases. However, the intricate network architecture posed challenges in terms of memory storage and processing power consumption, hindering its practical implementation in low-cost devices. To address this issue, their research focused on fine-grained disease classification based on a network, incorporating a “Reconstruction-Generation Model” during training. Furthermore, a refined disease identification strategy utilizing visual deep neural networks was developed for detecting peach and tomato disease leaflets, aiming to mitigate the limitations of deep neural networks in crop disease identification.

A primary focus is on the accurate identification of species, which is necessary to better control agricultural methods and avoid crop failures caused by pests and diseases [15]. The paper highlights the drawbacks of traditional identification approaches as well as the trend towards the application of deep learning methods. The article shares encouraging outcomes from experiments conducted on open datasets. Moreover, Liu et al. [16] introduced innovative network structures, such as the Global Activated Feature Pyramid Network (GaFPN) and Locally Activated Region Proposal Network (LaRPN), obtaining an 86.9% accuracy in pest recognition. Other studies by Liu et al. [17], Dawei et al. [18], Xia et al. [19], and Li et al. [20] demonstrated notable accuracies ranging from 75.46% to 93.84% in pest classification, utilizing different CNN models and transfer learning techniques. Furthermore, researchers like Sabanci et al. [21], Roy et al. [22], and Liu et al. [23] integrated CNN-based models with various architectures for disease and pest recognition, achieving detection rates exceeding 96%. D2Det’s model localizes pest regions with a detection score of 0.78 mAP, while a regional proposal network achieves a mAP of 0.78 on the AgriPest21 dataset [24]. A multi-feature fusion network, incorporating dilated convolution and deep feature extraction, yields 98.2% accuracy in classifying 12 types of pest diseases [25]. Attention models based on activation mapping and saliency discriminative guided models detect tiny pest regions with average accuracies of  $68.3 \pm 0.3$  and employ various branches for salient object detection in pest localization [26]. Fusion models, such as combining spatial channel networks with CNN and position-sensitive score maps, show promise in detecting pest illnesses with a mAP of 0.75 on the MPD-2018 dataset [17]. Compared to traditional pest identification techniques, deep learning, and computer vision provide exceptional accuracy. Deep learning models, especially CNN, autonomously learn hierarchical features from raw data, improving resilience and scalability in contrast to traditional methods that frequently rely on handmade features and presume feature independence. Large datasets and a variety of environmental circumstances are easily handled by these technologies, which makes it possible to discover hidden or subtle pests that conventional algorithms could miss. Deep learning and computer vision are revolutionary tools for contemporary pest detection systems because of their capacity to incorporate temporal analysis through designs like LSTM networks, which further expand their application in comprehending insect behavior.

Timely and precise identification of harmful insects and diseases is crucial, as highlighted by research [27] on pest detection in precision agriculture. Due to the mobility and hiding of pests, traditional approaches are hindered, which calls for the creation of sophisticated early warning systems. To meet

the growing demand for high-quality fiber, YOLO-JD represents substantial progress in the identification of diseases in jute plants [28]. Maize-YOLO is a method for real-time agricultural pest identification, particularly for pests that affect maize [29]. With a mean average accuracy (mAP) of 76.3% and a recall rate of 77.3%, Maize-YOLO, tested on a large-scale pest dataset, surpasses state-of-the-art models, providing accurate and quick pest diagnosis for better crop management.

### 3 Methodology

The YOLOCSP-PEST model is used for the localization and classification of different agricultural pests. We have optimized the backbone of YOLOv7, and specifically trained it on pest samples to identify several agricultural pest categories. The improved YOLO model detection, which consists of two main stages: classification and localization, is endowed with rich properties. The localization stage uses a bounding box regression technique to pinpoint the exact position of pest instances in the image. Subsequently, a multi-class classifier is used in the classification step to determine the precise kind of pest present in each bounding box. The IP102 dataset, which has pictures of pests from 102 classes, and a locally gathered agricultural pest dataset are used to train the suggested YOLOv7 model. Reducing the difference between the ground truth annotations and the anticipated bounding boxes is the goal of the training phase. The model may be taught to identify pests in freshly obtained photographs by feeding it new images and using a threshold to remove false positives from the projected bounding boxes.

YOLOCP-PEST takes the input image and then converts it into several rectangular regions that combine to form one grade and for each of these rectangular images we then do further processing. Then these rectangular sections are processed individually, and the YOLO network tries to find objects of interest within each of these rectangular sections and the predictions are made within these rectangular sections. For each of the items predicted we have their information like the forecasted bounding box coordinates, objectness score, and class probabilities for each possible object. Once the forecasting is done and the proposed objects are found, then these proposals are given to the non-maximum suppression (NMS). It makes sure that our result is accurate and free of any redundancies. NMS takes two bounding boxes that are overlapping more than a preset degree and of these two, only one is retained and that is the one that has more accuracy as compared to the other one, doing this will make sure that we get accurate bounding boxes around the objects with the most probability and accuracy at that area.

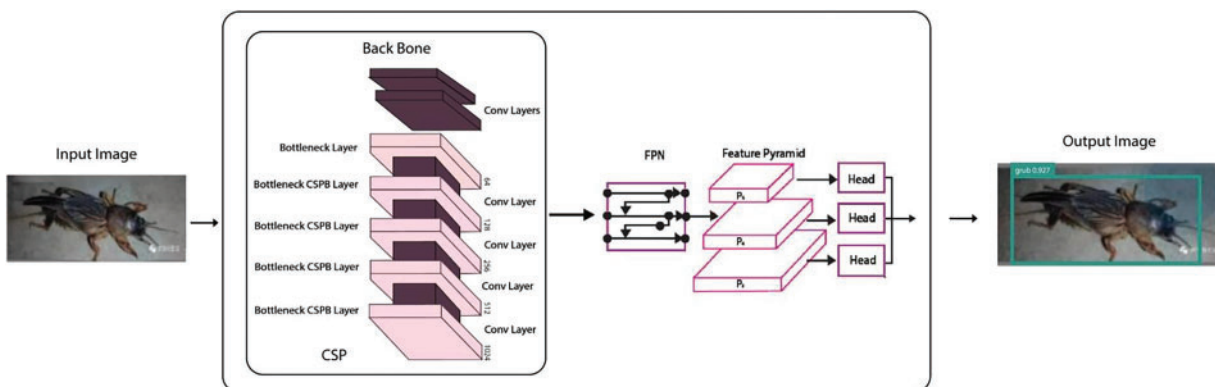
#### 3.1 You Look Only Once Version 7 (YOLOv7)

YOLO family has been in the machine learning outlook for a long time and YOLOv7 is the newest member of this family which has been praised much by the machine learning community and is known for its accuracy and fast results on multiple glasses. The YOLOv7 architecture consists of three main parts namely the backbone neck and head are doing their part to provide the best results. The backbone uses a DarkNet53 architecture to do the basic feature extraction from the image also defining the regions of interest within the input image [30]. The novel idea for YOLOv7 is the path aggregation network that is present in the neck region and the extracted features from the backbone are fed via the neck to the head which results in the final detection of the classes along with their detection score and accuracies. The Path Aggregation Network (PAN) gives scale invariance to the reduction which is very crucial for real life scenarios. Finally comes the head which is also known as the detection head in actuality it consists of two parts one is used for the output while the other one is used solely for the training of the neural network and is also known as the auxiliary head which plays an important part in reducing the error and training the network to give high accuracy output.

### 3.2 YOLOCSP-PEST

We have proposed the model YOLOCSP-PEST for the localization and division of multi-crop pests. We have altered the conventional YOLOv7 approach by utilizing CSPNET as the base network, tuned on the pest samples to recognize the crop pests of various categories, and given the name YOLOCSP-PEST. CSPNET is a good fit for the accurate detection and localization of multiple crop pest classes based on the better feature extraction and region of interest highlighted by CSPNET as the backbone of YOLOv7. The CSPNET backbone in the proposed YOLOCSP-PEST model improves feature learning by splitting feature maps and merging them through cross-stage connections, enhancing gradient flow, and reducing computation. This architecture boosts YOLOv7's efficiency and accuracy, making it ideal for real-time multiclass pest detection in complex agricultural scenarios. The YOLOCSP-PEST architecture consists of three parts after the input is given to it. First, we have the backbone YOLOv7 that serves the purpose of feature extraction from the input image, we use the CSPNET for this of the YOLOCSP-PEST architecture. These features flow into the neck, where bottom-up pathways upsample deeper, semantically rich features, and top-down pathways downsample shallower, detail-rich features. Feature fusion modules cleverly combine these at each level, enriching them with multi-scale information.

Bottleneck layers then refine the combined features for efficiency. Finally, two prediction heads take over: the lead head makes final bounding box and objectness predictions, while the optional auxiliary head assists in training by providing earlier predictions. Non-Max Suppression refines the predictions, and voila, you have object detections with bounding boxes and class labels! This intricate interplay of backbone, neck, and heads, empowered by innovations like CSPNET backbone, adaptive feature fusion, and efficient channel fusion, makes YOLOv7 a powerful object detection machine. Fig. 1 shows the methodology of the YOLOCSP-Pest model. YOLOCSP-PEST achieves faster inference speeds and higher accuracy, particularly in detecting small pests. Compared to traditional methods like ensemble models (e.g., genetic algorithm-based CNNs or Inception modules), YOLOCSP-PEST provides a more resource-efficient, real-time solution with reduced computational overhead. Its ability to balance precision and speed makes it more suitable for large-scale, real-time pest detection.



**Figure 1:** Methodology of YOLOCSP-PEST

### 3.3 Localization and Object Detection

In digital photos and videos, object detection is a computer technique that is associated with computer vision and image processing. Its main function is to find instances of semantic items of a certain class. Combining the ideas of image classification and localization, object identification is a challenging task.

An object detection method would provide bounding boxes around all things of interest in a picture and classify them.

Localization is the term used for the technique in machine learning that defines the exact location of a particular instance of an object we need a video or image. Similarly, in our YOLOCSP-PEST model, the localization is done on the individual portions of the grid that are made up of rectangular components. The localization consists of a bounding box coordinate which gives the exact location within the image of an object, and it also gives the score of that prediction which defines the likeliness of the presence of any one of the classes given during the training. From the predetermined set of categories, our neural network defines as probability score to each of the categories which later is given to the NMS, which selects the one with the highest score. This score represents how likely it is that the item within the bounding box belongs to that typical class we have proposed a powerful neural network that can accurately identify any object within an image and given the binding box using the localization with the accurate guess of the class resulting in good classification which applies to wide real time needs in the industries medical sector and surveillance systems.

### 3.4 Loss Function

The YOLOCSP-PEST loss function is essential for the model to be trained. It leads the model's parameter alterations to minimize the discrepancy between the actual labels for the input data and the model predictions. The following three terms make up the loss function:

- **Object Confidence Loss:** The object's existence in a cell is determined by Object Confidence Loss, which computes the difference between the item's anticipated confidence score. It guarantees that the model labels empty cells with low confidence and predicts the existence of objects with confidence.
- **Localization Loss:** It computes the discrepancy between the ground truth bounding box's actual coordinates and the bounding box's expected coordinates as predicted by the model. It provides advice to the model so that it can accurately predict the positions of observed objects.
- **Classification Loss:** It measures the deviation between the observed class labels of the detected objects and the predicted class probabilities by the model. This ensures that the model appropriately classifies the objects in the image or video.

The contributions of these three terms are added up to determine the YOLOCSP-PEST loss function:

$$\text{Loss} = \text{Object Confidence Loss} + \text{Localization Loss} + \text{Classification Loss} \quad (1)$$

During training, YOLOCSP-PEST learns to precisely locate and classify objects within an image or video by minimizing this loss function. Different components of the detection job, such as localization or classification accuracy, can be prioritized by adjusting the weighting of each word in the loss function.

## 4 Results and Discussion

We explore the details of implementation and assessments carried out to determine the effectiveness of the suggested approach in this section. To present YOLOCSP-PEST's effectiveness completely, we ran multiple experiments to calculate the outcomes of pest recognition and division, contrasting and comparing them to a range of different models. We used the IP102 Dataset for the experimentation of the result with the YOLOCSP-PEST Model. The IP102 Dataset contains more than 75,000 images of 102 distinct kinds of insect pests, grouped to mimic their natural environment. Fig. 2 shows the sample images of the IP102 Dataset. With its 19,000 images, it helps computers find flaws in a shot.



Figure 2: Samples from the IP102 dataset

Because it classifies these pests based on hierarchy, the IP102 dataset is interesting. Bugs that target a single crop, such as wheat or rice, are grouped. So, we have categories like “rice pests,” “wheat pests,” “corn pests,” and “cotton pests”. It covers a whole bunch of pests that might be found all over the place and at different times of the year pests like aphids, beetles, caterpillars, flies, grasshoppers, leafhoppers, moths, and weevils. The following Fig. 3 shows the data distribution of the IP102 dataset.

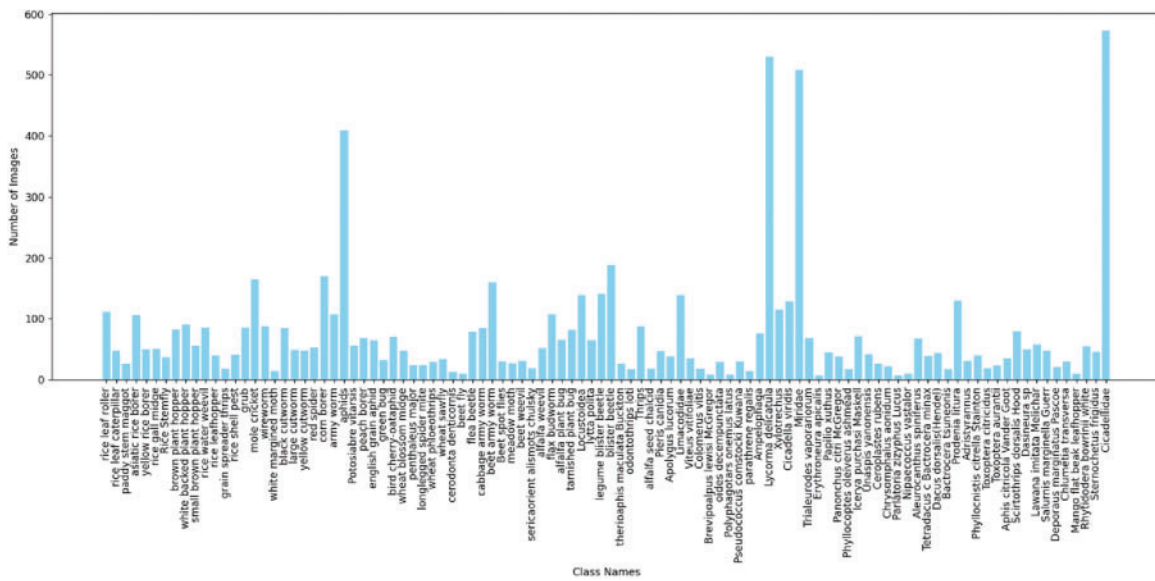


Figure 3: IP102 dataset distribution

### 4.1 Implementation Details

The proposed model is implemented in the TensorFlow environment of Python utilizing the Keras library for the building and training of the model. The hyperparameters used in the training of YOLO-SPPEST have been detailed in Table 1. To get optimum results according to our situations, we fine-tuned some of the hyper parameters namely the batch size, the learning rate, and the number of epochs before the training commenced. We used the stochastic Gradient Descent (SGD) and the Adam training optimizer



model in this implementation. While keeping the learning rate of 0.01, the number of epochs used to be 150 and we set the batch size to be 32 along with that we fixed the input dimensions of the picture being fed as input image at a fixed size of  $320 \times 320$ .

**Table 1:** Training parameters for the proposed model

Framework parameter	Value
Epochs	150
Batch size	32
Image size	320
Learning rate	0.01

#### 4.2 Evaluation Parameters

We've utilized several quantitative indicators: precision (P), recall (R), and mAP to evaluate the effectiveness of the proposed approach. These metrics are calculated as follows:

$$P = TP / (TP + FP) \quad (2)$$

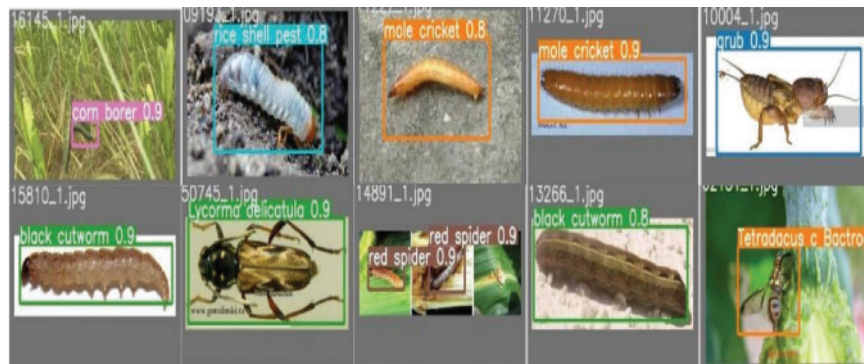
$$R = TP / (TP + FN) \quad (3)$$

Here, TP stands for true positive, TN for true negative, FP for false positive, and FN for false negative instances. In the context of identifying pests in images, a true positive (TP) occurs when the pest is accurately identified. If the pest is not visible in the photograph despite being categorized, it is considered a false negative (FN). On the other hand, if something is incorrectly categorized as a pest when it's not visible in the image, it is a true negative (TN); otherwise, it's a false positive (FP). The computation for mAP is depicted in Eq. (4) below, where AP represents the average accuracy for each category, ' $t$ ' signifies the analyzed image, and ' $T$ ' denotes the total number of test pictures.

$$mAP = \sum_{i=0}^T \frac{AP(t)}{T} \quad (4)$$

#### 4.3 Pest Localization Results

We experimented to assess our framework's ability to accurately locate pests within test samples, crucial for an effective automated pest recognition system. Utilizing the entire test photo set from the IP102 database, as depicted in Fig. 4, our findings reveal the framework's proficiency in identifying pests of diverse sizes, shapes, and colors. Despite challenging conditions like varied lighting, backgrounds, angles, and shifts, our method consistently detects pests effectively. Leveraging key point estimation for localization, our framework accurately distinguishes different pest types. Quantitatively, our mean Average Precision (mAP) score of 0.88 underscores the robustness of our approach in recognizing pests within this complex dataset, highlighting its high recall capability.

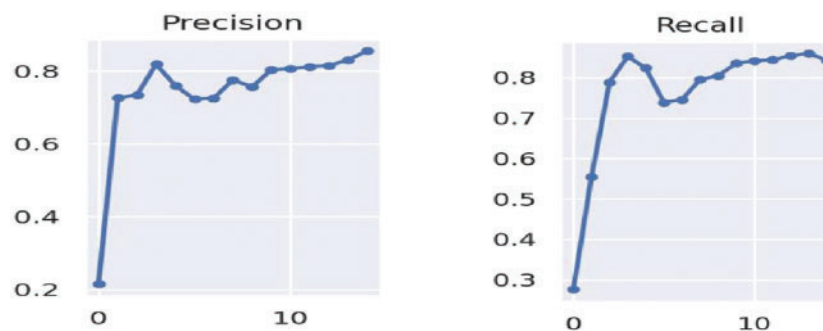


**Figure 4:** Localization visuals attained for the YOLOCSP-PEST

#### 4.4 Pest Classification Results

Accurately categorizing various pests is pivotal to demonstrating the reliability of a model. Given that different crop categories may host diverse pest types, it's essential to evaluate the effectiveness of our proposed procedure in classifying pests across multiple hierarchical crop groups. Thus, we conducted experiments to assess its performance in this regard.

The trained YOLOCSP-PEST framework undergoes evaluation on each test image within the IP102 dataset for this purpose. Fig. 5 showcases the precision, and recall outcomes, specifically focusing on crop-based pest classification. The results highlight that the proposed framework attains precision, and recall of 85.55%, and 84.25% across all crop-specific classes. The robustness of the key point computation approach employed contributes significantly to accurately and reliably delineating each pest class, thereby enhancing the model's performance in pest categorization. Hence, it's evident that YOLOCSP-PEST performs effectively in crop-wise pest identification, validating the efficacy of this approach.



**Figure 5:** Precision and Recall values for the YOLOCSP-PEST (1 unit in the  $x$ -axis is 10 epochs and  $y$ -axis represents precision and recall, respectively)

#### 4.5 Performance Comparison with Existing Approaches

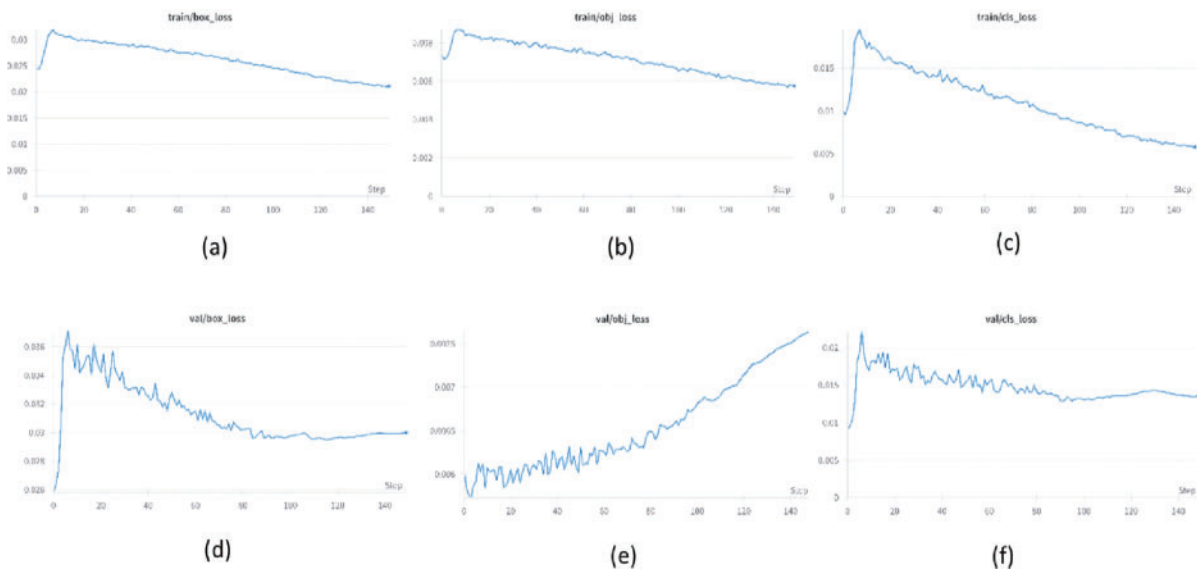
The proposed localization method's results are compared to the latest existing methods, as mentioned in Table 2. This table presents the localization results that are compared to the existing methods. The YOLOPest is an enhanced YOLOv5s algorithm achieving a mAP of 64.6 for accurate detection of small agricultural pests [31]. CornerNet with DenseNet-100 is used for localization on the IP102 dataset, which provides an mAP of 57.23 [32]. The Maize-YOLO achieves a mAP of 76.3, surpassing current YOLO-based algorithms, and enhancing maize pest detection accuracy significantly [29]. The multiple-scale attention

method is utilized for the recognition of pest regions with 77.04 mAP [33]. The paper [34] demonstrates achieving 54.19 mAP on the IP102 dataset, showcasing superior performance in real-time pest detection using deep convolutional neural networks for agricultural applications. The proposed integrated model achieves an 85.2 mAP in detecting pests and diseases, surpassing other comparative models in crop pest and disease identification [35]. The Faster-PestNet studies [36], with an average accuracy of 82.43%. Calculated features are utilized by the Faster-PestNet architecture to perform the localization and classification of the pest in the input images. This leads to the better performance of the Faster-PestNet model on the difficult to handle IP102 dataset regarding both localization and classification. Keeping this in mind the proposed YOLOCSP-PEST model gets even better results in comparison with existing models currently implemented in the same research area after we fine-tuned and trained it.

**Table 2:** YOLOCSP-PEST comparison with new approaches

Reference	Year	Method	mAP (%)
Song et al. [31]	2023	YOLO-Pest	64.6
Ayan et al. [32]	2020	CornerNetwith DenseNet-100	57.23
Yang et al. [29]	2023	Maize-YOLO	76.3
Zhang et al. [33]	2022	Multi-scale mixed attention mechanism	77.04
Doan [34]	2023	Deep convolutional neural network	54.19
Li et al. [35]	2023	YOLO	85.2
Ali et al. [36]	2023	Faster-PestNet	82.43
Proposed		YOLOCSP-PEST	88.40

The Fine-tuned YOLOCSP-PEST model training results with loss rate and validation outcomes are shown in Fig. 6.



**Figure 6:** Training/validation results (a) loss of training box, (b) loss of training object, (c) loss of training cls, (d) loss of validation box, (e) loss of validation object, (f) loss of validation cls

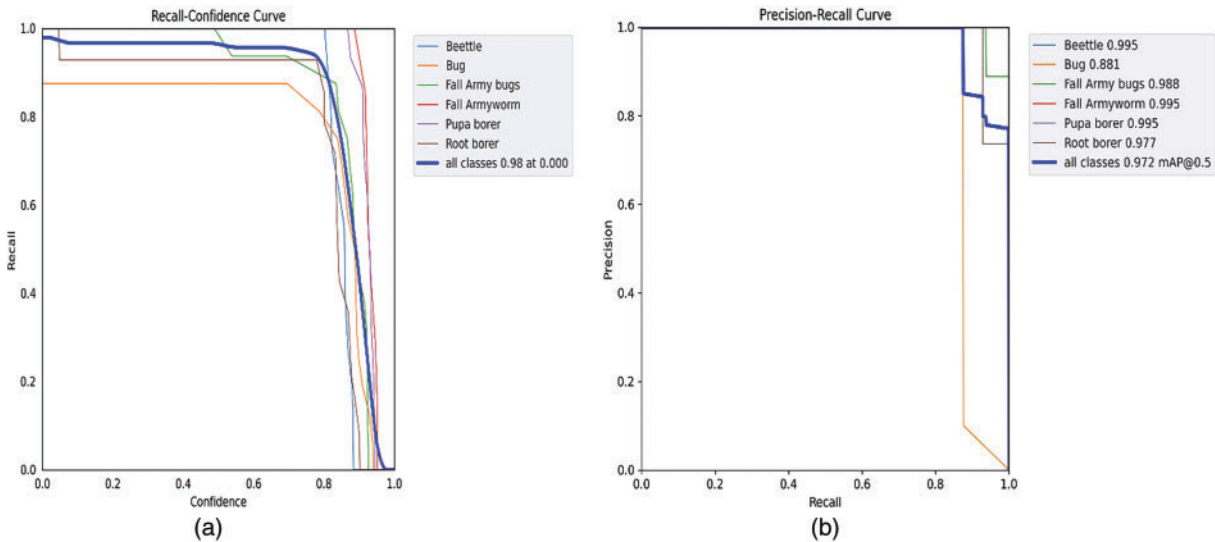
#### 4.6 Generalization Ability Testing

The Local Crop collection compiles a variety of pests that pose a major risk to important crops such as wheat, rice, sugarcane, and maize. To classify these pests into six unique categories “Bug,” “Pupa Borer,” “Root Borer,” “Beetle,” “Fall Army Bug,” and “Army Worm”, we worked closely with experts that we used to further test the robustness of our proposed model YOLOCSP-PEST. With the invaluable assistance of these specialists, these categories were created while considering the distinctive traits and behaviors of every pest. To guarantee precise pest identification in actual agricultural environments, the specialists provide detailed instructions for recognizing every pest and the crop that it is related to. The comprehensive classification procedure sought to enhance the dataset for accurate identification and efficient control of agricultural pests. dataset preparation involved three key stages: image collection with expert assistance to identify and classify pests, preliminary filtering by trained volunteers to remove irrelevant or damaged images and convert them to JPEG format, and professional annotation by agricultural experts using bounding boxes and labels to accurately classify pests in local crops. The preliminary data filtering process was a crucial step in ensuring the quality and relevance of the dataset for our research on pest detection and localization. To ensure consistency and accuracy in this process, we enlisted the help of four dedicated volunteers who underwent comprehensive training before commencing their tasks. For professional dataset annotation, we consulted with experts for better understanding and annotation to be accurate. As the identification of different types of pests found on local crops is outside our expertise, we had to consult with experts from the agriculture field who could accurately identify the pests. With their collaboration, we were able to both identify and classify the pests making the bounding boxes and the labels for each of the identified pests in our images using different annotation tools. The visual representation shown in Fig. 7 illustrates how efficiently are proposed approach identifies and localizes a diverse set of pests from realworld scenarios amid complex background settings.



Figure 7: YOLOCSP-PEST visual results on the local crops dataset

Additionally, Fig. 8a showcases the Recall-Confidence Curve and (b) the Precision-Recall Curve for the local crop dataset, highlighting our approach's exceptional recognition ability across all pest classes with a notably high recall rate.



**Figure 8:** (a) Recall-confidence curve, and (b) Precision-recall curve

So, after fine-tuning the YOLOCSP-PEST, we got better results with higher accuracy. The YOLOCSP-PEST shows a precision of 97.50%, recall of 94.89%, and mAP of 97.18% of the local crop dataset. The comprehensive analysis demonstrates YOLOCSP-PEST ability to detect a wide range of pest classes, along with its strong generalization abilities and capacity to address model overfitting issues.

## 5 Conclusion

YOLOCSP-PEST deep-learning system for agricultural pest identification and classification is presented in our research as a solution for pest localization and classification. Using the IP102 dataset, which consists of images captured in the field for pest detection, the YOLOCSP-PEST model serves as its core network to localize and classify the pests and give a solution for the detection of pests. YOLOCSP-PEST was used for the IP102 dataset giving the mAP of 88.40%. In many tests utilizing the difficult IP102 dataset, YOLOCSP-PEST has demonstrated efficacy in an optimized architecture created to quickly extract characteristics for pest detection. By correctly recognizing and categorizing a wide range of pests with different sizes, colors, orientations, and illumination variations, YOLOCSP-PEST proved its robustness in real-world pest monitoring. Its amazing capacity to generalize for the local crop dataset showed good identification rates across various pest classes. The thorough investigations YOLOCSP-PEST as a pest identification technique have given very good results for the detection of pests on crops. The results highlight the higher generalization capabilities of YOLOCSP-PEST on the local crop dataset. Both quantitative and qualitative studies confirm that the proposed model is effective in identifying a wide range of pests. In addition, we hope to adapt our proposed technique to other agricultural issues such as crop disease identification caused by pests. Further research could explore using transformer-based architectures for improved feature extraction, integrating hyperspectral imaging for better pest and disease differentiation, and utilizing self-supervised learning to lessen reliance on labeled data, all of which would increase the YOLOCSP-PEST model's scalability and robustness. Also, cross-validation could be explored in future work with larger datasets for more robust

evaluation. If we get multiple class datasets, then we could use this for more than pest detection. Beyond pest detection, the YOLO-CSP-PEST model can be adapted for crop disease identification by training it on datasets of diseased crop images. Due to its effective detection and classification of many categories, it is appropriate for real-time plant health monitoring, supporting precision agriculture and early disease control. Future research should concentrate on enhancing YOLO-CSP-PEST by looking at novel methods like feature fusion to boost classification results even further. It is still possible to extend the recommended technique's use to other agricultural areas, such as identifying crop diseases caused by pests, through more research and practical application.

**Acknowledgement:** We would like to thank King Saud University, Riyadh, Saudi Arabia to support this research through the Researchers Supporting Project under Grant RSPD2025R697.

**Funding Statement:** This research was supported by King Saud University, Riyadh, Saudi Arabia, through the Researchers Supporting Project under Grant RSPD2025R697.

**Author Contributions:** The authors confirm their contribution to the paper as follows: Conceptualization and Methodology: Farooq Ali; Data Collection: Farooq Ali, Huma Qayyum, Muhammad Javed Iqbal; Analysis of Results: Farooq Ali, Kashif Saleem, Iftikhar Ahmad; Draft Manuscript Preparation: Farooq Ali, Huma Qayyum, Muhammad Javed Iqbal; Supervision: Huma Qayyum, Muhammad Javed Iqbal. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Ethics Approval:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest to report regarding the present study.

## References

1. 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(7):101460. doi:10.1016/j.ecoinf.2021.101460.
2. Liu J, Wang X. Plant diseases and pests detection based on deep learning: a review. *Plant Methods.* 2021;17(1):22. doi:10.1186/s13007-021-00722-9.
3. Liu J, Wang X. Tomato diseases and pests detection based on improved Yolo V3 convolutional neural network. *Front Plant Sci.* 2020;11:898. doi:10.3389/fpls.2020.00898.
4. Kuzuhara H, Takimoto H, Sato Y, Kanagawa A. Insect pest detection and identification method based on deep learning for realizing a pest control system. In: 2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE); 2020 Sep 23–26; Chiang Mai, Thailand: IEEE; 2020. 709–14.
5. Tang Z, Lu J, Chen Z, Qi F, Zhang L. Improved Pest-YOLO: real-time pest detection based on efficient channel attention mechanism and transformer encoder. *Ecol Inform.* 2023;78(9):102340. doi:10.1016/j.ecoinf.2023.102340.
6. Nanni L, Manfè A, Maguolo G, Lumini A, Brahmam S. High performing ensemble of convolutional neural networks for insect pest image detection. *Ecol Inform.* 2022;67(2):101515. doi:10.1016/j.ecoinf.2021.101515.
7. Setiawan A, Yudistira N, Wihandika RC. Large scale pest classification using efficient Convolutional Neural Network with augmentation and regularizers. *Comput Electron Agric.* 2022;200(3):107204. doi:10.1016/j.compag.2022.107204.
8. Rimal K, Shah KB, Jha AK. Advanced multi-class deep learning convolution neural network approach for insect pest classification using TensorFlow. *Int J Environ Sci Technol.* 2023;20(4):4003–16. doi:10.1007/s13762-022-

9. Amin J, Anjum MA, Zahra R, Sharif MI, Kadry S, Sevcik L. Pest localization using YOLOv5 and classification based on quantum convolutional network. *Agriculture*. 2023;13(3):662. doi:10.3390/agriculture13030662.
10. Liu X, Min W, Mei S, Wang L, Jiang S. Plant disease recognition: a large-scale benchmark dataset and a visual region and loss reweighting approach. *IEEE Trans Image Process*. 2021;30:2003–15. doi:10.1109/TIP.2021.3049334.
11. Chen CJ, Huang YY, Li YS, Chang CY, Huang YM. An AIoT based smart agricultural system for pests detection. *IEEE Access*. 2020;8(9):180750–61. doi:10.1109/ACCESS.2020.3024891.
12. Peng Y, Wang Y. CNN and transformer framework for insect pest classification. *Ecol Inform*. 2022;72:101846. doi:10.1016/j.ecoinf.2022.101846.
13. Hassan SM, Maji AK. Plant disease identification using a novel convolutional neural network. *IEEE Access*. 2022;10:5390–401. doi:10.1109/ACCESS.2022.3141371.
14. Wu Y, Feng X, Chen G. Plant leaf diseases fine-grained categorization using convolutional neural networks. *IEEE Access*. 2022;10:41087–96. doi:10.1109/ACCESS.2022.3167513.
15. Lin S, Xiu Y, Kong J, Yang C, Zhao C. An effective pyramid neural network based on graph-related attentions structure for fine-grained disease and pest identification in intelligent agriculture. *Agriculture*. 2023;13(3):567. doi:10.3390/agriculture13030567.
16. Liu L, Xie C, Wang R, Yang P, Sudirman S, Zhang J, et al. Deep learning based automatic multiclass wild pest monitoring approach using hybrid global and local activated features. *IEEE Trans Ind Inf*. 2021;17(11):7589–98. doi:10.1109/TII.2020.2995208.
17. Liu L, Wang R, Xie C, Yang P, Wang F, Sudirman S, et al. PestNet: an end-to-end deep learning approach for large-scale multi-class pest detection and classification. *IEEE Access*. 2019;7:45301–12. doi:10.1109/ACCESS.2019.2909522.
18. Wang DW, Deng LM, Ni JG, Gao JY, Zhu HF, Han ZZ. Recognition pest by image-based transfer learning. *J Sci Food Agric*. 2019;99(10):4524–31. doi:10.1002/jsfa.9689.
19. Xia D, Chen P, Wang B, Zhang J, Xie C. Insect detection and classification based on an improved convolutional neural network. *Sensors*. 2018;18(12):4169. doi:10.3390/s18124169.
20. Li Z, Jiang X, Jia X, Duan X, Wang Y, Mu J. Classification method of significant rice pests based on deep learning. *Agronomy*. 2022;12(9):2096. doi:10.3390/agronomy12092096.
21. Sabanci K, Aslan MF, Ropelewska E, Unlersen MF, Durdu A. A novel convolutional-recurrent hybrid network for sunn pest-damaged wheat grain detection. *Food Anal Meth*. 2022;15(6):1748–60. doi:10.1007/s12161-022-02251-0.
22. Roy AM, Bose R, Bhaduri J. A fast accurate fine-grain object detection model based on YOLOv4 deep neural network. *Neural Comput Appl*. 2022;34(5):3895–921. doi:10.1007/s00521-021-06651-x.
23. Liu H, Zhan Y, Xia H, Mao Q, Tan Y. Self-supervised transformer-based pre-training method using latent semantic masking auto-encoder for pest and disease classification. *Comput Electron Agric*. 2022;203:107448. doi:10.1016/j.compag.2022.107448.
24. Wang R, Jiao L, Xie C, Chen P, Du J, Li R. S-RPN: sampling-balanced region proposal network for small crop pest detection. *Comput Electron Agric*. 2021;187:106290. doi:10.1016/j.compag.2021.106290.
25. Wei D, Chen J, Luo T, Long T, Wang H. Classification of crop pests based on multi-scale feature fusion. *Comput Electron Agric*. 2022;194(12):106736. doi:10.1016/j.compag.2022.106736.
26. Bollis E, Maia H, Pedrini H, Avila S. Weakly supervised attention-based models using activation maps for *Citrus* mite and insect pest classification. *Comput Electron Agric*. 2022;195:106839. doi:10.1016/j.compag.2022.106839.
27. Tian Y, Wang S, Li E, Yang G, Liang Z, Tan M. MD-YOLO: multi-scale Dense YOLO for small target pest detection. *Comput Electron Agric*. 2023;213(9):108233. doi:10.1016/j.compag.2023.108233.
28. Li D, Ahmed F, Wu N, Sethi AI. YOLO-JD: a deep learning network for jute diseases and pests detection from images. *Plants*. 2022;11(7):937. doi:10.3390/plants11070937.
29. Yang S, Xing Z, Wang H, Dong X, Gao X, Liu Z, et al. Maize-YOLO: a new high-precision and real-time method for maize pest detection. *Insects*. 2023;14(3):278. doi:10.3390/insects14030278.
30. Wang CY, Bochkovskiy A, Liao HM. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2023 Jun 17–24; Vancouver, BC, Canada: IEEE. doi:10.1109/CVPR52729.2023.00721

31. Song H, Yan Y, Xie M, Duan D, Xie Z, Li Y. Agricultural pest small target detection algorithm based on improved YOLOv5 architecture. 2023. doi:10.21203/rs.3.rs-3109779/v1.
32. Ayan E, Erbay H, Varçın F. Crop pest classification with a genetic algorithm-based weighted ensemble of deep convolutional neural networks. *Comput Electron Agric.* 2020;179(4):105809. doi:10.1016/j.compag.2020.105809.
33. Zhang W, Sun Y, Huang H, Pei H, Sheng J, Yang P. Pest region detection in complex backgrounds via contextual information and multi-scale mixed attention mechanism. *Agriculture.* 2022;12(8):1104. doi:10.3390/agriculture12081104.
34. Doan TN. Large-scale insect detection with fine-tuning YOLOX. *Inter J Membrane Sci Tech.* 2023;10(2):892–915. doi:10.15379/ijmst.v10i2.1306.
35. Li M, Cheng S, Cui J, Li C, Li Z, Zhou C, et al. High-performance plant pest and disease detection based on model ensemble with inception module and cluster algorithm. *Plants (Basel).* 2023;12(1):200. doi:10.3390/plants12010200.
36. Ali F, Qayyum H, Iqbal MJ. Faster-PestNet: a lightweight deep learning framework for crop pest detection and classification. *IEEE Access.* 2023;11:104016–27. doi:10.1109/ACCESS.2023.3317506.