Advanced Image Recognition Techniques for Crop Pest Detection Using Modified YOLO-v3

Dechao Guo[1]*, Hao Zhang[2]
School of Public Health and Management, Guangzhou University of Chinese Medicine, Guangzhou 510006, Guangdong, China[1]
Guangzhou Center for Disease Control and Prevention (Guangzhou Health Supervision Institute), Guangzhou 510440, Guangdong, China[2]

***Abstract*—Accurate and efficient detection of agricultural pests is crucial for crop protection and pest control. This study addresses the limitations of traditional pest detection methods, such as weak detection capabilities and high computational demands, by proposing an improved image recognition system based on the YOLO-v3 algorithm. The research focuses on enhancing pest detection accuracy through deep learning techniques, specifically by modifying the YOLO-v3 model with the ISODATA clustering algorithm, DenseBlock enhancements, and the ELU activation function. A dataset of 13,000 images representing six common crop pests was created and expanded using various image augmentation techniques. The modified YOLO-v3 model was trained and evaluated on this dataset, achieving a higher mean Average Precision (mAP) of 89.7% and faster recognition speed compared to Faster-RCNN, SSD-300, and the original YOLO-v3 model. Finally, the improved model demonstrated a recognition speed of 27 frames per second (fps), significantly outperforming other detection models in both accuracy and speed. The proposed method offers a superior solution for real-time pest detection in agricultural settings, combining high accuracy with computational efficiency. Future work will explore the application of optimization algorithms to further enhance the robustness and generalizability of the system across diverse pest detection scenarios.**

Keywords*—*Feature detection algorithm; YOLO-v3 network; image recognition technology; crop pest detection applications

I. Introduction

Identifying and detecting crop pests is a challenging task [1-3]. To address this, there are two main approaches: traditional machine learning-based methods and deep learning-based methods [4]. These methods rely on digital image processing and pattern recognition technology [5]. Two major steps in traditional machine learning-based pest identification and detection systems are feature extraction and pattern recognition [6]. Li et al. [7] proposed an algorithm for orchard pest gesture characteristic representation learning to identify automatic trapping target pests, and its recognition rate reached 86.7%. Han and He [8] studied a set of stationary fast identification and diagnosis methods on the identification of field pests, and achieved the effect of real-time identification and diagnosis. Liang et al. [9] for the specificity of rice pests, fused the global features of the image and local gradient direction histogram features, proposed a pest classification and identification method based on support vector machine, and obtained an accuracy rate of 91.4%. Sanghavi et al. [10] used six invariant moments to extract the shape features of pests, and ARTMAP neural network to classify the pests. Han et al. [11] designed a hierarchical automatic pest identification system, and the pest identification rate reached 93% under a variety of categories. Deep learning based crop pest identification method is an end-to-end extraction of high quality feature representation of pests utilizing picture detection and recognition algorithms based on deep learning techniques. Chen et al. [12] proposed an improved residual network pest image recognition method, using the improved convolutional neural network in depth of residual block, adding high-resolution convolutional layer and the corresponding channel, so that the recognition rate of 91.4%. Cheng et al. [13] for the specific pest detection problem, proposed a deep convolutional network based on grain storage pest image recognition method. Renault et al. [14] designed a kind of coarse and fine convolutional neural network and applied it to the field aphid detection and identification problem, which improved the detection and identification accuracy. Lü et al. [15] used deep learning algorithms to detect and identify 15 kinds of beetles in food, and its accuracy rate reached 83.3%. Crop pest detection is a sub-task of target detection, and despite the use of image recognition technology to solve the task of pest identification and detection in different scenarios, the accuracy rate is still limited, mainly in the following aspects [16]: 1) the existing detection and identification methods only focus on the whole picture classification, and less detection for tasks such as pest occurrence location and pest number; 2) the current method test validation only uses its own constructed dataset, and its expandability and generalizability need to be improved; 3) less research on pest images in complex backgrounds, and the practicality of existing methods is poor.

This work provides a detailed analysis of the technical and application challenges associated with crop pest identification, taking into account enhanced feature detection and deep learning algorithms. The proposed approach for field crop pest detection is based on these challenges. The primary contributions of this study are as follows:

- Gathering image data of crop pests in the field and creating the necessary dataset;
- Integrating deep learning algorithms to create a target detection model based on the improved YOLO-v3 algorithm [17] and using it to solve the pest detection problem;
- Utilizing the dataset CPXJ to confirm the efficacy of the suggested algorithm in this study. The findings indicate

*Corresponding Author.

that the improved YOLO-v3 algorithm has higher detection accuracy when compared to other recognition models.

- To address the limitations of existing crop pest detection methods, this study proposes an enhanced YOLO-v3-based image recognition approach.

The structure of this study is as follows: Section II details the acquisition and augmentation of the crop pest image dataset. It describes the construction and labeling of the dataset. Section III introduces the standard YOLO-v3 model and the specific improvements made, including the integration of ISODATA clustering, DenseBlock, and the ELU activation function. Section IV presents the experimental setup, parameter configurations, and comparative evaluation results with other detection models. Finally, Section V discusses the conclusion, summarizing the advantages of the proposed approach and outlining potential future research directions to enhance model generalizability and robustness in real-world applications.

II. CROP PEST DATASET ACQUISITION SELECTION

Since the pest detection problem studied in this study is a target detection problem, the first step is to analyze the collection of crop pest datasets.

A. Data Acquisition

The study data for this work were gathered over a five-month period, from May 2019 to October 2019, at the Institute of Agricultural Sciences' experimental base. In order to improve the robustness and generality of the validation process, the current time period is separated into three time nodes every morning, noon and afternoon. Through the collection and analysis, there are six prevalent crop pests in the test base [18], as indicated in Table I.

This research employs web crawler technology to acquire image data of six types of crop pests, respectively, because there aren't many crop pests in the experimental base. To maintain the diversity of the data set, this is done, and the specific schematic diagram of the original photographs is presented in Fig. 1.

TABLE I. DESCRIPTION OF EXPERIMENTAL CROP PEST DATA

No.	Name	Harm
1	Cabbage greenfly	Bok choy, oleander, cauliflower, etc.
2	Moth	Peppers, cabbages, apple trees, pear trees, etc.
3	*Morus albopictus* (type of grasshopper)	Apple trees, pear trees, date palms, etc.
4	Three-spotted blind stink bug	Tomatoes, corn, cotton, soybeans, etc.
5	Green stink bug	Apple trees, pear trees, cotton, cucumbers, etc.
6	Leafhopper	Cabbage, sugar beet, apple trees, etc.

Fig. 1. Crop pest image data.

B. Image Data Expansion

Since the collection and web crawler image dataset is insufficient for deep learning training, this study expands the dataset using techniques like panning, mirroring, adding noise, and making light and dark changes. This increases the model's robustness and generalization ability. The specific operation is as follows.

1) Panning method: pan the image 50 pixels to the upper right, lower right, upper left, lower left, the place after the panning is supplemented with black, after the panning, it can generate 5 different images, which contains the original image that has not been panned. The specific operation is shown in Fig. 2, and the panning equation is as follows [Eq. (1)]:

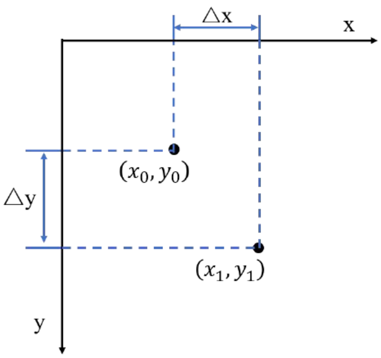


Fig. 2. Schematic diagram of translation

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta x \\ 0 & 1 & \Delta y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} \quad (1)$$

where, x_1 and y_1 denote the pixel position after translation, Δx and Δy denote the pixel translation amount, and x_0 and y_0 denote the original pixel position.

2) Mirroring method: each image is mirrored to the left, right, up and down respectively, and 5 different images (including the original image) are generated after mirroring, and the mirroring schematic diagram is shown in Fig. 3. The equation for image level mirroring method is as follows [Eq. (2)]:

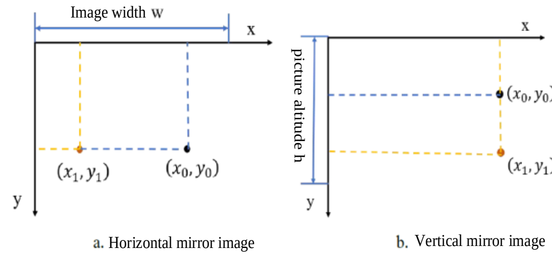


Fig. 3. Mirroring

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & w \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} \quad (2)$$

where, w is the image width.

The equation for calculating the image vertical mirroring method is as follows [Eq. (3)]:

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & h \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} \quad (3)$$

where, h denotes the image height.

3) Add noise method: add pretzel noise or Gaussian noise to each image data, the number of noise points of the added pretzel noise is a random number between 3000 and 5000.

4) Brightness and darkness transformation method: each image will be adjusted to different degrees of brightness and darkness, using four levels of brightness and darkness division, using OpenCV and Numpy [19] for each image matrix operation to get the image data with different degrees of brightness and darkness.

Through the above four sample expansion methods, the number of samples in the crop pest image dataset was expanded to 13,000, and specific examples of the expansion are shown in Table II.

TABLE II. EXAMPLE OF IMAGE DATA EXPANSION

Expansion Methods	**Original figure**	**Transformed image**		
Panning				
Mirroring				

Add noise				
Brightness and darkness changes				

C. Data Labelling

Firstly, the dataset was size-unified, compressed to 416×416 and saved in JPG format. Secondly, the dataset was separated into a training set and a testing set with a 4:1 ratio. Finally, the Labeling annotation tool [20] was used to annotate the dataset, as illustrated in Fig. 4. In this study, based on the Overall labeling method, a Non-overall labeling method is developed, and the labeling situation is specifically shown in Fig. 5.

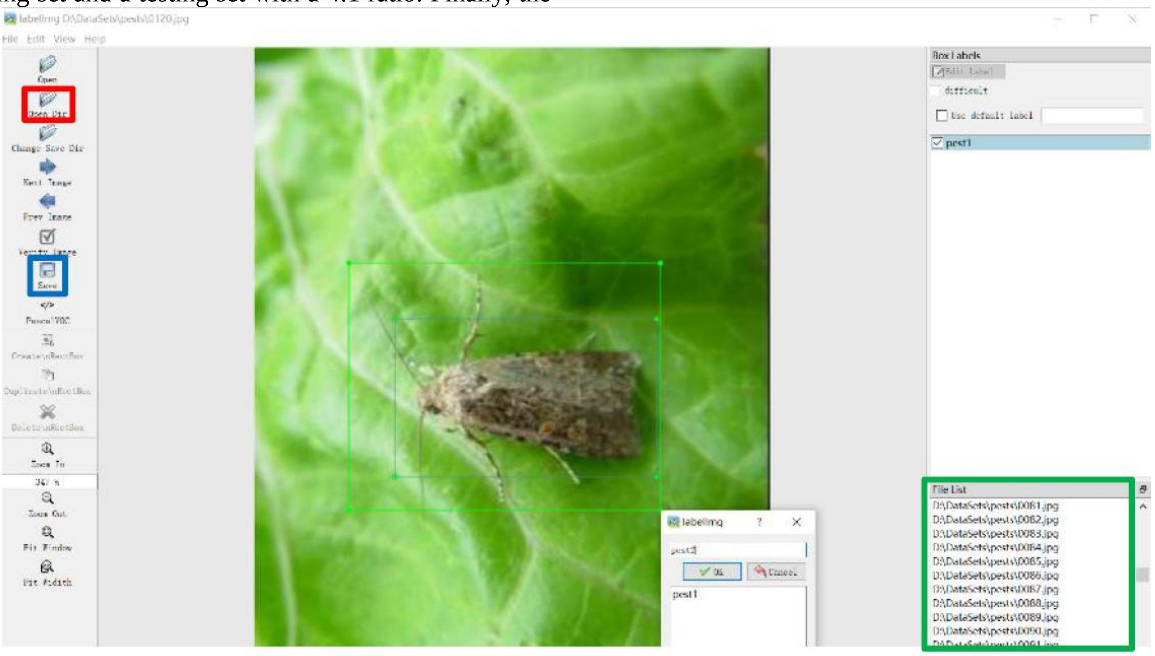


Fig. 4. Labeling annotation process

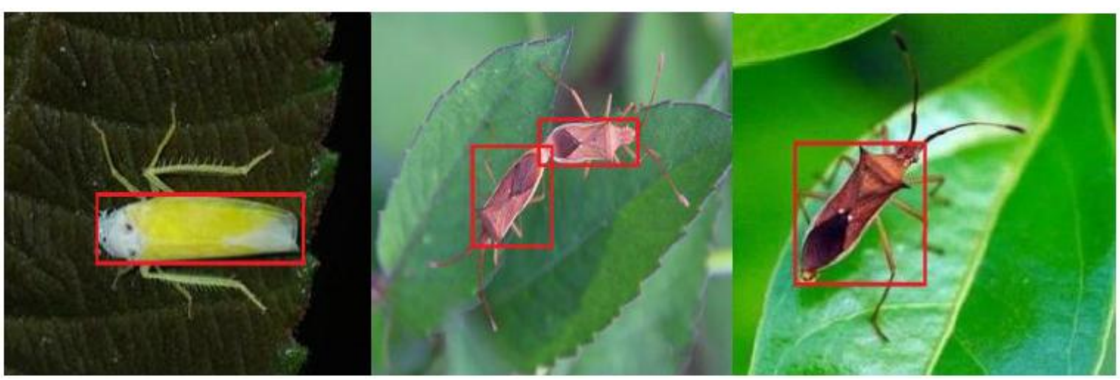


Fig. 5. Labeling

D. Construction of Data Sets

The acquired crop pest photos are increased by data, six crop pests have the same amount of data, and the six pest image data totals 13,000 images, which are then randomly assigned into a training dataset of 10,000 and a test set of 3,000 according to an estimated 4:1 ratio. The data image set and labels are produced as crop pest image datasets (CPXJ-Datasets).

III. Image Recognition Based on Improved YOLO-v3

A. YOLO-v3 Algorithm

1) YOLO-v3 algorithm structure: YOLO-v3 (You Only Look Once version 3) [21] is a popular target detection algorithm known for its fast detection speed and relatively high accuracy. The key features of YOLO-v3 include the use of multi-scale prediction to improve detection of targets of different sizes and the use of Darknet-53 [22] as its feature extractor, which is a deeper convolutional neural network than previous versions.YOLO-v3 is capable of predicting both bounding box and category probabilities, and provides better detection performance while maintaining real-time performance. The structure of the YOLO-v3 algorithm is shown in Fig. 6.

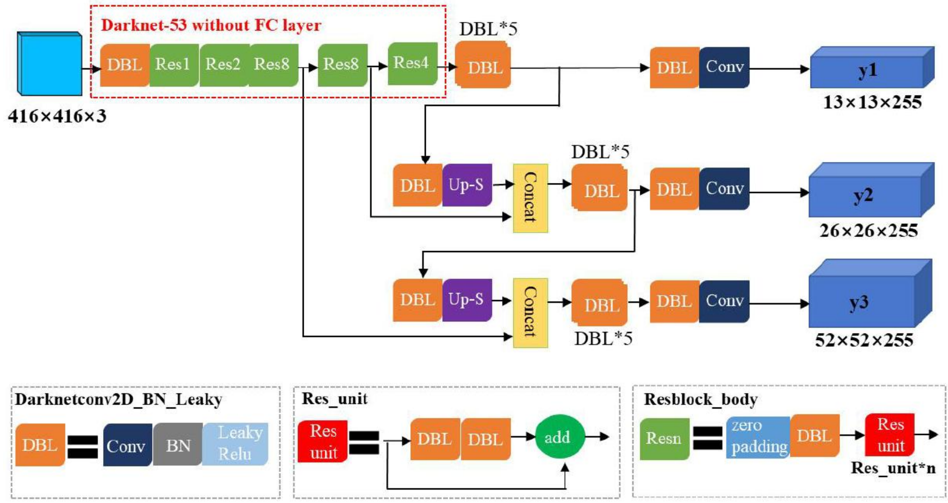


Fig. 6. YOLO-v3 algorithm structure

As seen from Fig. 6, the red dashed portion represents the YOLO-v3 algorithm feature extractor Darknet-53 [22], which is the main component of the method. The Resblock-body section of the algorithm is made up of various residual structures (Res-unit), DBL structures, and zero-padding structures.

2) K-means clustering algorithm: In order to avoid the detection model to detect the wrong target box in the training and learning time, and to speed up the model convergence, by using a K-means clustering algorithm [23] in the labelled ground true shape, an existence of a certain regularity can be found, specifically as shown in Fig. 7.

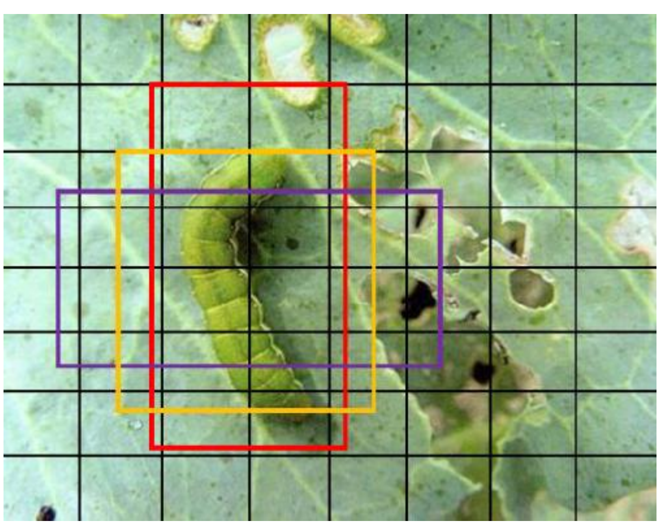


Fig. 7. Target box shape

3) Darknet53 feature extractor: The YOLOv3 target detection algorithm uses a deep convolutional neural network architecture called Darknet53 [22]. It's 53 convolutional layers, including multiple residual blocks, help to mitigate the issue of gradient vanishing in deep networks and facilitate network training. The structure of Darknet53 is depicted in Fig. 8. Darknet53 does not use pooling and fully connected layers, but downsamples the feature map by altering the step size of the convolutional kernel.

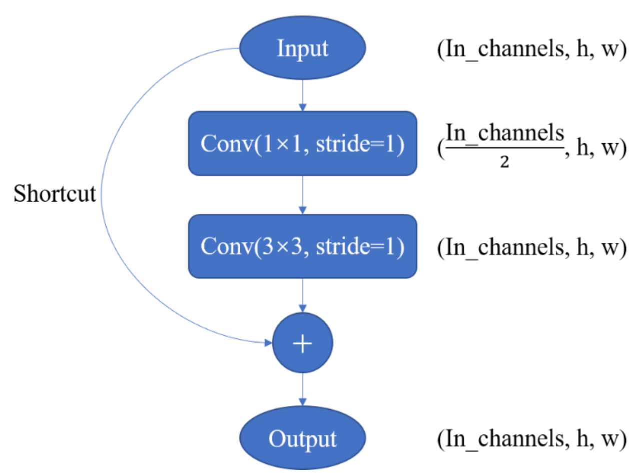


Fig. 8. Darknet53 structure.

4) RPN network: The RPN network is used by YOLO-v3 in order to avoid doing a lot of convolutional calculations [24]. Its primary goals are to extract the feature map that the network has acquired, extract multi-dimensional feature vectors from

Conv1, pass it through a number of target regions, including Conv3-FC1 and Conv4-FC2, and after each matrix target region has a regional target score. The integrated score is then passed on to the following RoI-Pooling operation. Fig. 9 depicts the RPN network.

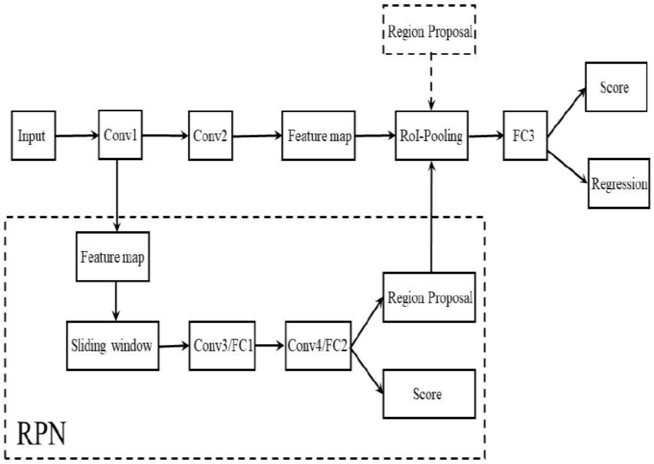


Fig. 9. RPN network structure

Combined with the crop pest detection data in this study, the image is input, and the target area is obtained through a computational process (Fig. 10).

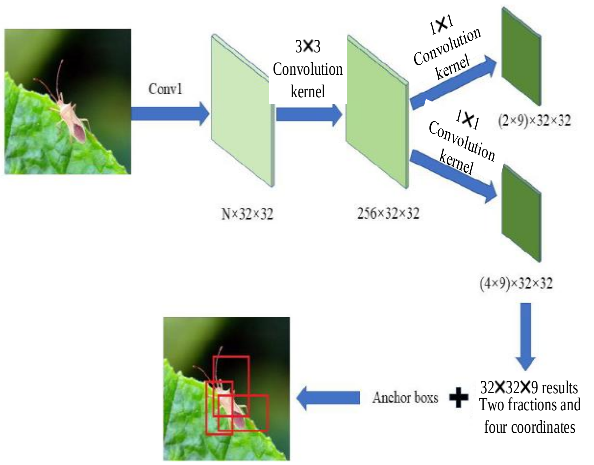


Fig. 10. RPN calculation flow.

5) Boundary box regression: In order to effectively detect ringed agricultural pest targets, the notion of Intersection over Union (IoU) is introduced, which is represented schematically in Fig. 11. In Fig. 11, it can be seen that the green box is the real box of the target using Labeling annotation tool and retrograde labeling, and the red box is the box predicted by the trained target detection model. IoU is the area intersection over union operation, and the specific form of calculation is shown in Fig. 12. From Fig. 12, it can be observed that the more accurate the projected box of the target detection prediction model is, the greater the IoU value. When the IoU value is more than 0.5, the projected box is deemed as accurate; otherwise, the red predicted box is fine-tuned to bring it close to the true green box.

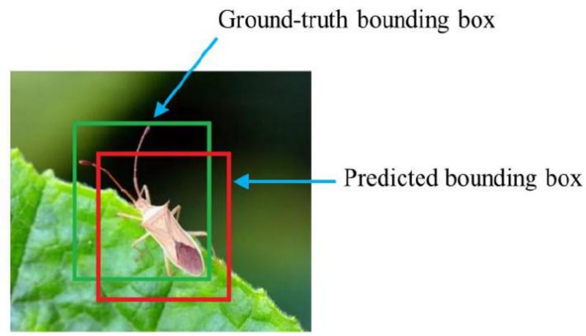


Fig. 11. Intersection over Union (IoU).

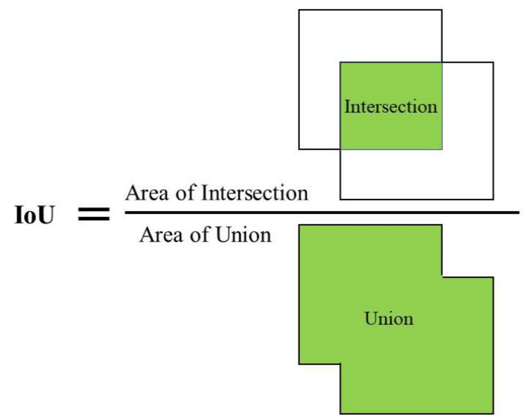


Fig. 12. IoU Calculation.

6) Loss function: The YOLO-v3 loss function consists of four components [25], i.e., (x, y) loss, (w, h) loss, confidence loss, and category loss, and the total loss is expressed as follows [Eq. (4)]:

$$L = L_{xy} + L_{wh} + L_{conf} + L_{class} \tag{4}$$

where, L is the total YOLO-v3 loss, L_{xy} is the (x, y) loss, L_{wh} is the (w, h) loss, L_{conf} is the confidence loss, and L_{class} is the category loss.

The bounding box position loss L_{xy} is calculated as follows [Eq. (5)]:

$$L_{xy} = \sum_{i=0}^{S^2} \sum_{j=0}^{B} \lambda_{ij}^{obj} \left[\left(x_i - \hat{x}_i\right)^2 + \left(y_i - \hat{y}_i\right)^2 \right] \tag{5}$$

where, λ_{ij}^{obj} is the j^{th} bounding box predicted by the i^{th} grid to detect the target to be detected, takes the value of 1, otherwise a smaller weight value of 0.1 or 0; S^2 is the total number of grids after the input avatar is rasterized; B is the number of bounding boxes predicted by individual grids, takes the value of 3; (x_i, y_i) denotes the predicted bounding box centroid position coordinates; $(\hat{x}_i, \hat{y}_i)$ denotes the actual bounding box centroid position coordinates.

The bounding box size loss L_{wh} is calculated as follows [Eq. (6)]:

$$L_{wh} = \sum_{i=0}^{S^2}\sum_{j=0}^{B}\lambda_{ij}^{obj}\left[\left(w_i - \hat{w}_i\right)^2 + \left(h_i - \hat{h}_i\right)^2\right] \tag{6}$$

where, $\left(w_i, h_i\right)$ indicates the coordinates of the predicted width and height of the bounding box and $\left(\hat{w}_i, \hat{h}_i\right)$ indicates the coordinates of the actual width and height of the bounding box.

The loss of confidence L_{conf} is calculated as follows [Eq. (7)]:

$$L_{conf} = \sum_{i=0}^{S^2}\sum_{j=0}^{B}\lambda_{ij}^{obj}\left(C_i - C_i^*\right)^2 + I_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\lambda_{ij}^{noobj}\left(C_i - C_i^*\right)^2 \tag{7}$$

where, I_{noobj} indicates that there is no loss in the control cell to prevent model instability due to gradient explosion. C_i denotes the actual detection target confidence, C_i^* denotes the detection target confidence, and $C_i^* = \Pr\left(obj\right) * IoU_{pred}^{truth}$.

Category losses L_{class} are calculated as follows [Eq. (8)]:

$$L_{class} = \sum_{i=0}^{S^2}\lambda_{ij}^{obj}\sum_{c\in classes}\left[p_i^*(c)\log\left(p_i(c)\right) + \left(1 - p_i^*(c)\right)\log\left(1 - p_i(c)\right)\right] \tag{8}$$

where, c denotes the category to which the detected target belongs, $p_i^*\left(c\right)$ denotes the actual probability that a target belongs to the category c when it is detected by the ith network, and $p_i\left(c\right)$ denotes the predicted probability that a target belongs to the category c when it is detected by the ith network.

B. *Improvement of the YOLO-v3 Algorithm*

In order to increase the detection accuracy of YOLO-v3 network, the advanced clustering algorithm ISODATA clustering algorithm, is employed for anchor boxes collection, the ELU activation function is used in YOLO-v3, and the Darknet53 structure is improved to adapt to the dataset CPXJDatasets.

1) ISODATA clustering algorithm: In order to overcome the shortcomings of K-means clustering algorithm, this study adopts ISODATA clustering algorithm [26] to cluster the anchor boxes.The flowchart of ISODATA clustering algorithm is shown in Fig. 13.

The ISODATA clustering algorithm clusters the anchor boxes to obtain 9 prior frames, the specific results are shown in Table III.

2) Improvement of Darknet53 structure: In order to reduce the amount of computation, in this study, DenseBlock [27] is added to Darknet53 to improve the performance of YOLO-v3. The structure of DenseBlock is shown in Fig. 14, which deepens the feature extraction network's ability of extracting features in Darknet53 with fewer parameters, which makes it easier to train. Before and after Darknet53 improvement is given in Fig. 15.

To test the effectiveness of the improved Darknet53 module, the original images are input and analyzed, and the results shown in Fig. 16 are obtained. From Fig. 16, it can be seen that the improved Darknet53 module not only adds more semantic information to the output feature map at each layer, but also enhances the expression ability of the feature map.

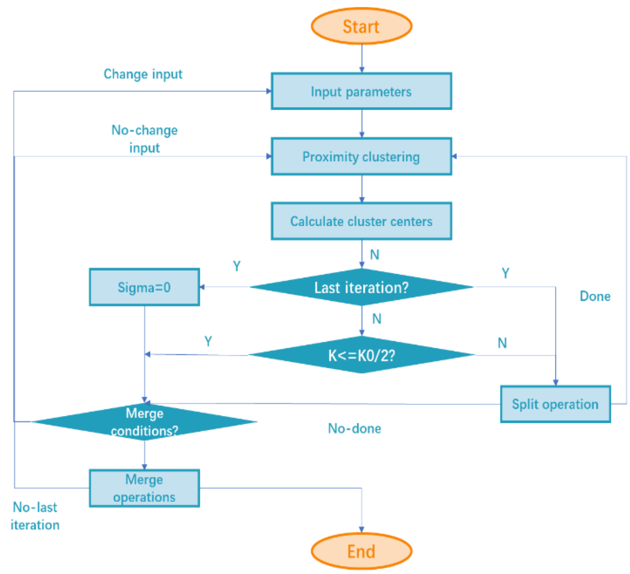


Fig. 13. Flowchart of ISODATA clustering algorithm.

TABLE III. PRIOR FRAME ASSIGNMENT AFTER CLUSTERING PROCESS

Characteristic graph	13*13	26*26	52*52
Experience the wild	oldest	middle	few
A priori framework	(240*100) (330*152) (412*386)	(166*112) (174*162) (192*243)	(46*49) (56*102) (107*146)

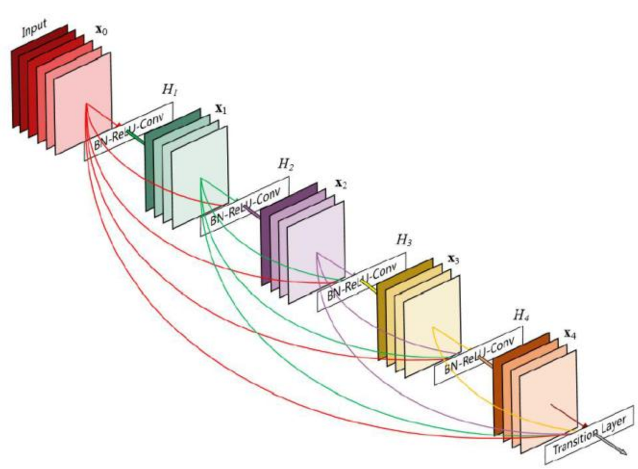
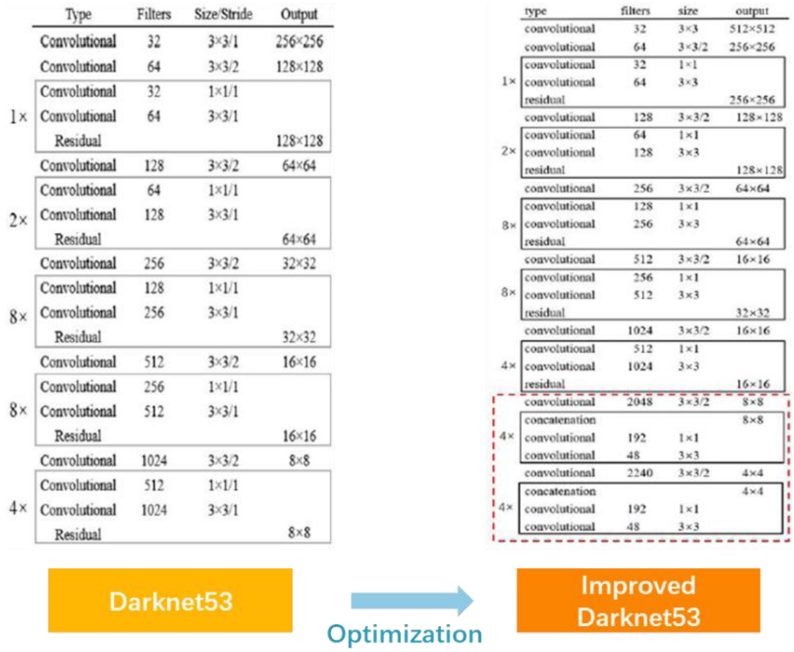


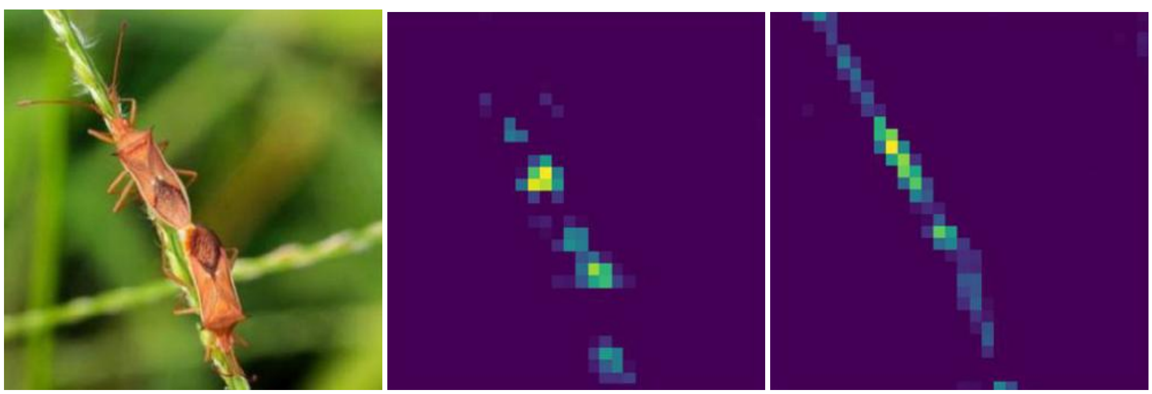
Fig. 14. DenseBlock structure.



	Type	Filters	Size/Stride	Output
	Convolutional	32	3×3/1	256×256
	Convolutional	64	3×3/2	128×128
1×	Convolutional	32	1×1/1	
	Convolutional	64	3×3/1	
	Residual			128×128
	Convolutional	128	3×3/2	64×64
2×	Convolutional	64	1×1/1	
	Convolutional	128	3×3/1	
	Residual			64×64
	Convolutional	256	3×3/2	32×32
8×	Convolutional	128	1×1/1	
	Convolutional	256	3×3/1	
	Residual			32×32
	Convolutional	512	3×3/2	16×16
8×	Convolutional	256	1×1/1	
	Convolutional	512	3×3/1	
	Residual			16×16
	Convolutional	1024	3×3/2	8×8
4×	Convolutional	512	1×1/1	
	Convolutional	1024	3×3/1	
	Residual			8×8

	type	filters	size	output
	convolutional	32	3×3	512×512
	convolutional	64	3×3/2	256×256
1×	convolutional	32	1×1	
	convolutional	64	3×3	
	residual			256×256
	convolutional	128	3×3/2	128×128
2×	convolutional	64	1×1	
	convolutional	128	3×3	
	residual			128×128
	convolutional	256	3×3/2	64×64
8×	convolutional	128	1×1	
	convolutional	256	3×3	
	residual			64×64
	convolutional	512	3×3/2	16×16
8×	convolutional	256	1×1	
	convolutional	512	3×3	
	residual			32×32
	convolutional	1024	3×3/2	16×16
4×	convolutional	512	1×1	
	convolutional	1024	3×3	
	residual			16×16
	convolutional	2048	3×3/2	8×8
4×	concatenation			8×8
	convolutional	192	1×1	
	convolutional	48	3×3	
	convolutional	2240	3×3/2	4×4
4×	concatenation			4×4
	convolutional	192	1×1	
	convolutional	48	3×3	

Fig. 15. Darknet53 before and after improvements.



(a) Input images (b) Darknet53 output (c) Improvement of Darknet53 output

Fig. 16. Analysis of Darknet53 module result output.

3) Spatial pyramid pooling: In order to enrich the features and increase the feature expression ability, YOLO-v3 introduces the SPP network, i.e., the spatial pyramid pooling structure (Fig. 17). The upgraded YOLO-v3 network is depicted in Fig. 17. The SPP network structure is added before the first fully connected input in the YOLO-v3 network structure, and the results of three times Max pooling are fused to obtain a fixed output for the input of the first fully connected layer, which is a method of fusing three kinds of features with different scales, which results in a wider range of the field of view of the convolution kernel. The fusion of the three characteristics is utilized to remove the effect of inconsistent effective feature information due to the individual variability of agricultural pests.

4) ELU activation function: The original YOLO-v3 network uses a nonlinear activation function Leaky ReLU, and the function image is displayed in Fig. 18(a). The Leaky ReLU activation function was used to avoid the effects that the traditional activation function brings to the model, although it solves the problem that neurons do not learn when they enter the negative region, the rate of neuron learning after Leaky ReLU activation is very slow, which leads to a longer training time. Therefore, the ELU activation function [Fig. 18(b)] is employed instead of Leaky ReLU to speed up the convergence of the network.The particular equation for the ELU activation function is as follows [Eq. (9)]:

$$f(x)=\begin{cases} x & x>0 \\ \alpha\left(\exp(x)-1\right) & x\leq 0 \end{cases} \tag{9}$$

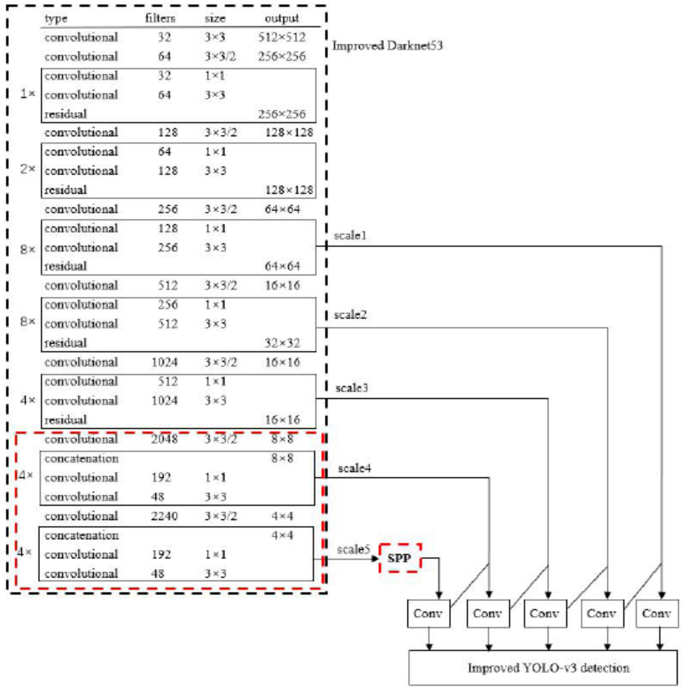
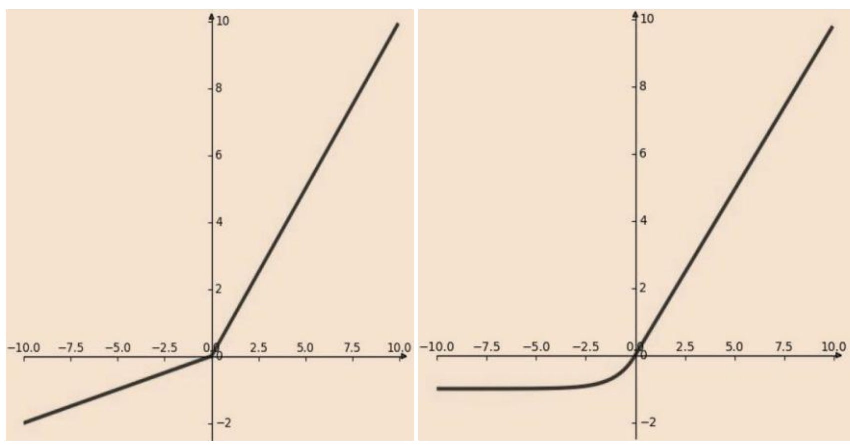


Fig. 17. Improved YOLO-v3 network structure.



(a) Leaky ReLU (b) ELU

Fig. 18. Analysis of the activation function of Leaky ReLU and ELU

IV. Experimental Analysis of Data

A. Experimental Setup

1) Experimental environment parameter setting: The experiments in this study use deep learning techniques to solve the crop pest detection task, and the specific experimental environment is shown in Table IV.

2) Network parameter setting: The validated YOLO-v3 network parameters are designed as shown in Table V. The detection models compared are Faster-RCNN, SSD-300, and YOLO-v3.

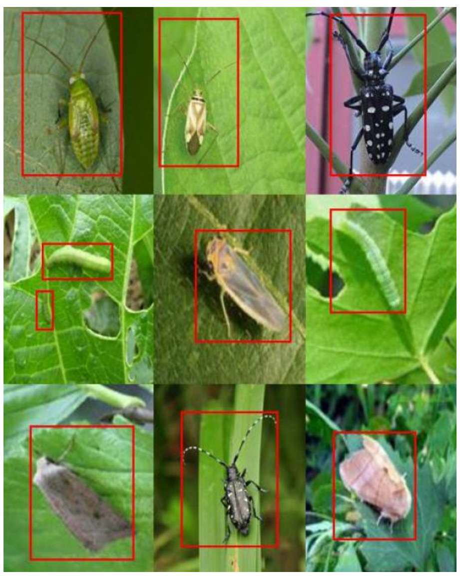
TABLE IV. Experimental Environment Parameter Settings

No.	Experimental Environment Project	Specific settings
1	Programming Development Environment	Python 3.7
2	operating system	Linux Ubuntu 16.72LTS
3	software platform	PyCharm 2019.3.3 Professional, Labelimg 1.8.3, OpenCV 4.2.0.34
4	Hardware Development Environment	Intel(R) Core(TM) i7-9750H CPU @2.60GHz 2.59GHz Processor
5	memory	GTX1080 6GB
6	Deep Learning Development Framework	Keras 2.3.1

TABLE V. ALGORITHM PARAMETER SETTINGS

No.	Parameter	Specific settings
1	Optimization methods	Batch stochastic gradient descent
2	Total number of iterations	30000
3	learning rate	0.01
4	weight decay value	0.0005
5	batch size	64
6	momentum factor	0.99

3) Experimental data set: The agricultural pest detection dataset, CPXJDatasets, contains 13,000 photos of six crop pests, namely, green blind stink bug, three-spotted blind stink bug, cabbage greenfly, leafhopper, moth and mulberry aspen (10,000 training datasets and 3,000 testing datasets). The training set is split into an integral labeling method and a non-integral labeling technique training set, as illustrated in Fig. 19.



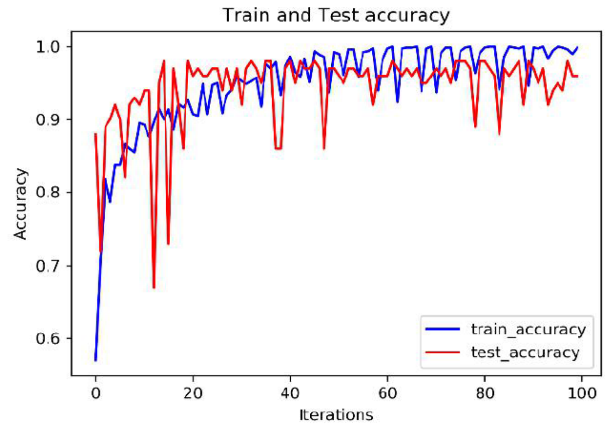
(a) Integral labeling method data set

(b) Non-integral labeling method training set

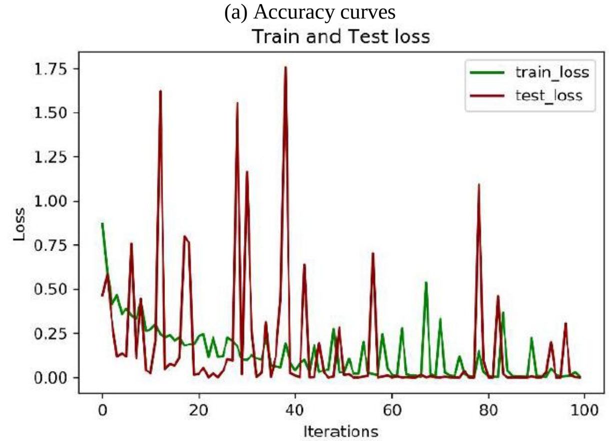
Fig. 19. Integral and Non-integral labeling method training set

B. Analysis of Results

Improved YOLO-v3 model is trained and tested using crop pest detection dataset, CPXJDatasets. The training phase of the network algorithm presented in this study is shown in Fig. 20. From Fig. 20, it can be seen that the accuracy converges to about 0.99 with the rise in the number of iterations, and the loss value drops to near 0 with the increase in the number of iterations.



(a) Accuracy curves



(b) Loss curve

Fig. 20. Accuracy and loss changes during the training process

To assess the efficacy of the enhanced YOLO-v3 network, this paper employs Faster-RCNN, SSD-300, and YOLO-v3 as comparison algorithms. The training and learning tests are employed to compare the results, as illustrated in Table VI. From Table VI, it can be shown that the pest detection accuracy of improved YOLO-v3 is better than other models and the recognition speed is faster than other network models.

TABLE VI. CONTRASTING NETWORK MODEL RESULTS

Test	Recognition rate			
	Faster-RCNN	SSD-300	YOLO-v3	Improvement of YOLO-v3
mAP	89.3 %	81.7 %	85.0 %	89.7%
recognition speed	12 f/s	36 f/s	20 f/s	27 f/s

Fig. 21 presents a schematic visualisation of the output of the convolutional layer findings of the output pest in order to facilitate a more thorough analysis of the experimental results. From Fig. 21, it can be seen that as the number of layers of convolutional network increases, the convolutional output results are overloaded from shallow features to deep semantic features to achieve the ultimate feature results.

The test detection results are provided in Fig. 22. In the CPXJDatasets dataset, 3000 pest photos were used as the test set, and the improved YOLO-v3 algorithm detected the presence of pests in the images, and the detection results met the detection requirements.

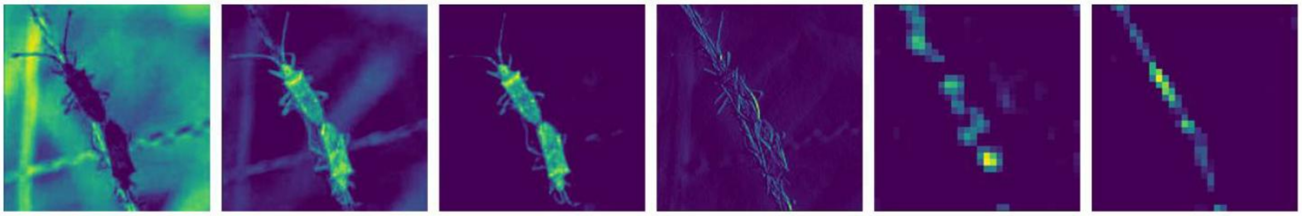


Fig. 21. Visualisation of the output results of the convolutional layer

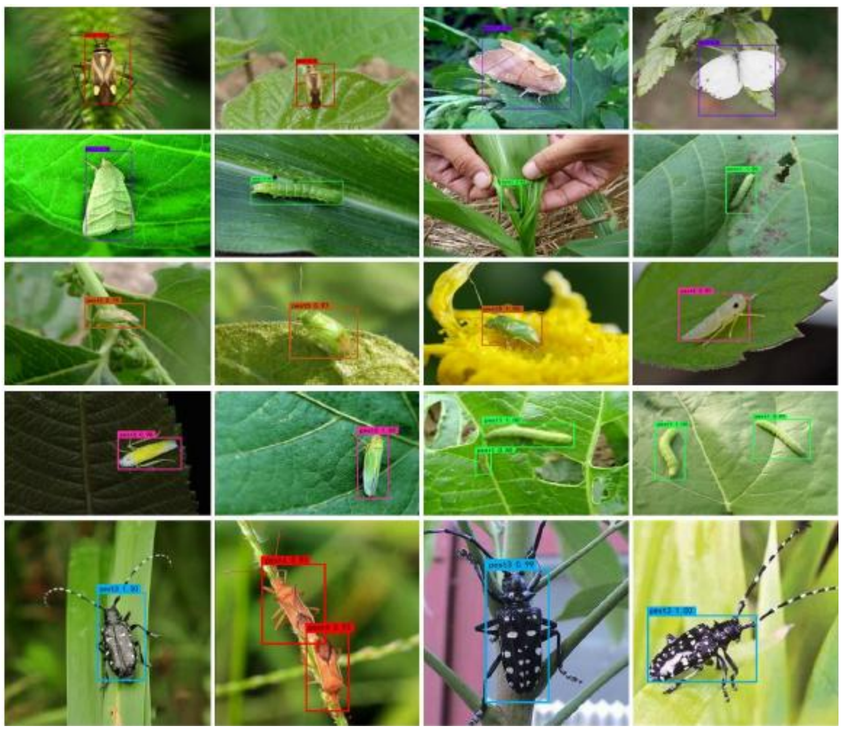


Fig. 22. Partial presentation of test results.

V. CONCLUSION

This study proposes an agricultural pest detection method that is based on an enhanced YOLO-v3 algorithm to address the issue of crop pest detection. The problem of crop pest detection is analyzed in this study, which also gathers data images from experimental fields, expands the dataset through the use of panning, mirroring, adding noise, and adjusting light and dark, combines deep learning algorithms, enhances the YOLO-v3 network from four angles, suggests a detection model based on the enhanced YOLO-v3 network, and compares and validates it using the created dataset, CPXJDatasets. The results reveal that, compared to the models of Faster-RCNN, SSD-300, and YOLO-v3, the pest detection accuracy and recognition speed of the new approach described in this study are better and faster than other models. In the next phase, to further increase the detection accuracy of the proposed method, the intelligent optimization algorithm is utilized to optimise the upgraded network and used on multiple crop pest datasets to improve the robustness and generalization of the system.

In future research, efforts will focus on integrating attention mechanisms and lightweight neural network architectures to further improve detection accuracy and computational efficiency, particularly on mobile or edge devices. Additionally, expanding the dataset to include more pest species and diverse environmental backgrounds will help enhance the model's robustness and applicability in real-world agricultural scenarios. Cross-domain transfer learning and semi-supervised learning techniques will also be explored to reduce reliance on large-scale labeled datasets and improve performance in low-resource settings.

ACKNOWLEDGMENT

This work is supported by Guangzhou Philosophy and Social Science Development Foundation (Grant No. 2024GZGJ272); Guangdong Research Center for Health

Service and Industrial Development of Chinese Medicine Foundation (Grant No. 2025YBA14,2025YBA05); Guangzhou Philosophy and Social Science Development Foundation (Grant No. 2023GZGJ64).

REFERENCES

[1] Meena S, Susank M, Guttula T, Chandana S H, Sheela J. Crop Yield Improvement with Weeds, Pest and Disease Detection[J]. 2023.

[2] Zhang J, Wang J, Zhao M. A Lightweight Crop Pest Detection Algorithm Based on Improved Yolov5s[J].Agronomy, 2023, 13(7).

[3] Wang X, Zhang S, Zhang T. Crop insect pest detection based on dilated multi-scale attention U-Net[J].Plant Methods, 2024, 20(1).

[4] Ali M A, Sharma A K, Dhanaraj R K. Heterogeneous features and deep learning networks fusion-based pest detection, prevention and controlling system using IoT and pest sound analytics in a vast agriculture system[J].Computers and Electrical Engineering, 2024, 116.

[5] Jiao L, Xie C, Chen P, Du J, Li R, Zhang J. Adaptive feature fusion pyramid network for multi-classes agricultural pest detection[J].Computers and Electronics in Agriculture, 2022, 195:106827

[6] Butera L , Ferrante A , Jermini M , Prevostini M, Alippi C. Precise Agriculture: Effective Deep Learning Strategies to Detect Pest Insects[J].IEEE/ CAA Journal of Automatica Sinica, 2022(2):9.

[7] Li W Y, Li M, Chen M X, Qian J P, Sun C H, Du S F. A machine vision-based feature extraction and classification method for crop multi-gesture pests[J]. Journal of Agricultural Engineering,2014,30(14):154-162.

[8] Han R Z, He Y. Remote automatic identification system of field pests based on computer vision[J]. Journal of Agricultural Engineering,2013,29(03):156-162.

[9] Liang Y, Qiu R Z, Li Z P, Chen S X, Zhang Z, Zhao J. Identification of major rice pests based on YOLO v5 and multi-source datasets[J]. Journal of Agricultural Machinery,2022,53(07):250-258.

[10] Sanghavi V B , Harshad B , Vijay D. Hunger games search based deep convolutional neural network for crop pest identification and classification with transfer learning[J].Evolving Systems, 2023, 14(4):649-671.

[11] Han D, Yoo D, Kim T. Analysis of social welfare impact of crop pest and disease damages due to climate change: a case study of dried red peppers[J]. Humanities and Social Sciences Communications, 2023, 10:1-13.

[12] Chen J, Chen L Y, Wang S S, Zhao H Y, Wen C J. Image recognition of garden pests based on improved residual network[J]. Journal of Agricultural Machinery,2019,50(05):187-195.

[13] Cheng X, Wu Y Z, Zhang Y H, Le Y. Image recognition of grain storage pests based on deep convolutional neural network[J]. Chinese Agronomy Bulletin,2018,34(01):154-158.

[14] Renault D, Elfiky A, Mohamed A. Predicting the insecticide-driven mutations in a crop pest insect: Evidence for multiple polymorphisms of the acetylcholinesterase gene with potential relevance for resistance to chemicals[J].Environmental Science and Pollution Research, 2022:1-19.

[15] Lü J, Nanda S , Shi-Min C , Yang M, Kang H E, Bao L Q. A survey on the off-target effects of insecticidal double-stranded RNA targeting the Hvβ'COPI gene in the crop pest Henosepilachna vigintioctopunctata through RNA-seq[J]. Journal of Agricultural Science:English Edition, 2022, 21(9):2665-2674.

[16] Yang Z, Jiang X, Jin G B J. MFSPest: a multi-scale feature selection network for light-trapped agricultural pest detection[J].Journal of Intelligent & Fuzzy Systems: applications in Engineering and Technology, 2023, 45(4):6707-6720.

[17] Zhu G, Xu Y, Sun Y Y, Zhang L. Pest identification and pesticide spraying system based on YOLO v3 under Darknet deep learning framework[J]. Agriculture and Technology, 2023, 43(10):33-38.

[18] Xiang Q, Huang X, Huang J T X. Yolo-Pest: An Insect Pest Object Detection Algorithm via CAC3 Module[J].Sensors, 2023, 23(6).

[19] Li Y, Huang P, Li H M. Research on part identification and localisation method of Jacquard machine based on OpenCV[J]. Dynamical Systems and Control, 2023, 12(3):157-164.

[20] Cui Y, Jiang X, Dai Y. Self-Labeling Learning Ensemble via Deep Recurrent Neural Network and Self-Representation for Speech Emotion Recognition[J].International Journal of Pattern Recognition and Artificial Intelligence, 2024, 38(09).

[21] Ahuja U, Singh S , Kumar M , Kumar K, Sachdeva M. COVID-19: Social distancing monitoring using faster-RCNN and YOLOv3 algorithms[J].Multimedia Tools and Applications, 2023, 82(5):7553-7566.

[22] Pathak D, Raju U S N. Shuffled-Xception-DarkNet-53: A content-based image retrieval model based on deep learning algorithm[J].Computers and Electrical Engineering, 2023.

[23] Sharma S, Goyal D.Enhanced security using video summarization for surveillance system using deep LSTM model with K-means clustering technique[J]. .Journal of Discrete Mathematical Sciences and Cryptography, 2023, 26(3):913-925.

[24] Cheng Q, Li J, Gao X L, Tang P R, Sheng L R, Wang W. A deep neural network lightweighting method based on deep sparse low-rank decomposition[J]. Control and Decision Making, 2023, 38(3):751-758.

[25] Ji S Y, Wang Y S, Zhai Y C. Improved YOLO V3-based water column signal detection algorithm at sea impact point[J]. Tactical Missile Technology, 2023(2):144-152.

[26] Tong Y X, Jin C, Li C. Wind power combination prediction model based on CEEMDAN-ISSA-BiLSTM 0 Introduction In order to build a new type of power system mainly based on new energy, wind power generation [J]. Electrical Engineering and Electricity, 2023(11):26-32.

[27] Xu C, Yang X, Lei P X. The improved deep plug-and-play super-resolution with residual-in-residual dense block for arbitrary blur kernels[J]. Pattern analysis and applications: paa, 2023, 26(4):1657-1670.