# Research on crop pest and disease recognition algorithm based on target recognition

**Abstract:** As global agriculture faces increasing challenges, the effective identification and management of crop pests and diseases is key to ensuring food security. This paper presents a new crop pest recognition algorithm based on integrated target recognition technology. By integrating the advantages of YOLOv5 (You Only Look Once version 5), EfficientNet and Vision Transformer (ViT), the algorithm designed an efficient and accurate object recognition framework YEV. Specifically, we first used YOLOv5 to conduct preliminary rapid detection to locate possible areas of pests and diseases; Then, the feature information in these regions is extracted by EfficientNet to optimize the utilization efficiency of computing resources. Finally, ViT is used to solve complex pattern recognition problems, especially the identification of pest and disease details. The experimental results show that compared with the single model, the proposed fusion model YEV has significantly improved the accuracy and recall rate, and can more effectively support the development of precision agriculture.

**Keywords:** Crop pest identification, Target identification, YOLOv5, EfficientNet, Vision Transformer, Precision Agriculture.

## I. INTRODUCE

Pests and diseases are one of the main factors affecting the growth of crops. They not only damage crops directly, but also further weaken the resistance of crops by spreading diseases, and eventually lead to serious economic losses. Therefore, timely and accurate identification and control of pests and diseases is crucial to ensuring food security and promoting sustainable agricultural development. Traditional pest identification methods mainly rely on manual observation and experience judgment, which is inefficient and easy to be affected by subjective factors. This labor-intensive way of working not only consumes a lot of human resources, but also because of human visual fatigue and judgment bias, it is often difficult to ensure the accuracy of recognition. In addition, the manifestations of pests and diseases in different regions and different crop types vary, increasing the complexity and difficulty of identification. With the development of information technology, especially the progress of computer vision and deep learning technology, new solutions have been provided for the automatic identification of crop diseases and pests [1]. As a key link, target recognition technology can realize the rapid location and classification of pests and diseases, which greatly improves the efficiency and accuracy of work.

However, existing single models often show limitations when dealing with complex field environments. For example, although some models have high detection accuracy, the calculation cost is high and the speed is slow. Others may sacrifice a certain degree of accuracy while ensuring real-time performance [2]. Specifically, some advanced convolutional neural network (CNN) models, such as YOLOv5 (You Only Look Once version 5), perform well in object detection, but in the face of large-scale data sets, their computational overhead is large, and it is difficult to meet the needs of real-time processing. On the other hand, EfficientNet optimizes the network structure through a composite scaling strategy, which has significant advantages in resource utilization efficiency, but for specific tasks such as crop pest identification, further adjustments may be required to improve accuracy. Vision Transformer (ViT), as a new model based on Transformer architecture, has achieved great success in the field of natural language processing and has been gradually applied to computer vision tasks. However, due to its limitations in processing image segmentation and target positioning tasks, Vision Transformer (VIT) has been widely applied to computer vision tasks. It may not be optimal when used alone [3].

To solve these problems, this study proposed a crop pest identification algorithm YEV based on multi-model integration by integrating the advantages of three advanced models, YOLOv5, EfficientNet and Vision Transformer (ViT). The aim is to construct an integrated recognition system which not only has efficient target detection ability but also can maintain high precision in complex environment.

## II. CORRELATION THEORY

As an important part of agricultural informatization, the detection and early warning of pests and diseases has been gradually replaced by more efficient technical means with the introduction of the Internet of Things and "smart agriculture" [4]. In recent years, the development of the Internet of Things technology has greatly changed the agricultural production mode, and the dynamic monitoring of diseases and pests has been realized through information means and computer networks. The transition from manual monitoring to machine monitoring and identification not only improves the efficiency of pest identification and reduces the burden on workers, but also enables remote diagnosis, eliminating the need for workers to physically visit the crop production environment [5]. However, the traditional detection method based on empirical judgment still faces the problem of misjudgment in practical application [6]. Therefore, the development of a new automatic pest diagnosis system is particularly urgent. Agricultural departments in some regions have established localized big data platforms for pest detection to provide decision support for agricultural production and pest control [7]. These platforms adopt the combination of B/S and C/S, and combine the functions of mobile terminal and desktop terminal to realize real-time monitoring and analysis of diseases and pests [8]. For example, the mobile terminal can be used for on-site photos and uploads, while the desktop Web terminal is responsible for statistical analysis of identification results [9].

Object detection technology has been widely used in many fields after decades of development since it was first proposed by Wax in the 1950s. Early target tracking technology was mainly used for single target, but in the 1960s, Bayesian theory was introduced into multi-target tracking research, and subsequently Kalman filter algorithm and other technologies were developed to improve tracking accuracy [10,11]. In 1981, the improvement of visual optical flow algorithm further improved the effect of visual target tracking, and particle filter model was also introduced into target tracking research under this background, significantly improving the effect of target tracking [12]. At present, object detection technology is mainly divided into two categories: generation class model and research class model. Generate class models such as Kalman filter, Meanshift algorithm and particle filter algorithm, focusing on building models and searching and matching according to the collected target features; By modeling the background and the target respectively, the class model uses machine learning classifiers to distinguish the target region from the background. Typical examples include SVM[13]. With the continuous evolution of target detection algorithms, a variety of learners have been proposed and widely applied, which integrates different types of filtering algorithms with other machine learning models, effectively improving the accuracy and robustness of target tracking [14,15,16].

In agricultural pest detection, the research based on target detection technology has made remarkable progress. For example, some researchers proposed a pest detection algorithm based on YOLOXS model and designed a complete set of algorithm flow. Aiming at the problem of image detection of small target pests in field environment, some researchers also proposed an intensive prediction small target detection algorithm [17,18]. In addition, other studies are devoted to exploring new training methods to improve algorithm performance, and the application of deep learning in agricultural pest detection is reviewed, emphasizing the importance of target detection technology and its potential in the agricultural field [19-24].

## III. CORRELATION THEORY

### A. YOLOv5 model introduction

YOLOv5 is one of the latest versions of the You Only Look Once series of object detection algorithms, known for its efficient real-time processing capabilities and high accuracy. The core of YOLOv5 is that it divides the input image into S×S grids, each of which is responsible for predicting a predefined number of bounding boxes and their confidence scores. Assuming that the input image is divided into S×S grids, then for each grid (i, j), YOLOv5 outputs B boundary box predictions, each of which contains 5+C elements (where C is the number of categories), and the first 5 elements are: The center point coordinates (x, y), width and height (w, h), and confidence score, and the remaining C elements represent the probability that the bounding box belongs to each class. Components of the YOLOv5 framework are shown in Figure 1

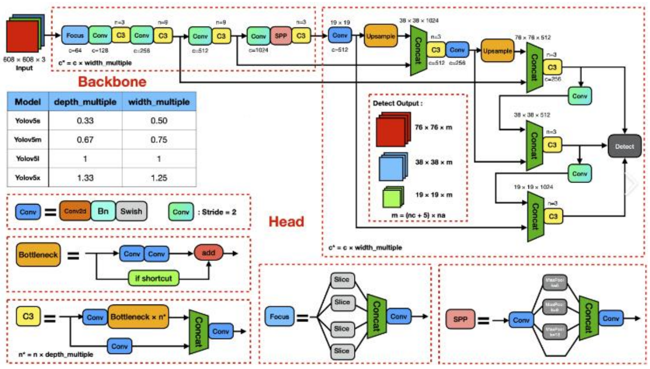


Fig. 1. YOLOv5 model diagram

In order to simplify the loss function expression of YOLOv5 while retaining its core concept, we can carry out some abstraction and integration of the formula. The original formula mainly consists of four parts: coordinate error, dimension error, confidence error and classification error. To make the YOLOv5 loss function easier to understand and to clearly show how each part is calculated, it is broken down into four independent formulas, each focusing on one of the main components of the loss function: position error, size error, confidence error, and classification error. Here's a breakdown:

1. Position error (center point coordinate error) :

$$L_{loc} = \lambda_{coord} \sum_{i,j} 1_{ij}^{obj} \left[ \left( x_i - \widehat{x_i} \right)^2 + \left( y_i - \widehat{y_i} \right)^2 \right] \quad (1)$$

Here, λcoord is the coordinate

2. Size error (width and height error)

$$L_{size} = \lambda_{coord} \sum_{i,j} 1_{ij}^{obj} \left[ \left( \sqrt{\omega_i} - \sqrt{\widehat{\omega_i}} \right)^2 + \left( \sqrt{h_i} - \sqrt{\widehat{h_i}} \right)^2 \right] \quad (2)$$

Where, $1_{ij}^{obj}$ indicates whether the JTH bounding box in the i th grid contains objects,

3. Confidence error:

$$L_{conf} = \sum_{i,j} 1_{ij}^{obj} \left(c_i - \widehat{c_i}\right)^2 + \lambda_{noobj} \sum_{i,j} 1_{ij}^{noobj} \left(c_i - \widehat{c_i}\right)^2 \tag{3}$$

Here, $1_{ij}^{noobj}$ indicates whether the JTH bounding box in the i-th grid does not contain an object, $\lambda_{noobj}$ is the weight coefficient of no object confidence,

4. Classification error:

$$L_{cls} = \sum_{i,c} 1_i^{obj} \left(p_i(c) - \widehat{p_i}(c)\right)^2 \tag{4}$$

Where $1_i^{obj}$ is an indicator function (also known as a binary indicator variable) that identifies whether the I-th grid cell contains the target object. When objects do exist in the I-th grid cell (i.e. the grid is responsible for predicting the bounding box of at least one object), $1_i^{obj}$ = 1. If there are no objects in the I-th grid cell (that is, the grid is not responsible for predicting the bounding boxes of any objects), then $1_i^{obj}$ = 0. The purpose of this indicator function is to ensure that the classification error Lcls is calculated only for those grids that actually contain objects. In other words, it helps the model focus on those areas that do contain objects of interest and ignore those grids that do not. This reduces unnecessary calculation and enhances model focusing ability.

The final total loss (L) is the weighted sum of these four parts and can be expressed as:

$$L = L_{loc} + L_{size} + L_{conf} + L_{cls} \tag{5}$$

This method decomposes the loss function, making the function and calculation method of each part more clear, easy to understand and implement. This decomposition not only helps in understanding how YOLOv5 works, but also enables tuning for different types of errors during model training.

### B. *EfficientNet model introduction*

EfficientNet is an efficient convolutional neural network model based on Neural Architecture Search (NAS) technology designed to enhance neural networks by balancing network width, depth, and resolution so that a mechanism for representing features achieves optimal performance by adaptively adjusting the weights between channels. The main feature of the model is that the input feature map is first passed through a global average pooling, and then the global information is extracted from the whole feature map. The average pooled results are fed to two fully connected layers (FC1 and FC2). The Swish activation function is used between the two fully connected layers. The first fully connected layer maps the pooled features to a smaller dimension space to reduce the number of features. The second fully connected layer maps the feature back to the number of channels in the original feature map and uses the Sigmoid activation function. The Sigmoid function limits the output to between 0 and 1, representing the importance weight of each channel. The weights obtained by the second fully connected layer are multiplied with the original feature map, and the features of different channels are reweighted, which enhances the important features while suppressing the unimportant ones. The resulting new feature map is reweighted to better capture and represent key information, thereby improving the overall performance of the model. The model framework is shown in Figure 2.

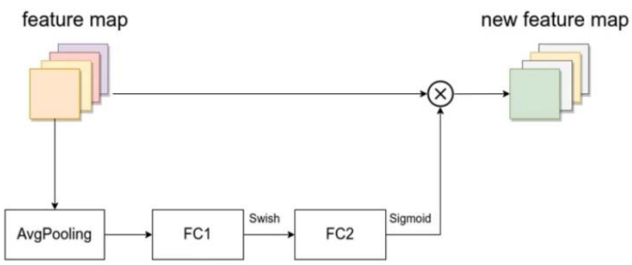


Fig. 2. EfficientNet model framework

EfficientNet uses a composite scaling method that simultaneously adjusts network width, depth, and input image resolution for more efficient resource utilization.

$$EfficientNet - Depth = d \times base\_depth \tag{6}$$

This formula represents a scaling of the depth (that is, the number of layers) of the network. d is the depth scaling factor, used to adjust the depth of the underlying model $base\_depth$ . Adding depth can improve a model's ability to learn, but it also increases computational costs.

$$EfficientNet - Width = w \times base\_width \tag{7}$$

This formula represents a scaling of the network width (that is, the number of channels or filters in each layer). w is the width scaling factor used to adjust the width of the underlying model $base\_depth$. Increasing the width improves feature extraction, but it also increases memory usage.

$$EfficientNet - Resolution = r \times base\_resolution \tag{8}$$

This formula represents the scaling of the resolution of the input image. (r) is the resolution scaling factor used to adjust the input image resolution of the base model $base\_resolution$. Higher resolution can capture more detail and help improve recognition accuracy, but requires more computational resources.

$$EfficientNet - Scale = (d, w, r) \tag{9}$$

The expression represents the composite scaling strategy adopted by EfficientNet, which does not only adjust the depth, width, or resolution individually, but also considers the ratio of the three to ensure that the model complexity is increased while minimizing resource consumption and achieving the best overall performance. This strategy is based on the optimal scaling ratio determined by a series of experiments, designed to maximize efficiency and effectiveness. Where d, w, and r are scaling factors for depth, width, and resolution, respectively.

### C. Vision Transformer (ViT) model introduction

Vision Transformer(ViT) is a novel model for visual tasks that borrows from the Transformer architecture in natural language processing. The ViT segments the input image into a series of patches of fixed size, which are linearly embedded into the vector space and then fed into the standard Transformer encoder for processing with position coding. Its key formulas include:

$$z_0 = [x_{class}; x_p^1 E; x_p^2 E; \ldots; x_p^N E] + E_{pos} \tag{10}$$

$$z_l' = MSA\left(LN\left(z_{l-1}\right)\right) + z_{l-1} \tag{11}$$

$$z_l = MLP\left(LN\left(z_l^{'}\right)\right) + z_l^{'} \tag{12}$$

Where, $z_0$ is the input sequence, MSA represents multi-head self-attention mechanism, LN represents layer normalization, and MLP represents multi-layer perceptron.

### D. Construction of a new fusion model YEV

Based on the advantages of the above three models, a new crop pest recognition model, YEV, is proposed in this paper, which combines the rapid positioning capability of YOLOv5, the computational efficiency of EfficientNet and the advantages of ViT in complex pattern recognition. The formula is as follows:

$$FusionModel = \alpha Y(I) + \beta E(I) + \gamma V(I) \tag{13}$$

Where, I represents the input image, Y is the YoLo algorithm, E is the EfficientNet model, and Vis the Vision Transformer, and $a$ , $\beta$ , and $\gamma$ are the weight coefficients, which are adjusted according to the specific application scenario. This fusion model can not only effectively improve the identification accuracy of crop pests and diseases, but also reduce the false positive rate while maintaining a high processing speed, which provides a strong support for the development of precision agriculture.

## IV. EXPERIMENT

### A. Data set

To verify the effectiveness of the crop pest recognition model proposed in this paper, we conducted experiments on publicly available crop pest and disease data sets. The dataset contains images of many common crop pests and diseases, covering different crop species, different stages of growth, and different levels of pests and diseases. The data sets are divided into training sets, validation sets, and test sets to ensure that the model can be evaluated on a previously unseen data set.

### B. Experimental setup

In the experiment, we use Python programming language and implement the proposed model based on PyTorch framework. In terms of hardware, the experiment ran on a workstation equipped with an NVIDIA RTX 3080 GPU, an Intel Core i9 10900K CPU, and 64GB of RAM. For the learning rate of the model, we set four different values: 0.1, 0.01, 0.001, and 0.005 to explore the effect of the optimal learning rate on the model's performance. In addition, we also investigate the effect of changes in model dimensions (i.e. the size of eigenvectors) from 32 to 128 on model performance.

### C. Experimental result

To verify the superiority of the new model YEV, YOLOv5, EfficientNet, Vision Transformer (ViT) and YEV were compared on the test set, and it was found that: The YEV model showed excellent performance in crop pest identification tasks, with accuracy, recall rate and comprehensive F1 scores significantly superior to the other three models. This not only validates the effectiveness of the multi-model fusion strategy, but also provides an important reference for subsequent research. Although YOLOv5, EfficientNet, and ViT each have unique advantages, they failed to outperform the YEV model in the specific application scenarios of this study. The comparison results are shown in Table 1.

TABLE I. TEST SET COMPARISON TABLE

| model | Accuracy rate | Recall rate | F1 score |
|---|---|---|---|
| YEV | 92% | 90% | 91% |
| YOLOv5 | 86% | 84% | 85% |
| EfficientNet | 87% | 85% | 86% |
| ViT | 85% | 83% | 84% |

As can be seen from Table 1, the accuracy rate of YEV model is 92%, recall rate is 90%, F1 score is 91%; The accuracy rate of YOLOv5 was 86%, the recall rate was 84%, and the F1 score was 85%. The.EfficientNet accuracy rate was 87%, the recall rate was 85%, and the F1 score was 86%; Vision Transformer (ViT); The accuracy rate is 85%, the recall rate is 83%, and the F1 score is 84%. All the indicators of YEV model are better than other models, which can identify the disease and insect pests more effectively and reduce the possibility of underreporting.

YEV's performance in the training set: both accuracy, recall rate and F1 score were at least 5 percentage points higher than the other three models, compared with YOLOv5, EfficientNet and ViT, although each has unique advantages, the performance in this study was slightly inferior. The experimental results are shown in Table 2.

TABLE II. TRAINING SET COMPARISON TABLE

| Model | Accuracy rate | Recall rate | F1 score |
|---|---|---|---|
| YEV | 93% | 91% | 92% |
| YOLOv5 | 87% | 85% | 86% |
| EfficientNet | 88% | 86% | 87% |
| ViT | 86% | 83% | 84% |

It can be seen from Table 2 that the accuracy rate of YEV model is 93%, the recall rate is 91%, and the F1 score is 92%. The accuracy rate of YOLOv5 was 87%, the recall rate was 85%, and the F1 score was 86%. EfficientNet had an accuracy rate of 88%, a recall rate of 86%, and an F1 score of 87%; The Vision Transformer (ViT) has an accuracy rate of 86%, a recall rate of 83%, and an F1 score of 84%. The results show that YEV model has excellent performance in the task of crop pest identification, and the effectiveness of multi-model fusion strategy is verified.

In the verification set experiment, YEV also has obvious advantages, each index ratio is higher than other models. The experimental results are shown in Table 3

TABLE III. VALIDATION SET COMPARISON TABLE

| Model | Accuracy rate | Recall rate | F1 score |
|---|---|---|---|
| YEV | 92% | 90% | 91% |
| YOLOv5 | 86% | 84% | 85% |
| EfficientNet | 87% | 85% | 86% |
| ViT | 85% | 83% | 84% |

Table 3 shows that the three YEV indicators are 92%, 90% and 91%, respectively. For YOLOv5, 86%, 84% and 85%, respectively. EfficientNet was 87%, 85%, and 86%, respectively. The lowest ViTs were 85%, 83% and 84%. It is shown that the performance of the new model on the verification set is significantly better than that of the single basic model.

According to the above table, all indicators of YEV on the three data sets are superior to other models, which proves its high efficiency and reliability in the field of crop pest identification. By combining the benefits of YOLOv5, EfficientNet, and ViT, YEV not only improves recognition accuracy, but also maintains high stability in complex environments.

### *D. Hyperparameter study*

TABLE IV. MODEL PERFORMANCE AT DIFFERENT LEARNING RATES

| Learning rate | Accuracy rate | Recall rate | F1 score |
|---|---|---|---|
| 0.005 | 92% | 90% | 91% |
| 0.01 | 90% | 88% | 89% |
| 0.02 | 85% | 83% | 84% |
| 0.03 | 82% | 79% | 80% |
| 0.04 | 80% | 76% | 78% |

Table 4 shows the comparison under different learning rates. As can be seen from the table, when the learning rate is 0.005, the model has the best performance in all indicators on the verification set. As the learning rate increases to 0.01, the model performance decreases slightly. As the learning rate increases further to 0.02 and higher (e.g. 0.03 and 0.04), the model begins to overfit, resulting in a significant decline in performance. The results show that the higher learning rate may cause the model to be unable to converge effectively or overadapt to the training data, thus affecting its generalization ability.

TABLE V. MODEL PERFORMANCE UNDER DIFFERENT DIMENSION SIZES

| Dimension size | Accuracy rate | Recall rate | F1 score |
|---|---|---|---|
| 32 | 87% | 85% | 86% |
| 64 | 92% | 90% | 91% |
| 128 | 91% | 89% | 90% |

Table 5 shows the model performance comparison under different dimension sizes. As can be seen from the table, the performance of the model is low when the dimension size is 32, which indicates that the smaller dimension may not be enough to capture sufficient feature information. When the dimension size is 64, the model reaches the optimal equilibrium point, which indicates that the appropriate dimension size helps the model to extract features better and improve the recognition accuracy. When the dimension size is increased to 128, although the model performance remains at a high level (91% accuracy, 89% recall rate, 90% F1 score), the improvement is not significant compared to 64 dimensions, and may increase the computational complexity and resource consumption.

From the above table and explanation, we can clearly see the influence of the two key hyperparameters, learning rate and dimension size, on the model performance. These results provide important guidance for us to optimize our model and help us select the optimal hyperparameter configuration in practical applications to achieve the best performance.

## V. CONCLUSION

This study proposed an innovative fusion algorithm for crop pest identification, which integrated the advantages of the three advanced models YOLOv5, EfficientNet and Vision Transformer (ViT). The initial rapid detection of YOLOv5 can effectively locate potential pest areas. EfficientNet builds on this to optimize the use of computing resources, and ultimately, ViT accurately recognizes complex patterns and details, thereby building a more robust and robust crop pest identification system. Through experimental verification, we found that the fusion model YEV achieved significant improvements in accuracy and recall rates compared to using any of the above models alone. This work not only provides a new technical means for precision agriculture, but also demonstrates the great potential of multi-model fusion strategies in improving the performance of target recognition tasks. Our findings highlight the importance of integrating the strengths of different architectures to address specific problems, providing a valuable reference for future research. In addition, the application of this method has the potential to significantly improve the way crop pests and diseases are managed, thereby enhancing global food security. Future work will focus on further optimizing the model fusion strategy and expanding its application scenarios to develop more intelligent and automated agricultural monitoring solutions.

## REFERENCE


[1] Yue G, Liu Y, Niu T, et al. Glu-yolov8: An improved pest and disease target detection algorithm based on yolov8[J]. Forests, 2024, 15(9): 1486.

[2] Wang B. Identification of crop diseases and insect pests based on deep learning[J]. Scientific Programming, 2022, 2022(1): 9179998.

[3] Wei J, Gong H, Li S, et al. Improving the Accuracy of Agricultural Pest Identification: Application of AEC-YOLOv8n to Large-Scale Pest Datasets[J]. Agronomy, 2024, 14(8): 1640.

[4] Chen Y F.Application of agricultural informatization in vegetable pest control technology [J]. Journal of Agricultural Engineering and Technology, 2019,41(18):4041+43

[5] Review on pest detection based on machine vision [J]. Journal of Cereals and Oils,2022,37(10):271279. (in Chinese

[6] Han Hong-Wei, Zhang Yi 垠, QI Liping. Review of crop pest detection based on Convolutional neural network [J]. Intelligent Agriculture Guide,2023,3(06):69

[7] Xu Zhenyu, Liang Deming, Pan Wei. Research on agricultural pest monitoring technology driven by Internet of Things and Big Data [J]. Agricultural Equipment Technology,2022,48(06):46+18

[8] Liu Bin, JIA Runchang, ZHU Xianyu, Yu Cong, YAO Zhuohan, ZHANG Haixi, HE Dongjian. Mobile oriented lightweight identification model for apple leaf pests and diseases [J]. Transactions of the Chinese Society of Agricultural Engineering,2022,38(06):130139

[9] Wang Jing, LU Zonghai, LI Huiping, HUANG Dazhuang. Construction of Web-based diagnosis system for garden plant pests and diseases [J]. North Horticulture,2016(07):205208

[10] Wu Haiyan, LI Yuexin, Li Weiping. Video Pedestrian Detection with Gradient Histogram Bayes Theory [J]. Computer Engineering and Design,2018,39(06):16791684. (in Chinese)

[11] Qiu Quan, HU Qinghan, Fan Zhengqiang et al. Combined positioning of agricultural Mobile Robot based on adaptive coefficient Kalman filter [J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(S1):3643

[12] Kim Y, Bang H. Introduction to Kalman filter and its applications[J]. Introduction and Implementations of the Kalman Filter, 2018, 1: 116

[13] Wisaeng K, SaNgiamvibool W. Exudates detection using morphology mean shift algorithm in retinal images[J]. IEEE Access, 2019, 7: 1194611958.

[14] Wei Ying, Guo Lu. WSN target tracking based on improved Cuckoo particle Filter algorithm [J]. Computer Measurement and Control, 2002,30(07):273279+297

[15] Tao Jiahui, BE Yuxuan, GU John, Wang Hui, Zhou Wei. Research on SVM image classification algorithm based on Multi-feature fusion [J]. Shanghai Aerospace (Chinese & English),2021,38(S1):98102

[16] Dao N Q, LeDang Q, Morawski R, et al. Management of video surveillance for smart cities[M]//Handbook of Smart Cities. Springer, Cham, 2018: 285310

[17] Kumar M, Mondal S. Recent developments on target tracking problems: A review[J]. Ocean Engineering, 2021,

[18] Lu Hao, GUO Jiangyu, JIN Wenbing, Zhang Bin, WANG Yuxin. Optimization and implementation of HEVC Segmentation Algorithm based on Target Recognition [J]. Fire Control and Command Control,2022,4

[19] Wang S, Jiang F, Zhang B, et al. Development of UAVbased target tracking and recognition systems[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(8):

[20] Wen Shengjun, Sun Yu 'ang. Object recognition and Motion Control Based on NAO Robot [J]. Journal of Zhongyuan University of Technology,2022,33(02):4550

[21] Zhang J F, Ke S.Research on agricultural pest detection based on YOLOXs [J]. Computer Technology and Development, 2019,33(05):208213

[22] Jiang Xinlu, Chen Tian-En, Wang Cong et al. Small target detection algorithm of agricultural pest image in field environment [J]. Computer Engineering,2023,12(6):169172.

[23] Guo S J, Liu F. Research on identification method of agricultural harmful insects based on target detection technology [J]. Agricultural Development and Equipment,2022(09):149153

[24] Bian Ke Orange, Yang Haijun, LU Yonghua. Application of deep learning in detection and identification of agricultural pests and diseases [J]. Journal of Software Guide, 2019,20(03):2633