

Article

MCD-Yolov5: Accurate, Real-Time Crop Disease and Pest Identification Approach Using UAVs

Lianpeng Li ^{1,*} , Hui Zhao ^{1,2} and Ning Liu ² ¹ School of Automation, Beijing Information Science & Technology University, Beijing 100192, China; zhhhui@bistu.edu.cn² Beijing Key Laboratory of High Dynamic Navigation Technology, Beijing Information Science & Technology University, Beijing 100192, China; ning.liu@bistu.edu.cn

* Correspondence: llp@bistu.edu.cn

Abstract: As the principal factor affecting global food production, accurate identification of agricultural pests and diseases is crucial in ensuring a sustainable food supply. However, existing methods lack sufficient performance in terms of accuracy and real-time detection of multiple pests and diseases. Accordingly, accurate, efficient, and real-time identification of a wide range of pests and diseases is challenging. To address this, we propose an MCD-Yolov5 with a fusion design that combines multi-layer feature fusion (MLFF), convolutional block attention module CBAM, and detection transformer (DETF). In this model, we optimize the MLFF design to realize the dynamic adjustment of feature weights of the input feature layer to (1) find an appropriate distribution of feature information proportion for the detection task, (2) enhance detection speed by efficiently extracting effective images and effective features through CBAM, and (3) improve feature extraction capability through DETF to compensate for the accuracy problem of multiple pest detection. In addition, we established an unmanned aerial vehicle system (UAV) for crop pest and disease detection to assist in detection and prevention. We validate the performance of the proposed method through an established UAV platform, and five indicators are employed to quantify the performance. MCD-Yolov5 can detect pests and diseases with a large improvement in detection accuracy and detection efficiency, obtaining an 88.12% accuracy. The proposed method and system provide an idea for the effective identification of pests and diseases.



Citation: Li, L.; Zhao, H.; Liu, N. MCD-Yolov5: Accurate, Real-Time Crop Disease and Pest Identification Approach Using UAVs. *Electronics* **2023**, *12*, 4365. <https://doi.org/10.3390/electronics12204365>

Academic Editor: Mahmut Reyhanoglu

Received: 6 September 2023

Revised: 9 October 2023

Accepted: 18 October 2023

Published: 20 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: crop disease and pest identification; Yolov5; unmanned aerial vehicle system (UAVs); deep learning

1. Introduction

Demand for sustainable food supplies is increasing as the world's population continues to grow. The threat of plant diseases to the world's food supply has increased significantly, directly affecting global economic health. Over the past few years, pests and diseases worldwide have led to significant declines in crop yields and quality, particularly in fruits, grains, and vegetables [1]. According to the Food and Agriculture Organization (FAO) [2], current international pests and diseases reduce potential crop yields by an average of 40%, while many farmers in developing countries suffer yield losses of up to 100%, and crop pests and diseases cause losses of about one-tenth of the world's total food production [3]. Crop pests and diseases have regional characteristics and are random, unstable, and growing in time, leading to considerable difficulties in diagnosing their occurrence.

However, traditional approaches have been characterized by such drawbacks as time-consuming, lagging diagnosis, and limited scope by chemical methods and expert systems [4]. Therefore, a low-cost, efficient, and accurate method is in substantial demand.

Information and artificial intelligence (AI) techniques in agriculture are prevalent worldwide [5]. Different detection methods have been analyzed in domestic and international research, covering migration learning [6], convolutional neural networks (CNN) [7],

generative adversarial networks (GAN) [8], and hyperspectral techniques [9]. These studies have achieved some results in realizing the effective identification and extraction of various types of crop information. Sharma et al. [10] analyzed pest and disease identification measures, such as image processing, and machine learning in a large range of complex backgrounds with substantially reduced performance. References [11,12] studied the key technologies involved in image acquisition, data processing, and result analysis for crop pest identification. They explained that existing studies have faced various problems, such as difficult feature extraction and low detection performance, which do not meet the application requirements. Therefore, research on rapid detection methods for a wide range of crops in complex backgrounds, in which existing research has difficulty ensuring real-time detection accuracy, is still in its infancy. In addition, the increase in climate change and intensification of global trade flows will bring additional issues [13,14].

Given the current issues, we provide a detailed overview of recent studies and conduct a comparative analysis in Section 2. Motivated by the drawbacks of current research, we focus on deep learning, propose an unmanned aerial vehicle system (UAV)-based crop pest and disease identification using the MCD-Yolov5, and provide a performance test on a UAV platform. In summary, the main contributions of this paper are as follows.

- For the problems of accuracy and real-time recognition of multiple crop pests and diseases, we integrate the MLFF, CBAM, and DETF modules and propose the MCD-Yolov5 method. The MLFF module employs an adaptive learning approach to dynamically adjust the weights of each input feature layer. This allows for the intelligent allocation of feature information to enhance the capabilities of the detection task and improve overall feature extraction performance. CBAM is used to improve detection speed. Lastly, DETF improves the extraction of effective information on UAV crop pests and diseases, enhances the accuracy of network training and detection;
- To address the issue of identifying crop pests and diseases in a wide range, we combine the proposed algorithm to build a UAV-based crop pest and disease detection system. We likewise provide details of the system's component units and important parameters, including a 5G module, power subsystem (PS), environment perception subsystem (EPS), and aircraft subsystem (AS), to enhance image acquisition speed through the maneuverability and binocular vision of a UAV;
- To verify the effectiveness of the proposed algorithm and system, we set up a physical experimental platform and joint data set, and conduct performance validation experiments. The experimental results showed that the proposed method performs well in detecting pests and diseases of nine crops with good accuracy and real-time performance.

The rest of this paper is organized as follows. Section 2 introduces the related work. Section 3 gives a detailed description of the MCD-Yolov5 model. Section 4 established a UAV detection system and conducted a verification and comparison experiment about the performance of five algorithms. Finally, conclusions and future work are summarized in Section 5.

2. Related Work

Computer vision and AI technology, specifically deep learning, have made significant advancements in crop disease identification research. This progress can be attributed to the rapid development of deep learning technology, which has been boosted by the availability of large datasets, improved computational power, and advanced algorithms. Using the powerful image classification and target recognition ability of deep learning, diseased crops, and damaged parts can be directly separated from the images. Moreover, deep learning networks are insensitive to external environmental conditions and can be applied to practical agricultural production activities.

As a deep learning technique, CNN is the most popular image recognition classifier and is rapidly becoming the method of choice because it has shown excellent capabilities in image processing and classification [15]. CNN's architecture includes convolution, ReLu, pooling layer, fully connected layer, and classification layer. At present, there are various

widely available CNN architectures, including AlexNet, VGGNet, Inception, ResNet, and Dense Net [16]. In the study by Chen et al. [17], deep CNNs were employed to recognize symptoms of four cucumber diseases: downy mildew, anthracnose, powdery mildew, and target leaf spot. The researchers achieved impressive outcomes, attaining a recognition accuracy of 93.4%. A variety of crop pest solutions were derived based on CNN.

To improve the accuracy of detection models, Abhilasa et al. [18] proposed a combination of hybrid activation functions to improve the accuracy of CNN models. The hybrid activation function is tested and trained on different data sets, and the results show that this function has higher accuracy than ReLU activation function. In [19], SiamRPN was used to identify mobile pests, and an attention mechanism was introduced with an accuracy of 94.2%.

To solve the problem of crop disease detection in large-scale cultivation, Hua [20] proposed a new algorithm called faster R-CNN for crop disease detection based on multi-feature decision fusion. The growth of cucumber seedlings was captured and analyzed using machine vision image acquisition, image processing, and analysis techniques. The results of the study showed that model minus can effectively distinguish crop diseases. Large-scale detection often suffers from the problem that agricultural pest images are often blurred. Jin [21] proposed GAN with quadruple attention and residual and dense fusion mechanisms to transform low-resolution pest images. The experimental results using the proposed GAN demonstrate an increase of 182.89% in recall and a significant improvement in classification accuracy after reconstruction. The aforementioned research discussed the idea of using images taken based on aircraft for identifying soybean leaf diseases, thereby introducing the concept of solving large-scale crop identification.

To improve detection speed, Sustersic [22] provides a deep learning model named SSD, which follows a one-stage approach and was trained to detect and identify 14 different crops as well as 26 crop diseases. The trained model exhibited excellent performance, achieving a remarkable accuracy of 92.35% when evaluated on the test dataset. The Yolo series network combines detection accuracy and detection speed with excellent performance in detecting crop pests and diseases [23]. Tian et al. [24] proposed a deep learning-based method for apple anthracnose damage detection using Densenet to optimize the feature layer of the low-resolution Yolov3 model, thereby significantly improving the utilization of neural network features and enhancing the detection results of the Yolov3 model. The experimental results showed that the model achieved 95.57% of mean average precision (mAP). Gao [25] developed a DCNN model that merged deep learning and Google data analysis. They used an improved Yolov4 network to classify and accurately recognize images based on their quality levels, achieving a 95% recognition accuracy. The study provides a scientific basis for detecting the imaging capability of sensors and objectively evaluating the image quality of crop pests and diseases. Reference [6] proposed a Yolo-based pest detection and identification diagnostic system with an accuracy of 93.84%. The performance of their method was comparable to that of human experts and traditional neural networks. They also used their model to identify two weeds with an accuracy of 98.92%. Yao et al. [26] introduced CDMA into Yolov5 for dynamic target tracking, successfully improving the detection accuracy and real-time performance of the algorithm. The lighter CNNs such as SqueezeNet and NanoDet make it difficult to balance real-time and detection accuracy for crop images taken in flight. The information comparison of existing algorithms are as shown in Table 1.

Table 1. Information comparison of existing algorithms.

Ref	Model	Datasets	Accuracy	Pros and Cons
[17]	DCNN	Self-built	93.40%	High complexity
[19]	SiamPRN	IDADP	94.2%	High accuracy
[20]	Faster R-CNN	AI Challenger	95.35%	High time cost
[22]	SSD	Plant-Village	92.35%	High speed
[25]	Yolov4	IoT datasets	95.00%	High speed

Given the in-depth research on the application of deep learning techniques in crop pest identification tasks, several scholars have identified the limitations of deep learning techniques. Mekhalfi et al. [27] concluded that the Yolov5 neck network uses a single-scale feature map, which limits the expression of feature information and constrains detection accuracy. Jubayer et al. [28] explained that there are many UAV aerial photography crop targets and pests. However, how to immediately extract the identification of crop pest expectations is still limited by the backbone network structure. Hence, the network should be optimized and improved. At present, the main issues are as follows:

- Rapid extraction and detection of a large range of crops is difficult. Given that there are various types of crop pests and diseases, and that the pathological image features of different farming crops are different, improving the extraction ability and recognition accuracy of multiple crop pest and disease features should deal with the primary problem;
- Existing methods indicate that the performance is far from the capability of CNN detection. The current challenge is to achieve rapid and accurate detection of crop pests and diseases.

At present, UAV-based deep-learning detection schemes are the most effective methods for large-scale disease identification. Yolov5 as a lightweight CNN for pest and disease identification has some advantages in detection speed. Therefore, this paper proposes a UAS crop pest and disease identification using MCD-Yolov5, thereby providing a new method to solve the problems of accurate and rapid identification of plant diseases.

3. MCD-Yolov5 Model

For the UAV aerial photography crop pest target detection task, a network of certain depth is required to extract feature information owing to the complex background and small targets. We chose the Yolov5 algorithm with specification L as the baseline method, which consists of three parts: backbone, neck, and head networks.

This section optimized the established Yolov5 by introducing MLFF and CBAM to improve the feature extraction capability and provided DETF to enhance the real-time detection requirements of the backbone network, as shown in Figure 1.

3.1. MLFF

The Yolov5 feature extraction network utilizes both shallow and deep layers to extract different types of information. The shallow layers capture detailed texture features, while the deep layers focus on extracting broader semantic information. Each layer of feature maps plays a unique role in pest and disease identification. To enhance the feature extraction capability, we combine the information from different layers to incorporate contextual understanding. Additionally, we dynamically adjust the weights of each feature layer through adaptive learning, effectively allocating the appropriate proportion of feature information for the detection task.

The detection system can successfully detect instances of text similarity by utilizing various techniques, including but not limited to feature map multi-scale fusion, which involves generating a new feature map through the combination of feature maps from multiple layers. Initially, the fused feature maps of the four layers P_1, P_2, P_3, P_4 should be normalized to obtain a consistent size and number of channels. The adjusted feature maps are integrated into the channel dimension to obtain a fused feature map P ,

$$P = Fu([P_1, P_2, P_3, P_4]) \quad (1)$$

Channel attention of the feature map is calculated by Squeeze and Excitation functions to alleviate the degradation of feature capability expression caused by information

redundancy and reduce non-relevant information interference. Global average pooling is performed based on input $P \in \mathbb{R}^{H \times W \times C}$ to obtain the global features for each channel:

$$Z_c = F_s(P_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j), \quad (2)$$

where F_s is the excitation map and Z_c denotes the global feature space of the c th channel of the feature map P .

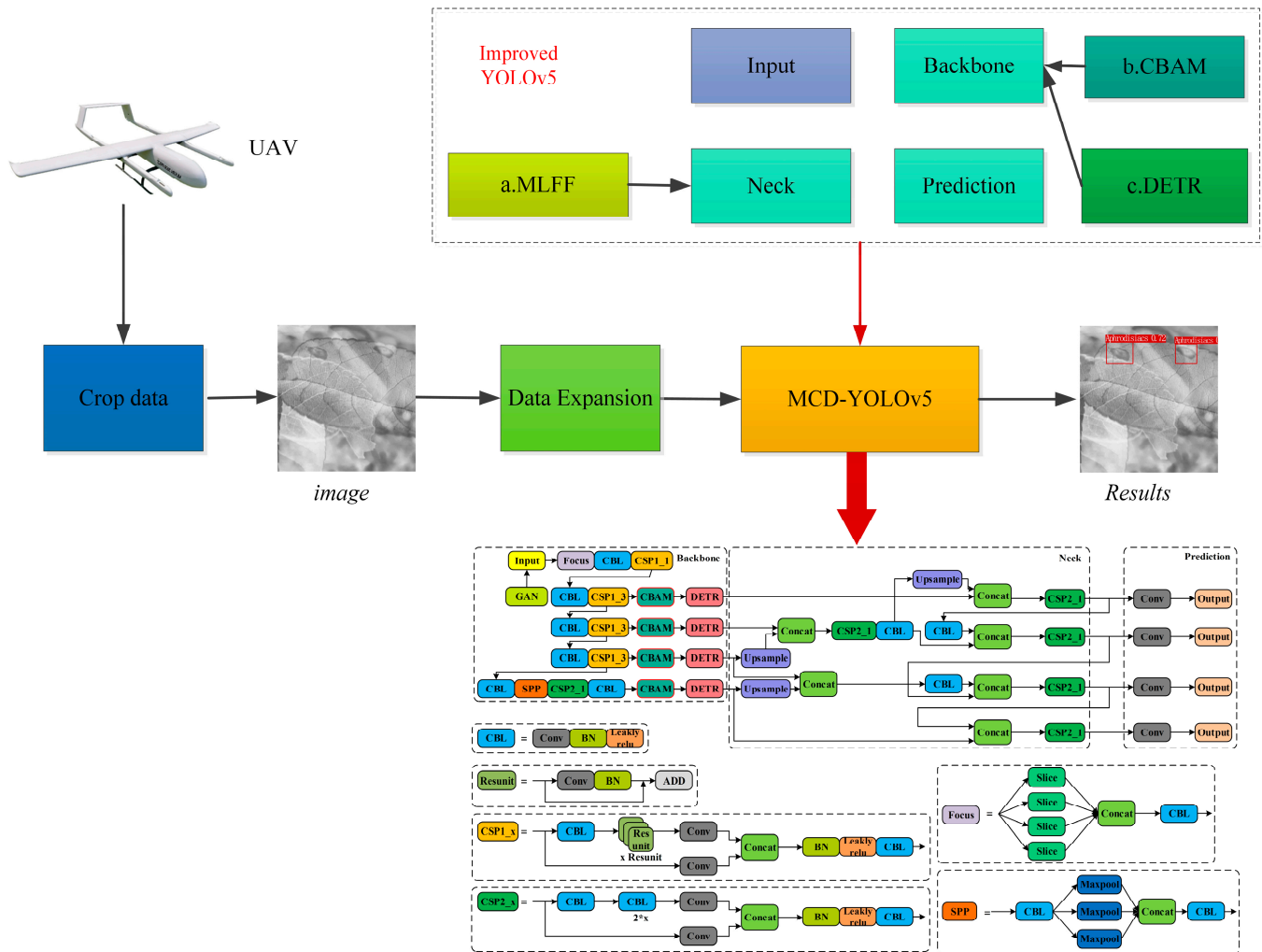


Figure 1. The structure of MCD-Yolov5.

We fit the nonlinear relationship between the channels through two fully connected layers and use the Sigmoid activation function to generate the corresponding weights s . The intermediate feature f is as follows:

$$f = F_e(x, W) = \sigma(W_2 \delta[W_1, z])), \quad (3)$$

where F_e represents the excitation mapping, σ denotes the Sigmoid function, δ means the nonlinear activation function ReLU6, and $[\cdot, \cdot]$ is the rotational splicing transformation. Weight s is multiplied with the feature map P to obtain the feature map U .

First, f is split into two independent features along the middle in the horizontal and vertical directions. Second, two convolutional layers with a convolution kernel of 1 and a nonlinear function activation layer are entered to obtain the weights in the two directions.

Lastly, the Softmax function is used to normalize the weights of the feature map in space to obtain the weight matrix.

$$W \in \mathbb{R}^{4 \times H \times W} \quad (4)$$

Weights are expanded into three-dimensional arrays $\alpha, \beta, \gamma, \lambda \in \mathbb{R}^{1 \times H \times W}$ with the same dimensionality as the input features along their respective directions, to obtain the important weight parameters of each layer of the feature map and complete the rescaling of the features to obtain the new feature map after fusion L_3 .

$$L_3 = \alpha P_1 + \beta P_2 + \gamma P_3 + \lambda P_4 \quad (5)$$

As illustrated in Figure 2, MLFF fuses the input feature maps, $\{P_2, P_3, P_4, P_5\}$ aggregates more contextual information from feature layers with different perceptual fields, and obtains a weighted fusion of feature maps at each layer. $\{L_2, L_3, L_4, L_5\}$ fully exploits the multidimensional features of different depth feature layers, which can better supervise the feature fusion process of the network and make the fused features balance powerful semantic information and rich texture details and shape information.

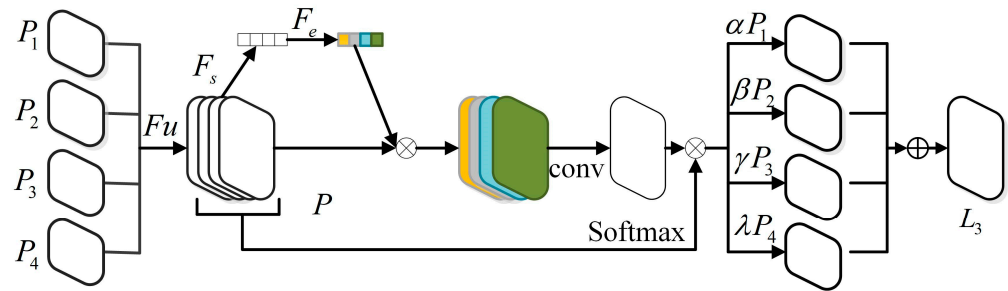


Figure 2. Multi-layer feature fusion schematic.

3.2. CBAM

Channel attention is concerned with the relationship between individual channels of features, whereas spatial attention is concerned with the relationship between individual pixels within the feature map channels. In the backbone network, to extract effective images and features, we designed the channel attention module and spatial attention module as the attention mechanism CBAM introduced into Yolov5 to improve the feature extraction capability and detection speed of the network-trained model. The specific model structure is shown in Figure 3.

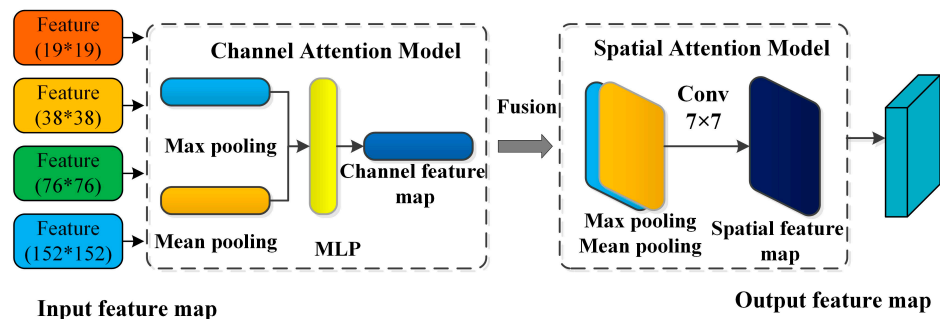


Figure 3. CBAM network structure diagram.

On the basis of max and mean pooling, the channel attention module (CAM) accumulates the input feature maps P of crop pests and diseases, and extracts thereafter the spatial information of the feature maps. The obtained spatial background description feature maps P_{avg}^c and P_{max}^c are processed through the multilayer perceptron (MLP) [16] to obtain the attention channel feature maps H_c .

The CAM formula is as follows:

$$\begin{aligned} H_c(P) &= \sigma(MLP(AvgPool(P)) + MLP(MaxPool(P))) \\ &= \sigma(\alpha_2(\alpha_1(P_{avg}^c)) + \alpha_2(\alpha_1(P_{max}^c))) \end{aligned} \quad (6)$$

For the spatial attention module (SAM), the input object is the output feature map of CAM. Through the average and max pooling of the input feature maps, the aggregated channel information of P_{avg}^s and P_{max}^s can be obtained. A 7×7 convolution method is used to generate a two-dimensional spatial feature map to improve the extraction of features. The spatial attention is calculated as follows:

$$\begin{aligned} H_s(G) &= \sigma(f^{7 \times 7}([AvgPool(P); MaxPool(P)])) \\ &= \sigma(f^{7 \times 7}([P_{avg}^s; P_{max}^s])) \end{aligned} \quad (7)$$

3.3. DETF

To further improve the extraction of effective information on UAV crop pests and diseases, a DETF module was established to improve the model for multiple pest and disease target feature learning. As shown in Figure 4, DETF consists of an encoder, decoder, and prediction unit. In the encoder, N objects are converted into embedded outputs by dimensioning the feature map of the backbone output. That is, the 2-dimensional feature map is convolved by a 1×1 convolution kernel to obtain multiple one-dimensional vectors, which are fed to the transformer-based encoder and decoder together with the position vectors. These embedded output vectors are independently encoded into N categories and prediction frames in parallel by a feedforward network with shared weights. Moreover, the entire input is predicted in parallel based on the transformer encoder-decoder structure.

Afterward, we applied a sequence transformation to the feature map in order to reduce the spatial dimension from $H \times W$ to HW . To encode the positions within the two-dimensional feature map, position coding was applied. This alternative representation allows for effective plagiarism detection. The position-coding formula is as follows:

$$PE_{(pos, 2i)} = \sin(pos/10,000^{2i/d}) \quad (8)$$

$$PE_{(pos, 2i+1)} = \cos(pos/10,000^{2i/d}) \quad (9)$$

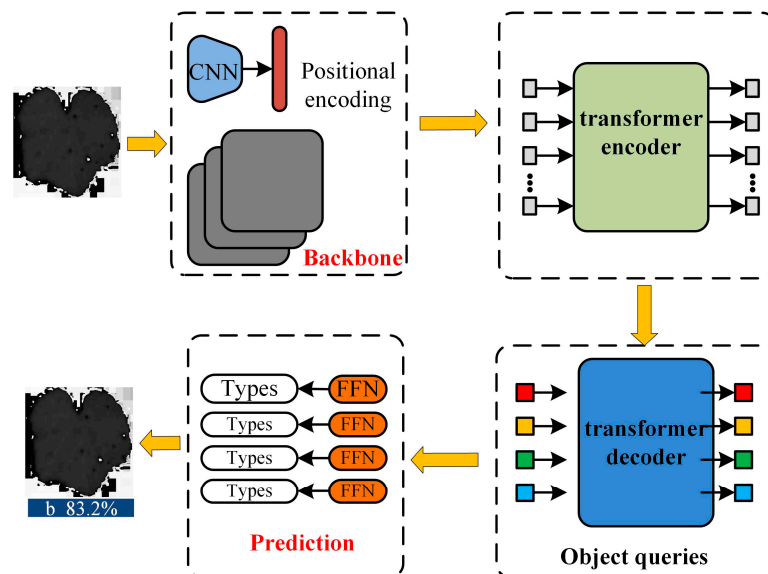


Figure 4. DETF schematic.

During the encoding stage of the transformer-based model, the attention matrix size is $(H \times W)(H \times W)$. The attention matrix in DETF maps to two different points in the feature block represented by each token. Both the number of input tokens in the encoder and the feature image elements define a frame, giving the DETF model unique advantages for target detection tasks. In addition, the presence of pests and diseases disrupts the product surface texture continuity. A DETF model with strong global feature learning capability can likewise tap into richer surface texture features. Hence, pest and disease detection is easily implemented.

4. System and Experiment

This section proposes a UAV detection system for crop pest identification and describes in detail the hardware and software system components. Accordingly, we conducted performance verification experiments to compare the proposed algorithm with existing algorithms including DaSiamRPN [19], Faster R-CNN [20], SSD [22], and YOLOv5. Based on the experimental results, we analyzed the superiority of the proposed algorithm in terms of performance.

4.1. UAVs

As shown in Figure 5, the proposed UAV system consists of a human–computer interaction handle (HCID), server, 5G module, PS, EPS, and AS. PS comprises four flight batteries, four motors, and eight propellers. EPS includes a binocular camera, phased array radar, inertial measurement unit (IMU), and infrared sensor.

In the established UAV, our system is based on DJI's UAV T40. T40 agricultural flying UAV adopts a co-axial dual-rotor design with a 50 kg load capacity, equipped with a dual fog spraying system, wise map system, and active phase control Chen radar and binocular vision perception system, integrating flying defense and aerial survey in one, precision agriculture. On this basis, we introduced a 5G module and infrared camera, server, IMU, and other software and hardware, focusing on agricultural pest detection. We optimized and upgraded software and hardware according to the demand for pest and disease information acquisition. Table 2 illustrates the hardware and software.

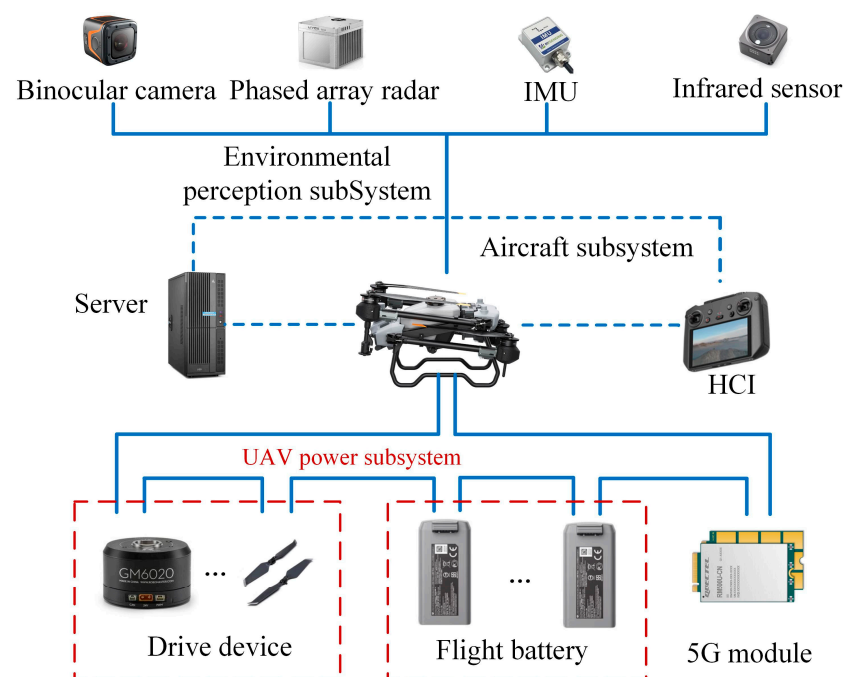


Figure 5. UAV system.

Table 2. UAV System Parameter.

Category	Name	The Main Parameter	Application
Hardware device	UAV body	Positional accuracy: ± 10 cm Volume: $2800 \times 3150 \times 780$ mm ³ Weight: 38 kg Range: 360°	The UAV
	HCID	Frequency: 5.615~5.850 GHz Signal: GPS + BDS WIFI protocols: WIFI 6	Trajectory planning and flight control
	Sever	CPU: Xeo GPU: RTX 3090 $\times 2$ RAM: 256 G	Storing and processing data
	Drive device	Volume: $\Phi 100 \times 33$ mm ³ KV value: 48 RPM/V Power: 4000 W/rotor Propellers Diameter: 54 inch Rotor number: 8 Model: BAX601 Capacity: 30,000 mAh	Driving UAV movement
	5G Module	Model: 5G RM500U Volume: $30.0 \times 52.0 \times 2.3$ mm ³ Temperature: -40 °C~ $+85$ °C	Remote communication
	Binocular camera	Range: 0.4–25 m FOV: level 90°, vertical 106°	Collecting all the displacement data.
	Radar	Model: RD2484R Range: 1.5–50 m FOV: level 360° vertical ± 45 °	Auxiliary perception
	IMU	Model: Honeywell HMS-MM-10 Range: ± 500 °/s, 16 g Volume: $40 \times 40 \times 15$ mm ³ Bandwidth: 200 Hz	UAV attitude balancing
	Infrared sensor	Model: LRCP20680 Resolution: 1920×1080	UAV obstacle avoidance
Operating system	Windows10	Windows10_19042.1165	The system in Sever and HCI
Software	Python3	Pycharm Community Edition 2021	Developing various algorithms
	Matlab	Matlab 2018 b	Data processing

UAV collects crop image data through low-altitude flight, and the active phased array radar fuses with an ultra-high definition binocular vision system for 360-degree omnidirectional sensing. The radar assists in sensing the crop operation area, the camera collects crop pest information in real time, and relevant data are transmitted to HCI and the server through the 5G module. The server dynamically stores and processes data according to the acquired data to complete the crop pest identification.

4.2. UAVs

A joint data set based on plant village and UAV data collection was built, and a comparison algorithm was set up to carry out performance tests, introducing evaluation indexes, such as recall, precision, and *mAP*. The performance of the algorithm was quantified and analyzed based on the experimental results. Figure 6 shows the experimental flow.

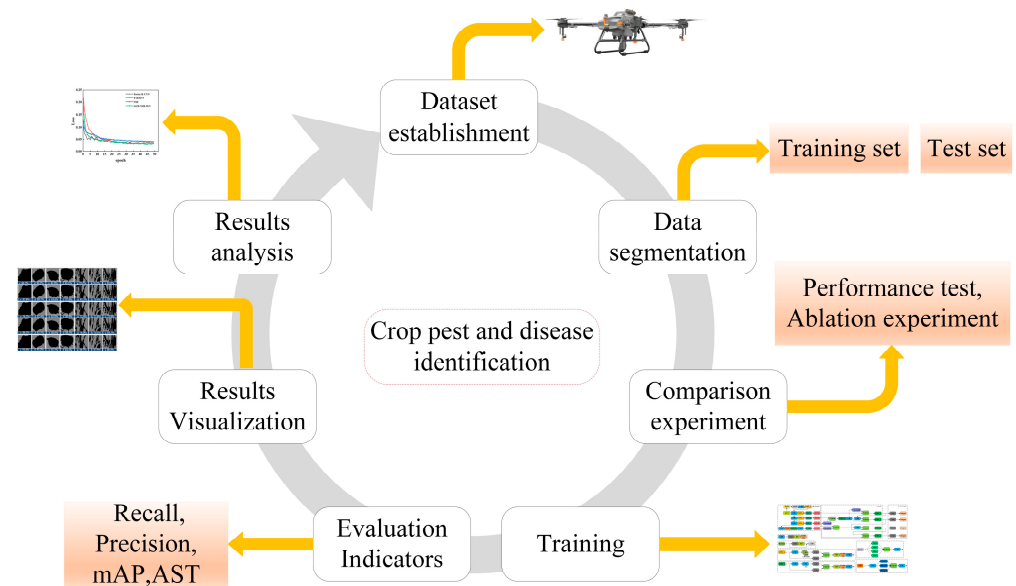


Figure 6. The flow of experiments.

4.2.1. Dataset

The dataset used in this paper is a fusion data set that integrates the public dataset plant village and drone collection of the rice data. Plant village spans 14 plant species with 26 categories of pest and disease leaf images and healthy leaf images. In this experiment, we selected 1250 images of common corn gray spot, grape leaf blight, pumpkin powdery mildew, strawberry banana leaf disease, potato late blight, and apple black rot (as shown in Figure 7), for a total of 7500 images in 6 categories, as the experimental data set. Moreover, we used 70% of the data for training, 20% for the test set, and 10% for the validation set.

UAV collected three types of pests as shown in Figure 7 (i.e., rice bacterial leaf blight, brown spot, and leaf smut), and obtained 100 corresponding pictures each through image processing.

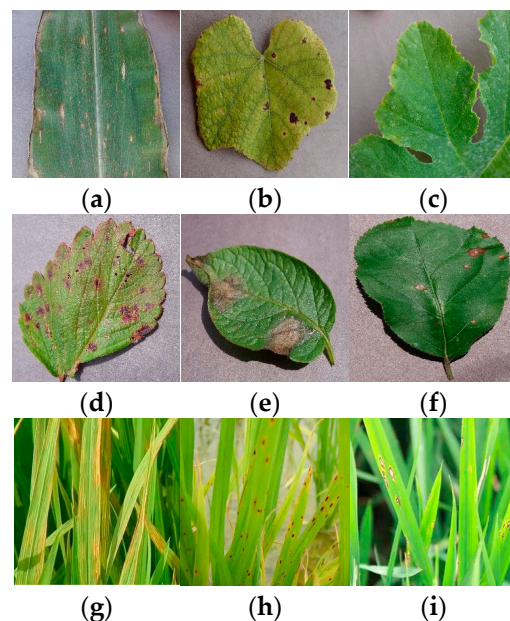


Figure 7. Sample images of the experimental data set. (a) Leaf spot of maize; (b) Grape leaf blight; (c) Powdery mildew; (d) Strawberry leaf disease; (e) Potato late blight; (f) Apple black rot; (g) Bacterial leaf blight; (h) Brown spot; (i) Leaf smut.

4.2.2. Setting and Training

By considering the evaluation of the performance of MCD-Yolov5, we selected SSD and faster R-CNN as comparison algorithms based on the detection accuracy from Table 1 and code reproducibility.

Given that the MCD-Yolov5 method proposed in this chapter was optimized on the basis of Yolov5, it was chosen as one of the comparison methods.

To guarantee the consistency of the algorithm, we selected the commonly used parameter values to complete the parameter settings for each algorithm. The initial learning rate is 0.01, weight decay takes 0.0005, batch_size is 8, and the intersection over Union (IOU) is 0.2. Moreover, we retained the algorithm parameters to maintain the performance of the algorithm.

Four commonly used evaluation indicators, namely, accuracy, precision, recall, mean average precision, and average prediction time, are proposed, represented as *ACC*, *PRE*, *REC*, *mAP* and *AST*, respectively. The definitions are as follows:

$$ACC = TP + TN / (TP + TN + FP + FN), \quad (10)$$

$$PRE = TP / (TP + FP), \quad (11)$$

$$REC = TP / (TP + FN), \quad (12)$$

$$mAP = \frac{\sum_{i=1}^k AP_i}{k} \quad (13)$$

In Equations (10)–(13), *TP* means true positive, *TN* indicates true negative, *FP* represents false positive, and *FN* is false negative. *AP* is the area under the curve of *PRE* and *REC*.

For *AST*, the total number of samples is *u* and the time taken by the model to make a classification prediction for the *i*th sample is *t_i*. *AST* is calculated as follows:

$$AST = \frac{\sum_{i=1}^n t_i}{u}, \quad (14)$$

4.2.3. Experimental Results and Analysis

The classification models of the four methods can be obtained through the training set based on the parameter settings in B. To verify the feasibility of the proposed algorithm, we validated the proposed algorithm based on the test set.

As illustrated in Figure 8, the precision and recall of the MCD-Yolov5 target detection network model stabilized and reached the fitted state after about 20 epochs of training. mAP@0.3 was determined by the target detection precision and recall, which also tends to be flat, thereby verifying the accuracy of the MCD-Yolov5 target detection network model.

The performance of the algorithms was further tested based on the test set for the comparative algorithms (i.e., DaSiamRPN, SSD, Faster R-CNN, and Yolov5), and the corresponding pests were numbered according to Figure 7.

Visualization of the results for a specific sample is shown in Figure 9 for the detection of mAP@0.3 by the five algorithms. The output of visualization provides a visual comparison of the performance of the proposed algorithm with other comparative algorithms.

- (1) Under the same training and test sets, the proposed MCD-Yolov5 algorithm has the highest identification AP (i.e., 95.7%), thereby demonstrating the superiority of our proposed algorithm;
- (2) The performance of Yolov5 and SSD was approximated, slightly higher than that of faster R-CNN and lower than that of DasiamRPN.
- (3) The first four methods have higher image data recognition accuracy for the plant village data set than that of the UAS acquisition, reaching 7.6%. The possible reason

is due to such factors as occlusion and illumination. MCD-Yolov5 proposed in this paper does not show a significant difference in performance in this respect, with an average difference of only 1.5%.

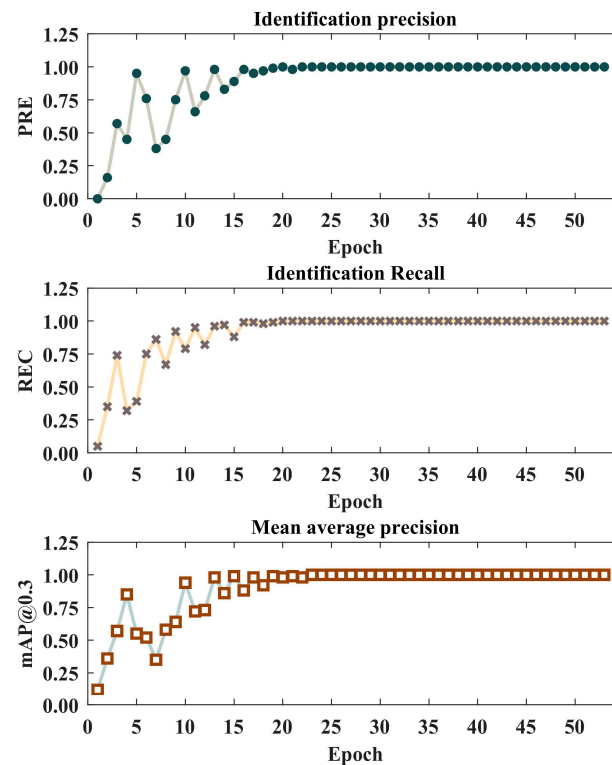


Figure 8. Identification results of the MCD-Yolov5.

The current challenge is to achieve rapid and accurate detection of crop pests and diseases.

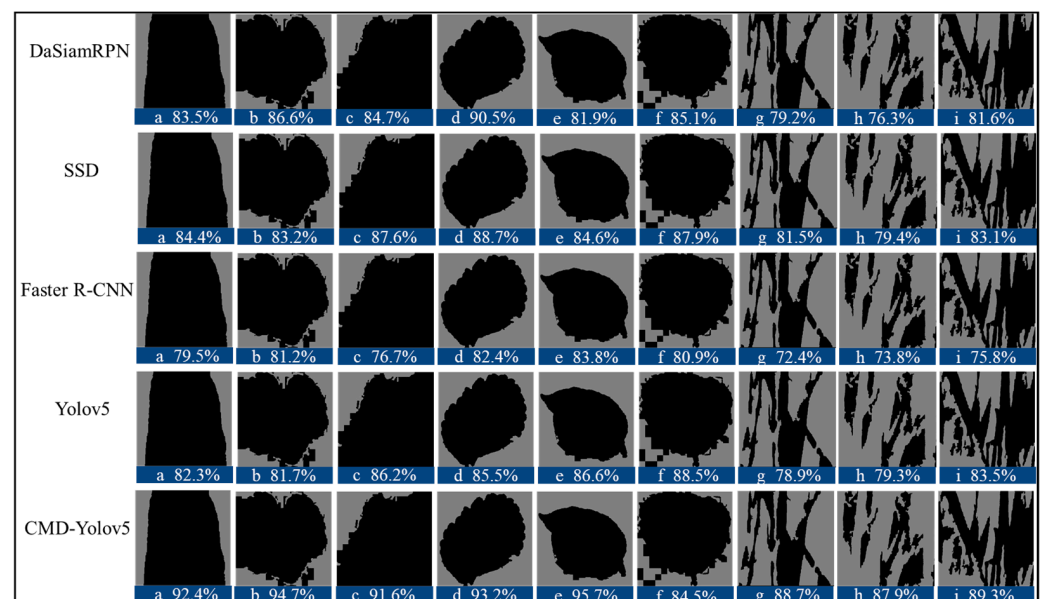


Figure 9. Visualization of identification results.

Table 3 shows the identification effect of each algorithm's performance on the test set. Moreover, the *mAP* of the four methods approximately 85%, and *PRE* did not exceed 80%. We can obtain the following conclusions.

Table 3. Test Results.

Algorithm	ACC	PRE	REC	mAP	AST (ms)
DaSiamRPN	83.60	79.25	77.19	88.29	3.67
SSD	80.07	74.12	74.07	85.37	1.83
Faster R-CNN	79.99	77.56	74.72	84.93	12.54
Yolov5	88.76	75.79	86.48	85.99	2.45
MCD-Yolov5	97.52	86.57	95.09	94.43	0.74

- (1) In terms of *ACC*, *PRE*, *REC*, *mAP*, and *AST*, the proposed MCD-Yolov5 has more performance advantages. In particular, *mAP* is 12.5% higher than the faster R-CNN;
- (2) The identification time of faster R-CNN is considerably beyond the others, but its performance is not superior. Faster R-CNN is a two-stage network that has a slower training process compared to Yolov5 and SSD, which directly outputs classification and localization results. This eliminates the need for a separate proposal generation stage, making the training process faster;
- (3) The performance of the *ACC*, *PRE*, *REC*, and *mAP* of the proposed algorithm has a significant increase, 8.76 %, 10.78%, 8.61%, and 8.44%, respectively, compared to the Yolov5, which proves the effectiveness and accuracy of the provided MCD module;
- (4) The MCD-Yolov5 outperforms other models in terms of both speed and accuracy, showcasing its efficient use of resources, faster convergence, and superior recognition capabilities.

To further validate the effectiveness of the proposed CBAM, MLFF, and DETF, ablation experiments were conducted on the validation set. The Yolov5 algorithm was used as the benchmark algorithm for the experiments, and each module was introduced separately for comparison. Among them, the CBAM and MFFL modules relied on the lightweight depth feature extraction network, which cannot be introduced directly. Therefore, the ablation experimental algorithms include Yolov5, D-Yolov5, MD-Yolov5, CD-Yolov5, and MCD-Yolov5.

The results of the ablation experiments are shown in Table 4. Evidently, the MCD proposed in this paper improves the discrimination accuracy of the benchmark algorithm, in which the multilayer fusion module has the highest effect, followed by the location information attention and the weakest depth feature extraction network. Moreover, the algorithm of this paper with the introduction of all three modules achieves the best among the five algorithms in terms of accuracy and robustness. The experimental results of the validation and test sets show that the optimized design of the CBAM, MFFL, and DETF modules can effectively improve the performance of crop pest identification, and the algorithm has good portability and scalability.

Table 4. Results of ablation experiment.

Model	MLFF	CBAM	DETF	mAP	AST (ms)
Yolov5	N	N	N	84.92	2.57
D-Yolov5	N	N	Y	87.25	3.13
MD-Yolov5	Y	N	Y	92.71	3.72
CD-Yolov5	N	Y	Y	88.36	0.58
MCD-Yolov5	Y	Y	Y	95.12	0.84

The MCD-Yolov5 demonstrates its ability to strike the perfect balance between resource efficiency and detection accuracy, making it an ideal choice for crop disease and pest identification.

5. Conclusions

Intelligent identification of agricultural pests and diseases is an important method to control pests and diseases. This paper proposes an MCD-Yolov5 agricultural pest identification algorithm based on the Yolov5 network target tracking algorithm, which integrates MLFF, CBAM, and DETF, and establishes UAV detection and control systems, including AS, EPS, and PS. CBAM uses channel information and location information to generate weights to optimize the feature extraction capability of the backbone network, DETF module to improve the model for multiple pest target feature learning, and MFFL to enhance the feature extraction capability of the proposed model. We established a joint data set based on UAV data collection and plant village and conducted algorithm performance validation experiments. The results show that the average prediction accuracy of this algorithm reached 88.12% and detection efficiency improved by 145%, which is superior to the comparative algorithms, such as SiamRPN and Yolov5. The method and system proposed in this paper provide references for solving agricultural pests and diseases.

In future research, we will build a 5G base station at the experimental site to realize online identification of pest species by UAV and develop mobile receiving devices to complete long-distance real-time pest identification. In addition, we will explore the balance between detection accuracy and detection in real time using more lightweight models.

Author Contributions: Conceptualization, L.L.; methodology, L.L.; software, H.Z.; validation, L.L. and N.L.; formal analysis, L.L.; investigation, L.L.; resources, L.L.; data curation, L.L.; writing—original draft preparation, L.L. and H.Z.; writing—review and editing, L.L. and N.L.; visualization, H.Z.; supervision, H.Z.; project administration, N.L.; funding acquisition, L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Beijing Natural Science Foundation Grant No. 4214071 and Beijing Municipal Education Commission 2023 Research Program General Project Foundation Grant No. KM202311232017.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Xia, L.; Robock, A.; Scherrer, K.; Harrison, C.S.; Bodirsky, B.L.; Weindl, I.; Jägermeyr, J.; Bardeen, C.G.; Toon, O.B.; Heneghan, R. Global food insecurity and famine from reduced crop, marine fishery and livestock production due to climate disruption from nuclear war soot injection. *Nat. Food* **2022**, *8*, 586–596. [[CrossRef](#)] [[PubMed](#)]
2. Sharma, R. Artificial intelligence in agriculture: A review. In Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 6–8 May 2021; pp. 937–942.
3. Lengai, G.; Muthomi, J.W.; Mbega, E.R. Phytochemical activity and role of botanical pesticides in pest management for sustainable agricultural crop production. *Sci. Afr.* **2020**, *7*, e00239.
4. Shaffer, L. Inner Workings: RNA-based pesticides aim to get around resistance problems. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 32823–32826. [[CrossRef](#)] [[PubMed](#)]
5. Han, M.; Liu, L.N.; Munir, S.; Bashir, N.H.; Wang, Y.; Yang, J.; Li, C.-Y. Crop diversity and pest management in sustainable agriculture. *J. Agric. Sci.* **2019**, *18*, 1945–1952.
6. Wang, K.; Chen, K.; Du, H.; Liu, S.; Xu, J.; Zhao, J.; Chen, H.; Liu, Y.; Liu, Y. New image dataset and new negative sample judgment method for crop pest recognition based on deep learning models. *Ecol. Inform.* **2022**, *69*, 101620. [[CrossRef](#)]
7. Liu, L.; Xie, C.; Wang, R.; Yang, P.; Sudirman, S.; Zhang, J.; Li, R.; Wang, F. Deep Learning based automatic multiclass wild pest monitoring approach using hybrid global and local activated features. *IEEE Trans. Ind. Inform.* **2021**, *17*, 7589–7598. [[CrossRef](#)]
8. Deng, H.; Zhang, Y.; Li, R.; Hu, C.; Feng, Z. Combining residual attention mechanisms and generative adversarial networks for hippocampus segmentation. *Tsinghua Sci. Technol.* **2022**, *27*, 68–78. [[CrossRef](#)]
9. Su, P.; Liu, D.; Li, X.; Liu, Z. A saliency-based band selection approach for hyperspectral imagery inspired by scale selection. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 572–576. [[CrossRef](#)]
10. Sharma, R.P.; Ramesh, D.; Pal, P.; Tripathi, S.; Kumar, C. IoT-enabled IEEE 802.15.4 WSN monitoring infrastructure-driven fuzzy-logic-based crop pest prediction. *IEEE Internet Things* **2022**, *9*, 3037–3045. [[CrossRef](#)]
11. Su, J.; Yi, D.; Su, B.; Mi, Z.; Liu, C.; Hu, X.; Xu, X.; Guo, L.; Chen, W.-H. Aerial visual perception in smart farming: Field study of wheat yellow rust monitoring. *IEEE Trans. Ind. Inform.* **2021**, *17*, 2242–2249. [[CrossRef](#)]

12. Nazari, L.; Zinati, Z. Gene expression classification for biomarker identification in maize subjected to various biotic stresses. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2023**, *20*, 2170–2176. [[CrossRef](#)] [[PubMed](#)]
13. Tonnang, H.E.; Sokame, B.M.; Abdel-Rahman, E.M.; Dubois, T. Measuring and modelling crop yield losses due to invasive insect pests under climate change. *Curr. Opin. Insect Sci.* **2022**, *50*, 100873. [[CrossRef](#)] [[PubMed](#)]
14. Shi, Y.; Han, L.; Huang, W.; Chang, S.; Dong, Y.; Dancey, D.; Han, L. A biologically interpretable two-stage deep neural network (BIT-DNN) for vegetation recognition from hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–20. [[CrossRef](#)]
15. Wittstruck, L.; Jarmer, T.; Trautz, D.; Waske, B. Estimating LAI from winter wheat using UAV data and CNNs. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [[CrossRef](#)]
16. Nagasubramanian, G.; Sakthivel, R.K.; Patan, R.; Sankayya, M.; Daneshmand, M.; Gandomi, A.H. Ensemble classification and IoT-Based pattern recognition for crop disease monitoring System. *IEEE Internet Things J.* **2021**, *8*, 12847–12854. [[CrossRef](#)]
17. Chen, X.; Ye, X.; Li, M.; Lou, Y.; Li, H.; Ma, Z.; Liu, F. Cucumber leaf diseases detection based on an improved faster RCNN. In Proceedings of the 2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 4–6 March 2022; pp. 1025–1031.
18. Abhilasa, S.; Srilakshmi, A.; Geetha, K. Classification of agricultural leaf Images using hybrid combination of activation functions. In Proceedings of the 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 6–8 May 2021; pp. 785–791.
19. Reddy, K.N.; Bojja, P. A novel method to solve visual tracking problem: Hybrid algorithm of grasshopper optimization algorithm and differential evolution. *Evol. Intell.* **2022**, *15*, 785–822. [[CrossRef](#)]
20. Hua, S.; Xu, M.; Xu, Z.; Ye, H.; Zhou, C. Multi-feature decision fusion algorithm for disease detection on crop surface based on machine vision. *Neural Comput. Appl.* **2022**, *34*, 9471–9484. [[CrossRef](#)]
21. Jin, Q.; Lin, R.; Yang, F. E-WACGAN: Enhanced generative model of signaling data based on WGAN-GP and ACGAN. *IEEE Syst. J.* **2019**, *14*, 3289–3300. [[CrossRef](#)]
22. Sustersic, T.; Rankovic, V.; Milovanovic, V. A deep learning model for automatic detection and classification of disc herniation in magnetic resonance Images. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 6036–6046. [[CrossRef](#)]
23. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
24. Tian, Y.; Yang, G.; Wang, Z.; Li, E.; Liang, Z. Detection of apple lesions in orchards based on deep learning methods of cycleGAN and yolov3-dense. *J. Sens.* **2019**, *2019*, 7630926. [[CrossRef](#)]
25. Gao, M.; Ma, S.; Zhao, L.; Zhang, Y. Soybean pest recognition based on YOLOv4 and graph cut algorithm. In Proceedings of the 2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC), Virtual, 12–14 November 2021; pp. 212–216.
26. Yao, H.; Liu, Y.; Li, X.; You, Z.; Feng, Y.; Lu, W. A detection method for pavement cracks combining object detection and attention mechanism. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 22179–22189. [[CrossRef](#)]
27. Mekhalafi, M.; Nicolo, C.; Bazi, Y.; Al Rahhal, M.M.; Alsharif, N.A.; Al Maghayreh, E. Contrasting YOLOv5, transformer, and efficient Det detectors for crop circle detection in desert. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 22179–22189.
28. Jubayer, F.; Alam Soeb, J.; Mojumder, A.N.; Paul, M.K.; Barua, P.; Kayshar, S.; Akter, S.S.; Rahman, M.; Islam, A. Detection of mold on the food surface using YOLOv5. *Curr. Res. Food Sci.* **2021**, *4*, 724–728. [[CrossRef](#)] [[PubMed](#)]

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