Computer Science and Information Systems 23(1):369–396 https://doi.org/10.2298/CSIS251109012Y

FNNMFF: Crop Pests and Diseases Detection Based on Fuzzy Neural Network and Multilevel Feature Fusion in Remote Sensing Images

Shoulin Yin[1], Hang Li[2,⋆], Mirjana Ivanovic[3,⋆], Tao Chen[1], and Lin Teng[1]

[1] College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China
[2] College of Artificial Intelligence, Shenyang Normal University, Shenyang 110034 China
lihang@synu.edu.cn
[3] Faculty of Sciences, University of Novi Sad, Novi Sad, Serbia
mira@dmi.uns.ac.rs

Abstract. To solve the problem that the detection effect of crop pests and diseases is not ideal due to the complicated image background and the interference of irrelevant factors, this paper proposes a novel crop pests and diseases detection based on fuzzy neural network and multilevel feature fusion in remote sensing images. Firstly, the model is based on YOLOv5 and extracts the semantic level information of different depth features from the convolutional neural network, and then combines the weight aggregation module to learn the weight of each layer feature adaptively. Then the learned weights are loaded to the segmentation graphs obtained by sampling on each feature layer to obtain the final segmentation results. In this model, a fuzzy learning module is added to the skip connection part to remove noise features and alleviate the uncertainty between classes. The traditional cross entropy loss involves activating the output value with the Softmax function and calculating a weighted cross entropy loss with the label. If the weight of the cross entropy loss term is not adjusted, the model will tend to update the weight related to the background, which makes it difficult to deal with the category imbalance in remote sensing images. Therefore, we use focus loss to alleviate the problem of class imbalance in images. The results on public data sets show that the accuracy rate of the proposed model in this paper is over 95%, the recall rate is over 85%, and the average accuracy is 91.2%. In terms of F1, compared with other advanced methods, the presented method has achieved improvements of 6.6%, 8.5%, and 7.7% respectively. It shows that the new model has strong robustness and generalization for crop pest detection.

Keywords: Remote sensing images, fuzzy neural network, crop pests and diseases detection, multilevel feature fusion, focus loss.

1. Introduction

Globally, the world is home to more than eight billion people, and safeguarding food security is fundamental to every nations economic stability, public well-being and social

⋆ Corresponding authors

cohesion. Transboundary pests and diseases rank among the most destructive threats to agriculture: they occur in countless forms, spread rapidly across regions and climates, inflict severe yield losses and can erupt into large-scale outbreaks. Left unchecked, they endanger global food supplies, rural livelihoods and the achievement of the Sustainable Development Goals [1-3]. Over 70% of the world's food crops are distributed in vast, uniform rain-fed plains or irrigated areas (such as the black soil belt in Northeast China, the middle and lower reaches of the Yangtze River Plain, and the corn belt in the United States). These areas have highly consistent ecological backgrounds, and remote sensing pixel-scale (meters to sub-meters) can cover hundreds or thousands of farmlands, making "spatial extrapolation" possible. Within the same ecological zone, the occurrence patterns of pests and diseases are driven by climate, soil, and cultivation systems, and are highly replicable. Therefore, as long as a remote sensing-ground verification framework is established in representative ecological areas, it can provide decision support for the entire ecological zone by using the representative areas to represent the entire zone, greatly reducing the cost of ground surveys.

Diseases and pests such as rice stem borer, fall armyworm, wheat stripe rust, and potato late blight, which are considered high-risk, can cause a yield loss of $\geq$30% within 5 to 7 days if they break out. The average coverage per person per day through traditional manual field surveys is only $1-2km^2$. In contrast, satellites or drones can obtain images covering hundreds to thousands of km^2 in a single pass, with time resolution reaching "daily" or even "hourly". In scenarios where the epidemic spreads rapidly and the affected area is large, only remote sensing methods can meet the time requirements of "early detection, early warning, and early response".

In recent years, the increasing global climate change has led to an increasing trend in the distribution range and harm degree of crop pests and diseases. According to statistics, in 2020, the global yield losses of wheat, rice and corn caused by pests and diseases reached 22.6%, 31.1% and 23.7%, respectively. The global food production loss caused by pests and diseases accounts for about 14% of the global total food production, among which, the food production loss caused by pests accounts for about 10% of the global total food production. Diseases and insect pests occur in different stages of crop growth and development, and the diseases and insect pests occur in different stages of crop growth. Therefore, diseases and pests should be closely monitored at different growth stages of crops and actively controlled to avoid major losses [4-6].

Precise prevention and control of diseases and pests is an effective measure to reduce losses caused by diseases and pests, improve crop yield and agricultural product quality, reduce pesticide pollution and waste, ensure the safety of agricultural products, protect the ecological environment and increase farmers' income. Accurate monitoring of pests and diseases is the premise and basis of accurate prevention and control of pests and diseases. At present, the control of crop pests and diseases in our country is mainly based on the judgment result of whether the pests and diseases occur and the degree of harm. Manual monitoring method is time-consuming, time-poor and subjective [7-9], which is easy to cause problems such as leakage prevention of diseases and pests and low prevention and control efficiency. Moreover, this pest control method generally ignores the difference in the degree of damage caused by diseases and pests in different regions, and adopts a uniform spraying method, resulting in resource waste, pesticide residues, agricultural product damage and environmental pollution.

The development of modern agriculture and the demand for precision medicine applications have put forward new requirements for pest monitoring. Crop diseases and pests mainly act on the metabolic system of crops, causing damage to the physiological function and structure of the internal system of crops. External manifestations are wilted, speckled, mildew powder and other obvious characteristics. Internal manifestations are changes in biophysical and chemical characteristics, such as the concentration or content of pigments such as chlorophyll and carotene, as well as a decrease in nutrient composition. These changes will lead to changes in the spectral characteristics of crops. Therefore, the biophysical properties and spectral characteristics of crops can be monitored through machine vision, hyperspectral technology and remote sensing technology, which is the research hotspot of crop pest monitoring [10,11].

With the rise of computer vision technology and artificial intelligence, image processing technology has been vigorously developed, because of low cost, easy to implementation, these technologies are widely used in agricultural pest detection. Traditional agricultural image processing mainly adopts machine learning algorithms such as support vector machine [12,13] and K-nearest neighbor [14], but the traditional machine learning technology needs to rely on manual feature selection and extraction, which has weak generalization ability and great environmental restrictions, and is not suitable for application in the actual orchard environment. Deep learning is a specific field of machine learning. Compared with machine learning, deep learning can automatically extract feature information, reduce the burden of feature extraction engineering and manual extraction prone to errors, suitable for processing large-scale data sets, and provide better prediction and generalization capabilities. Common target detection models in deep learning include Faster RCNN [15], SSD [16], YOLO series [17,18], etc., which have been widely applied to agricultural pest detection research. Li et al. [19] proposed a new Faster R-CNN architecture to detect seven different strawberry diseases and achieved good detection accuracy. Cheng et al. [20] constructed a lightweight convolutional neural network MEAN-SSD to reduce the size of the model, and successfully detected five apple leaf diseases on mobile devices. Faisal et al. [21] identified the diseases of citrus fruits by combining YOLOv4 and EfficientNet models. Ma et al. [22] detected cucumber diseases based on YOLOv5n, and introduced coordination attention mechanism and a Transformer structure to improve the detection accuracy and robustness of the model. Zhu et al. [23] used the improved YOLOv5s to achieve rapid detection of apple fruit diseases. Majeed et al. [24] realized the classification of monocotyledonous plants and dicotyledonous plants by identifying and detecting the morphological characteristics of leaves through image digitalization, and the accuracy rate reached 60%-80%. Fan et al. [25] studied a weed recognition method based on color features, which was not easily affected by factors such as shooting distance and occlusion. The test results showed that the recognition rate of wheat and weed by this method reached 54.9% and 62.2% respectively. Zhang et al. [26] proposed an improved single-step multi-frame detector model based on lightweight convolutional neural networks combined with feature information fusion mechanism. The VGG16 network in SSD model was replaced by a lightweight pre-basic network to improve the speed of image feature extraction. The deep semantic information and shallow semantic information were integrated to improve the detection accuracy of small crops and weeds. The final results showed that the average accuracy was 88.27%, and the detection speed and parameter number had been greatly improved. Yang et al. [27] proposed a maize leaf pest

detection model for YOLOv3-Corn. This model used Darkknet-53 as the feature extraction network, and used the clustering algorithm to select prior boxes and match them to the detection layer for target identification. The accuracy and recall rate of model detection were 93.31% and 93.08%, respectively. Guan et al. [28] studied the detection problem of corn seedlings, improved the Cascade R-CNN model, used residual network and feature pyramid as feature extractors, and made the average accuracy of corn seedling detection reach 91.76%.

At present, most studies on crop pests and diseases tend to focus on classification tasks [29] with few object categories, and the data sets used are taken in a laboratory environment based on a simple background. Due to the similar characteristics among crop pests and diseases, complex natural environment, and occlusions of branches and leaves, the existing models are prone to problems such as missed detection and false detection, and can not meet the requirements of multi-species and multi-target pest detection in a natural environment. To solve the above problems, this paper proposes a novel crop pests and diseases detection based on fuzzy neural network and multilevel feature fusion in remote sensing images, aiming to assist crop farmers in detecting diseases and pests, and provide an algorithm basis for intelligent products such as orchard drug application robots. The main contributions are as follows.

1. We introduce the CXV2 backbone block that replaces C3 in YOLOv5s, boosting feature diversity under complex field background.
2. We design a lightweight Fuzzy Learning Module (FLM) inserted in skip-connections to suppress mixed-pixel uncertainty at object boundaries.
3. We equip the neck and head with CARAFE up-sampling and DyHead, achieving accurate multi-scale pest localization in 0.5-2m resolution UAV images.
4. The resulting FNNMFF detector, trained with EIOU loss and SimAM attention, improves mAP by 1.9% and F1 by 6.6% while using only 1.29M parameters, enabling real-time mobile deployment.

2. Related Works

The traditional management mode of crops during planting mainly relies on manual observation of pests and diseases, which has some problems, such as strong subjectivity and heavy workload. With the continuous development of artificial intelligence and big data technology, the use of machine vision technology based on deep learning to extract image shape, color, texture and other feature information can effectively solve the problem of low accuracy and low efficiency of disease and insect detection during crop planting, which is also an important research direction of intelligent crop planting management [30]. At present, many scholars have carried out a lot of research on the application of deep learning and machine vision technology to crop management.

In order to realize accurate monitoring of crop pests and diseases, it is necessary to select an appropriate algorithm structure and build a relationship model between the selected spectral characteristics and the types of pests and diseases and the degree of harm. At present, classical statistical models, computer image processing methods, machine learning methods and deep learning algorithms are widely used in remote sensing monitoring and forecasting of crop pests and diseases.

2.1. Classical Data Statistical

Classical statistical analysis methods mainly include regression models, principal component analysis, cluster analysis, discriminant analysis and other methods [31-33]. Due to its advantages of simple form and clear mechanism, it is widely used in the monitoring and research of crop pests and diseases. Santos et al. [34] analyzed the spectrum of correlation coefficients by analyzing regression from partial least squares (PLS), and determined the effective wavelength by a stepwise multiple linear regression (SMLR) procedure. These analytical methods revealed several important regional wavelengths and produced predictive models of disease severity based on absorption spectra. The optimal model was used to predict disease severity in the validated dataset, with the root-mean-square difference (RMSD) of 4.9% and the coefficient of determination R2 of 0.82. Deng et al. [35] analyzed the hyperspectral data of wheat canopy stripe rust at different growth stages, and used statistical methods such as t-test and correlation analysis to screen the characteristic bands. A set of feature band screening methods for disease diagnosis was established. The four wavelengths suitable for early detection of stripe rust were 576nm, 705nm, 712nm, 1416nm and the five wavelengths suitable for intermediate detection were 558nm, 632nm, 375nm, 696nm, 712nm. Lin et al. [36] obtained canopy spectra and disease indices of infected winter wheat at different growth stages, and extracted the first five principal components using principal component analysis within the range of 350nm 1350nm. And the first three principal components of the first order differential spectrum in the blue (490nm 530nm), yellow (550nm 582nm) and red (630nm 673nm) bands were used to construct the disease inversion model for detecting winter wheat stripe rust. Chattopadhyay et al. [37] used ASD spectrometer to measure two kinds of hyperspectral data of wheat leaf stripe rust and mildew. Through correlation analysis and independent t-test, the spectral characteristic bands with obvious difference in sensitivity of the two diseases were selected, and the disease discrimination model and disease severity inversion model were constructed. Kolipaka et al. [38] selected wheat 25 planting yield index and DI to conduct linear, stepwise regression and BPNN modeling, and screened the optimal wheat leaf rust monitoring model at different growth stages. When studying cotton canopy leaves, Li et al. [39] found that the multivariate stepwise regression model established based on the first derivative spectrum of cotton leaves could effectively estimate the chlorophyll content of cotton leaves.

2.2. Computer Image Processing Methods

With the rapid development of science and technology, image processing technology has been widely used in the identification and diagnosis of crop pests and diseases. After crops are infected by pathogenic bacteria, a series of metabolic changes will occur, resulting in changes in the external characteristics of crop leaves such as color, shape, texture and spectrum. Different diseases have different plaque characteristics caused by different pathogens. Therefore, it is possible to determine whether crops are infected with diseases through the image information of crop leaves [40].

The research on pest recognition and diagnosis based on traditional image processing technology mainly includes four parts: image preprocessing, image feature extraction, image feature reduction and pest pattern recognition model construction. Tran et al. [41] used the method of spectral band combination to study the changes between the red,

green and blue spectral channels of wheat plants and determine the area of wheat diseased leaves, thus realizing the distinction between healthy plants and diseased plants. Kumar et al. [42] classified wheat leaf rust based on infrared thermal imaging edge detection algorithm. Mumtaz et al. [43] proposed an improved threshold median filtering algorithm, which could effectively remove noise and retain image details.

In the aspect of image feature extraction, edge detection is an important research method in computer vision, and it is also a key technology for analyzing crop images. Edge detection is not only an important method of digital image processing, but also a key technology of crop image analysis. Maraveas et al. [44] introduced the ant colony algorithm into the thermal image processing of maize drought, and proposed to apply edge detection to infrared thermal image analysis. The purpose of feature selection was to select more favorable features for pest identification. Principal component analysis was widely used as a common linear dimension reduction method. Masood et al. [45] proposed a method for color image recognition of maize leaf pests and diseases by using a manifold learning algorithm in RGB color space. Siddiqa et al. [46] proposed a nonlinear dimensionality reduction method for multispectral images, aiming at the problem of high data dimension of multispectral images resulting in high data processing complexity in image color reproduction. After the feature selection of plant pests and diseases, support vector machine and artificial neural network are generally used to complete the classification of plant pests and diseases. Chodey et al. [47] proposed a pest control method based on image gray frequency and artificial neural network. Gupta et al. [48] discussed a rice canopy image segmentation algorithm based on the combination of support vector machine and maximum inter-class variance method to solve the difficult problem of rice canopy image segmentation caused by variable light intensity in the field.

Traditional image processing has poor migration ability among plant pests and diseases, and the performance of lesion segmentation and pest recognition is not ideal. The emergence of convolutional neural networks effectively improves the migration ability of plant pests and diseases, and its end-to-end feature extraction capability provides technical support for real-time and accurate diagnosis of plant leaf diseases and pests. Chen et al. [49] proposed a low-resolution rice pest recognition network based on SCResNet (Self-calibrated convolutions and ResNet block for ResNet50) in order to solve the problem of low quality pest images captured in the natural environment of rice fields. By pruning the low-resolution image, constructing and training the ESRGAN network to generate Super-resolution (SR) image equivalent to the original image, the accuracy of rice pest identification was further improved. Khanna et al. [50] used convolutional neural network and computer vision technology to realize image recognition of a total of 12 leaf diseases of three crops (soybean, corn and ginseng), and improved the existing convolutional neural network model to significantly improve the model recognition rate. Under the framework of deep learning Keras, Guerrero-Ibanez et al. [51] trained the deep convolutional neural network with the normalized data of three rice diseases and established three rice disease recognition models. Imanulloh et al. [52] proposed a convolutional neural network image recognition model of maize pests and diseases based on transfer learning in order to realize image recognition of maize pests and diseases under complex field background. Du Al-Timemy et al. [53] combined the deep feature fusion algorithm with the convolutional neural network to remove the redundant features of the feature image on the premise of better retention of relevant information, and realized the rapid detection of tomato leaves.

Sahasranamam et al. [54] proposed a rice pest and disease identification method based on the Internet of Things and deep convolutional neural network, aiming at the problem that the rice pest and disease identification method based on image segmentation and convolutional neural network was sensitive to light and background.

At present, the application of computer image processing technology is very wide. In the process of agricultural production, image processing technology can be used for real-time monitoring of crop pests and diseases. Compared with manual monitoring, image processing technology can obtain pest information accurately and quickly, making diagnosis more accurate and effective. In the future, the combination of image processing technology and computer technology will make the monitoring of agricultural crop pests and diseases develop in the direction of automation and intelligence, which will inevitably improve the efficiency of agricultural production.

2.3. Machine Learning Methods

In recent years, machine learning methods have been gradually applied in the fields of canopy height (CH) and vegetation coverage (VFC) of crop diseases monitored by remote sensing and pest stress. The main components of machine learning methods and satellite-based spectra calculated by Sentinel-2A/B data are classical machine learning methods and deep learning methods. A classical learning machine uses machine learning models, including partial least squares regression (PLSR), support vector regression (SVR), random forest regression (RFR), and extreme learning regression (ELR), to evaluate the potential of drone data to replace manually sampled data and predict field disease indices (DI). The results show that the performance of SVR method is slightly better than other methods, mainly including support vector machine, decision tree and K-nearest neighbor algorithm, etc. The required training sample data is small, the equipment performance requirements are lower, and the model is easier to understand. Zou et al. [55] obtained the root mean square error (RMSE) of 1.89% by analyzing the UAV. In addition, the combination of canopy structure (CS) and vegetation index (VIs) improved the prediction accuracy compared with a single type feature (RMSEs 2.86% and RMSEVIs 1.93%). In order to distinguish wheat powdery mildew from aphids on a regional scale, Zhang et al. [56] adopted BPNN and support vector machine (SVM) methods based on the original training data set to conduct preliminary tests on crop growth and environmental parameters. The results showed that the overall accuracy of the two-time growth index model, SMOTE-BPNN model, BPNN model and SVM model was above 80% in combination with the growth and environmental parameters of different crop periods. The combination of SMOTE and a BP neural network could effectively improve the identification accuracy of small individual diseases and pests.

Deep learning is an emerging branch of machine learning that enables automated feature learning and pattern recognition by building deep neural network models. The main feature of deep learning is that it has a multi-layer network structure that can learn an abstract hierarchical representation of the data to extract more advanced and abstract features. In some complex scenarios, deep learning methods have shown powerful capabilities in monitoring crop pests and diseases. Irmak et al. [57] used AlexNet and GoogleNet to train 14828 images of tomato leaves infected with diseases, understand the symptoms of the leaves and locate the disease area through visualization, and realized the classification of tomato leaf images containing nine diseases. Roy et al. [58] used hyperspectral and

UAV remote sensing technologies to monitor cotton root rot, and the results showed that both could accurately identify the susceptible area under unsupervised classification. Sun et al. [59] used UAV remote sensing technology to monitor potato late blight and used five classification methods to classify disease areas, and the results showed that linear support vector machine and random forest algorithm had the best classification effect. Nayak et al. [60] proposed a transfer learning method to improve the problem of insufficient sample quantity of the rice lesion image set. The algorithm was based on pre-trained DDC deep transfer learning, and the rice image set is transferred to a pre-trained convolutional neural network model on a large dataset for retraining. Aiming at the problem that different training samples differ greatly, the DDC algorithm is proposed for optimization, and the final training results are compared with those directly trained on the training set, which verifies the effectiveness of the proposed algorithm and proves that the algorithm has stronger representation ability. Nguyen et al. [61] collected the data information of wheat powdery mildew, stripe rust, aphids, total erosion and other diseases, and used pattern recognition and machine learning algorithms to distinguish, estimate and predict various disease types.

Kumar et al. [62] proposed a recognition model based on a convolutional neural network, which combined batch normalization and global pooling techniques and was able to effectively identify 26 diseases of 14 different plants. The model was less affected by changes in leaf spatial position and had high recognition accuracy and robustness. Zhang et al. [63] made a comparative analysis of discrimination algorithms such as Extreme learning machine (ELM), RBF neural network, random forest, support vector machine and K-nearest neighbor (KNN) to identify the healthy and diseased areas of rape leaf disease parts.

In summary, deep learning has many advantages over traditional machine learning, including a high degree of automation, fast computing speed, and the ability to learn features autonomously. Through deep learning, tasks and features can be jointly modeled to extract more accurate information. However, in more complex scenarios, there may be a low recognition rate.

3. Proposed Pest and Disease Detection Model

Among the YOLO series models, YOLOv5 has the advantages of high detection accuracy, fast speed, small model and easy deployment, and it is often used in various environmental target detection scenarios. According to the model size, YOLOv5 is divided into five versions. Through comparative experiment analysis on self-built data sets, YOLOv5s is selected as the baseline model in this study. The YOLOv5s model is mainly divided into four parts: input layer, backbone network, neck network and detection head. The input layer premanipulates the detected image and passes the image into the model. The backbone network is used to extract the features in the input image, and the CSPDarknet53 architecture is used to effectively improve the efficiency of feature transmission. The neck network is the feature fusion layer. Feature pyramid network (FPN) [64] and path aggregation network (PAN) [65] are used to fuse feature maps of different scales to improve the detection ability of different feature layers. The detection head is used to map the extracted features and output the location, category and confidence information of the predicted target.

In order to improve the detection performance of crop pests and diseases in a natural complex environment, a new crop detection model FNNMFF is proposed in this study. A CXV2 module is designed to replace the C3 module in the original model, and the dynamic detection head (DYHEAD) is used to replace the original detection head, and the upsampling mode in the neck network is replaced by the CARAFE upsampling operator. The detection accuracy of crop pests and diseases is improved on the basis of not significantly increasing the calculation amount of the model. The network structure of FNNMFF is shown in Figure 1.

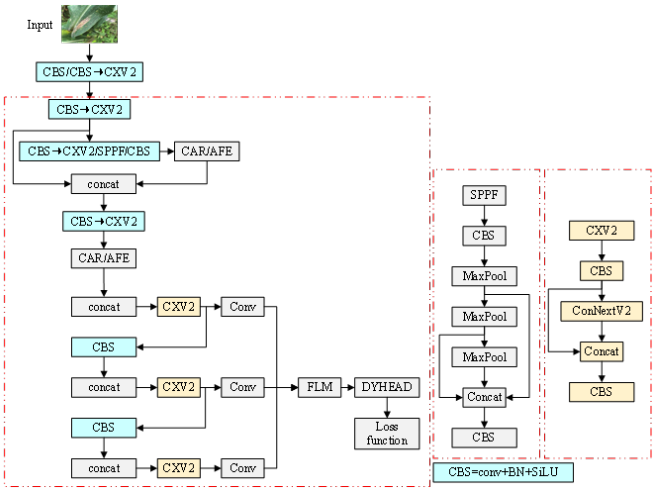

Fig. 1. Proposed FNNMFF model. Given a UAV or satellite remote-sensing image, the backbone first extracts multi-level features via the improved CXV2 blocks; the neck then fuses them through CARAFE up-sampling and PAN/FPN, while the embedded Fuzzy Learning Module (FLM) suppresses uncertain edges; finally, the Dynamic Head (DYHEAD) outputs pest-disease bounding boxes

3.1. CXV2 Module

In order to improve the feature extraction ability of the model, obtain channel features with more resolution, and enhance the generalization ability of the model, the ConvNeXxtV2 model is introduced in this study, and a CXV2 module is designed to replace the C3 module in YOLOv5s. ConvNeXxtV2 adopts the full convolutional mask autoencoder (FCMAE) mechanism, which combines the mask autoencoder MAE mechanism with self-supervised learning technology to improve the training speed and accuracy of the model. The structure of the ConvNeXtV2 module is shown in Figure 2. The global response normalization (GRN) layer is added to the original architecture of ConvNeXxt, which can solve the feature crash problem in the training process. The combination of deep convolution operation to enhance the feature competition between channels, im-

prove channel contrast and selectivity, is conducive to improving the diversity of extracted features in complex background images.

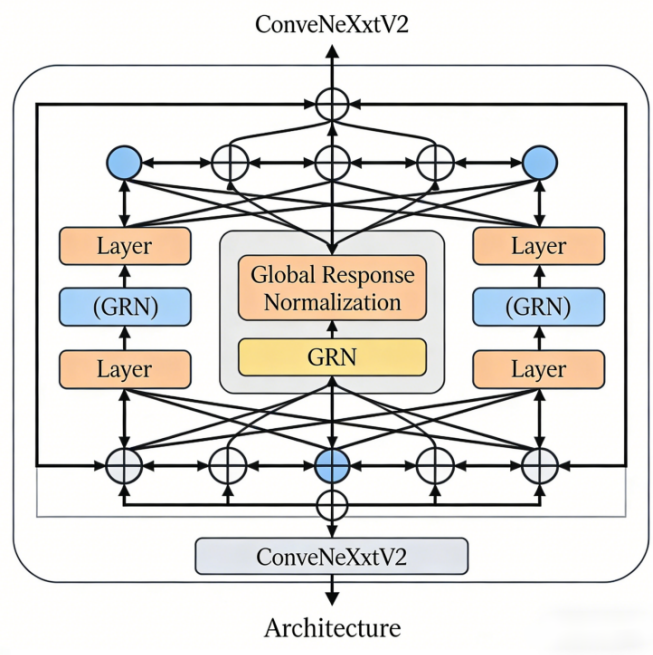

Fig. 2. ConvNeXxtV2 module structure

The CXV2 module uses a residual-like structure. Input features are first extracted through a 1×1 standard convolution block CBS to reduce the number of channels, and then transmitted to two different branches. One of the branch features is passed into the ConvNeXxtV2 module, which enhances feature extraction through module processing. The features of the other branch remain unchanged. Finally, all branches are spliced, and the number of channels is recovered by 1×1 standard convolution block operation to output features that enhance diversity and improve the feature expression ability of the model.

3.2. Dynamic Detection Head (DYHEAD)

Different diseases and pests have similar pathological characteristics, and in the natural environment, they are easily confused with the background. The scale and different spatial locations of the diseased leaves also bring challenges to feature extraction and target detection. To solve the above problems, a dynamic head is added in this study, and the frame of the dynamic head is shown in Figure 3. The features of the original feature pyramid network are scaled back and reshaped into a three-dimensional tensor $F \in R^{L \times S \times C}$

with the same scale, and different attention mechanisms are used in each independent dimension. Here L is the feature hierarch; S is the product of width and height; C indicates the number of channels.

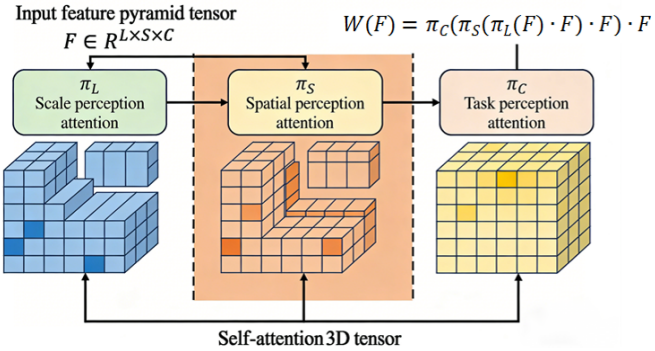

Fig. 3. Dynamic detection head

In order to avoid the high computation caused by the fully connected layer, the dynamic detection head adopts the mode of 3 sequential attention, and its self-attention equation is shown in formula (1), each attention module corresponds to only one dimension.

$$W(F) = \pi_C(\pi_S(\pi_L(F)\cdot F)\cdot F)\cdot F. \tag{1}$$

Where π_L is the scale perception attention function, increasing the perception ability of different scale features. π_S is the spatial perception attention function, which enhances the spatial position perception ability of the model. π_C is a task perception attention function to enhance the perception of different task objectives.

3.3. Carafe Upsampling Module

In the YOLOv5s model, feature pyramid is an important part of neck feature extraction and fusion. The model needs to match and fuse deep semantic information with shallow semantic information by upsampling. The original YOLOv5s model uses the nearest neighbor upsampling method, which only determines the upsampling kernel based on the spatial position of pixels, resulting in a small perceptual domain and is unable to make full use of the semantic information of feature maps. In this study, the upsampling module of CARAFE is used to replace the upsampling operator of the original model.

In the kernel prediction module, 1×1 convolution kernel is first used to compress the number of channels, and the original number of channel C is reduced to C_m, which reduces the subsequent calculation and improves the efficiency. After that, the pixelshuffle is used to reassemble the height, width and channel number of the feature map. The convolution layer of $K_{encoder} \times K_{encoder}$ is used to predict the upper sample kernel for the feature map after the compressed channel number, and the shape of the predicted upper sample kernel is $\sigma H \times \sigma W \times K_{up}^2$.

3.4. Fuzzy Learning Module (FLM)

In remote sensing images, small targets and boundary pixels are affected by mixed pixels and noise, which leads to more uncertainty in segmentation results. Compared with the traditional precise mathematical statistics method, fuzzy theory can deal with the uncertainty problem more effectively. In this study, a new FLM is designed and proposed to improve the segmentation accuracy of small targets and boundary regions. As part of a network, FLM can be seen as a fuzzy layer described by multiple parameters in a Gaussian membership function. The FLM structure is shown in Figure 4, where $H \times W \times C$ represents the size of the input feature map, and H, W and C represent the height and width of the feature map and the number of channels of the input feature map respectively.

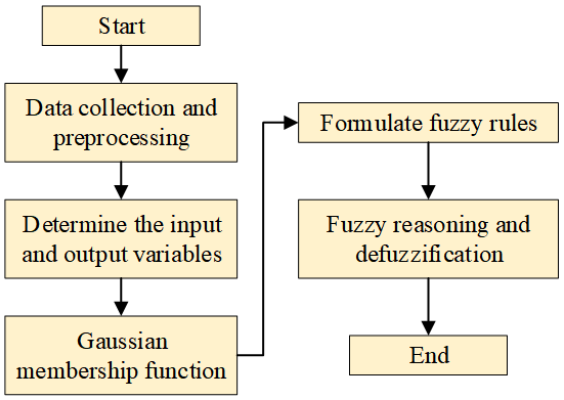

Fig. 4. FLM structure

In FLM, each Gaussian membership function assigns a fuzzy semantic tag to a feature point, and the Gaussian membership function is expressed as:

$$G_g^c(n_1, n_2) = e^{\frac{n_2^2+n_1^2}{22}}. \tag{2}$$

Where g and σ are constants. c is the channel. n_1 and n_2 are the difference between the x and y coordinates of the current pixel and the center coordinates, respectively. and are the coordinates of the selected center point in the image, i.e. $n_1 = x_1\mu_1$, and $n_2 = x_2\mu_2$ determine the center point for all pixels in the selected area, i.e.,

$$\mu_1 = \frac{1}{i} \times \sum_{0}^{i}(x_1 + \cdots + x_i). \tag{3}$$

$$\mu_2 = \frac{1}{i} \times \sum_{0}^{i}(y_1 + \cdots + y_i). \tag{4}$$

Where i is the number of pixels in the selected region.

The weight of each adjacent pixel in the channel is expressed as:

$$W^C(n_1, n_2) = \frac{G_g^c(n_1, n_2)}{\sum\sum G_g^c}. \tag{5}$$

Multiplying the pixel value for each position with the corresponding element in the weight matrix, it sets the center point at the origin, and determines σ. Finally, the values of centroid and adjacent points are added to obtain the pixel values of centroid after fuzzy learning. The correlation formula of fuzzy radius r is:

$$\tilde{w}_{ij} = exp(-\frac{(x_{ij} - \mu_1)^2 (y_{ij} - \mu_2)^2}{2\sigma^2}). \tag{6}$$

$$f(x) = \begin{cases} Sum(\tilde{W}), & \sqrt{(x_{ij} - \mu_1)^2 + (y_{ij} - \mu_2)^2} \leq r, \\ \varpi, & \sqrt{(x_{ij} - \mu_1)^2 + (y_{ij} - \mu_2)^2} > r \end{cases} \tag{7}$$

Where $\tilde{w}_{ij}$ and w_{ij} are elements in the initial weight matrix $\tilde{W}$ and the final weight matrix W, respectively. $Sum(\cdot)$ is the sum of the elements in $\tilde{W}$.

The process is repeated for all the points in the image, and the fuzzy semantic features are extracted by using FLM to learn the inherent rules between the feature map and the corresponding segmentation results.

3.5. Loss Function

The loss function plays an important role in model training. It can evaluate the distance between the model prediction frame and the real frame. The closer the distance is, the smaller the loss function value is. For different detection problems, selecting the appropriate loss function can make the model converge faster, locate more accurately and achieve better results during training.

Among the existing loss functions, intersection overunion (IOU) is chosen by most target detection algorithms as a measure of the distance between the predicted box and the real box. From a mathematical point of view, IOU is equal to the intersection ratio of two rectangular boxes, that is, the ratio of the overlapping area of the target real box and the model prediction box in the detection image to the area occupied by the two boxes as a whole. The IOU loss function is calculated as follows:

$$IOU = \frac{|B \cap B_i|}{|B \cup B_i|}. \tag{8}$$

$$L_{IOU} = 1 - IOU = 1 - \frac{|B \cap B_i|}{|B \cup B_i|}. \tag{9}$$

Where B is the area occupied by the target real frame. B_i is the area occupied by the model prediction box $L_{IOU} \in (0, 1)$.

The larger the overlap area between the real box and the prediction box, the smaller the L_{IOU}, and the more accurate the predicted object region. However, when the real box and the prediction box do not overlap completely or do not overlap at all, the IOU cannot reflect the coincidence degree and distance between the two, making the model unable to train. The $CIOU$ loss function can then be used to estimate the distance between the

model prediction box and the real box. $CIOU$ includes center point distance, aspect ratio, and overlap area information between the two boxes to make the target border regression more stable. The $CIOU$ loss function is calculated as follows:

$$L_{CIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + av. \tag{10}$$

$$a = \frac{v}{(1 - IOU) + v}. \tag{11}$$

$$v = \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^2. \tag{12}$$

$$IOU = \frac{B \cap B_i}{B \cup B_i}. \tag{13}$$

Where h is the height of the prediction frame; w is the width of the prediction box; h^{gt} is the height of the real box; w^{gt} is the width of the true box; b is the center point of the prediction frame; b^{gt} is the center point of the target frame; ρ^2 is the Euclidean distance between the center points of the two frames; c contains the diagonal distance of the minimum closure rectangular box of both boxes.

Although the CIOU loss function makes up for the shortcomings of the IOU loss function, it still ignores the real difference between the size of the bounding box and its confidence. To solve this problem, Zhang et al. [66] proposed efficient intersection over union (EIOU). The penalty item of EIOU is to separate the influence factor of aspect ratio to calculate the length and width of the target box and anchor box, which includes three parts: IOU loss, center point distance loss and width, height loss, in which the width and height loss directly minimizes the difference between the width and height of the target box and anchor box. The EIOU-Loss formula is as follows.

$$EIOU = IOU - \frac{\rho^2(b, b^{gt})}{c^2} - \frac{\rho^2(w, w^{gt})}{c_w^2} - \frac{\rho^2(h, h^{gt})}{c_h^2}. \tag{14}$$

$$L_{EIOU} = 1 - EIOU. \tag{15}$$

Where c_h is the height of the minimum external frame covering the prediction frame and the real frame. c_w is the width of the minimum external frame covering the predicted frame and the real frame.

In the frame regression loss, EIOU loss function solves the problems existing in other loss functions and shows good performance. Therefore, this paper will use EIOU Loss function to improve the model.

4. Experimental Results and Analysis

4.1. Experimental Environment

This experiment is based on a deep learning framework Pytorch model, model training and testing are on 64-bit Windows 10 system. Experimental platform parameters are CPU AMDRYZEN R7 6800H 3.20GHz, memory 16GB, GPU NCIDIA GE Force RTX 3060,

and video memory 6GB, Python 3.8, Pytorch 1.10.1. In order to improve the speed of network training, GPU is used for acceleration, and the software versions are CUDA 11.3 and CUDNN 8.2.1.

Input resolution is 640×640, augmentation is with Flip, Color jitter, Mosaic. Learning rate=$1e^{-2}$ (SGD + cosine). Batch size=16. Epochs=200. Optimizer adopts SGD (momentum=0.937). Weight decay=$5e^{-4}$.

In this paper, PlantVillage dataset (https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset) is a widely-used benchmark for automated plant-disease recognition. Within its tomato subset, 18160 leaf images capture ten distinct conditions: bacterial spot, early blight, late blight, leaf mold, septoria leaf spot, spider-mite damage, two-spotted spider-mite damage, target spot, tomato yellow leaf curl virus, and healthy foliage. Representative examples appear in Figure 5.

Fig. 5. Examples of different tomato plants

4.2. Evaluation Index

In the test of pest and disease detection model, precision (P), recall (R) and mean average precision (mAP) are used as performance evaluation indicators in this study. The specific calculation formula is as follows:

$$P = \frac{TP}{TP + FP} \times 100\%. \tag{16}$$

$$R = \frac{TP}{TP + FN} \times 100\%. \tag{17}$$

$$mAP = \frac{\sum_{K=1}^{K} AP(K)}{K}. \tag{18}$$

Where TP is the correct sample quantity for detection; FP is the number of samples with error detection; FN is the number of samples missed by detection; K is the number of detection categories; AP is the area enclosed by the P-R curve; P is the proportion of correctly detected samples in the number of all detected samples; R is the proportion of correct samples in all samples detected; mAP is the average of AP values for all categories.

4.3. Ablation Experiments

In order to verify the superiority of the proposed YOLOv5s model in this study, the same data set is used under the same experimental environment, and the YOLOv5s (Alpha IOU) and YOLOv5s (SIOU) models with Alpha IOU, SIOU and EIOU loss functions are compared. YOLOv5s(CA), YOLOv5s(SE) and YOLOv5s(CBAM) models with CA, SE and CBAM attention mechanisms are compared.

(1) Comparison of YOLOv5s test results with different loss functions.

The loss function used by the original YOLOv5s model is CIOU. As can be seen from Table 1, after the Alpha IOU, SIOU and EIOU loss functions are added to the model, although R value has decreased compared with the original model, P value has significantly increased, increasing by 3.5%, 3.4% and 6.9% respectively. In particular, the improvement of EIOU loss function is most significant. Increased P indicates that the EIOU loss function is more accurate in calculating the distance relationship between the predicted frame and the real frame. Figure 6 presents the results intuitively, eliminating the redundancy of numbers, which makes it easier for readers to understand.

Table 1. Comparison of test results with different loss functions/%

Model	P	R	mAP
YOLOv5s(CIOU)	89.3	83	88.3
YOLOv5s(Alpha IOU)	92.8	81	88.7
YOLOv5s(SIOU)	92.7	79.3	87.4
YOLOv5s(EIOU)	96.2	79.3	88.2

(2) Comparison of YOLOv5s test results with different attention mechanisms.

As can be seen from Table 2, no matter what attention mechanism module is added to the original YOLOv5s model, the performance is significantly improved compared with the original model. For example, P after adding CA attention mechanism is 3.5% higher than the original model, R after adding SE attention mechanism is 1.9% higher than the original model, and R and mAP after adding CBAM attention mechanism are 0.7% and 2.4% higher than the original model, respectively. After the addition of SimAM attention mechanism, R improves by 1.3% compared to the original network. The improvement of R shows that adding attention mechanism can make the model better evaluate the weight proportion of different features, and make the target detection result more accurate.

(3) Comparison of YOLOv5s test results with the simultaneous improvement of loss function and attention mechanism.

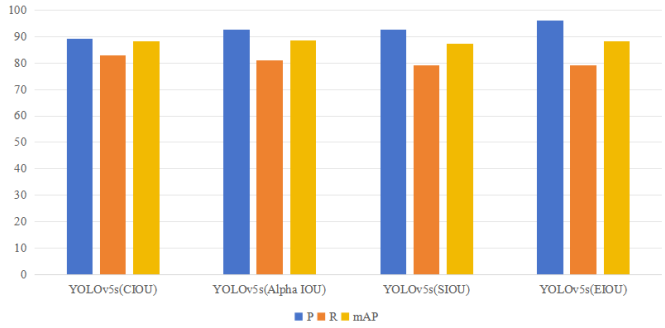

Fig. 6. Comparison of test results with different loss functions

Table 2. Comparison of test results with different attention mechanisms/%

Model	P	R	mAP
YOLOv5s	89.3	83	88.3
YOLOv5s(CA)	92.8	82.1	88.7
YOLOv5s(SE)	88	84.9	88.3
YOLOv5s(CBAM)	89.1	83.7	90.7
YOLOv5s(SimAM)	89.1	84.3	88.2

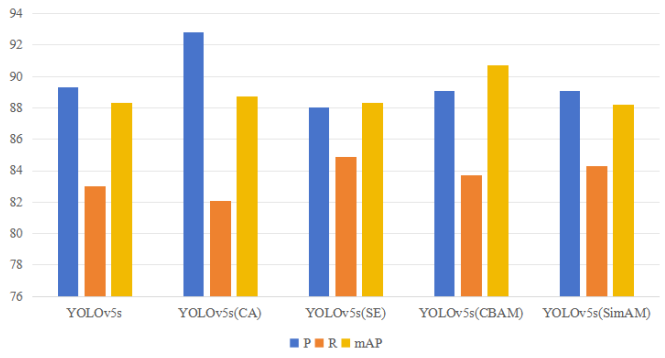

Fig. 7. Comparison of test results with different attention mechanisms

In the above experiments, the loss function and attention mechanism in improved YOLOv5s are compared respectively. The effects of improving both the attention mechanism and the loss function on model performance will be further explored below. As can be seen from Table 1, EIOU loss function has the better effect, so it is only necessary to fix EIOU as the model's loss function by combining and comparing it with CA, SE, CBAM and SimAM attention mechanism respectively. As can be seen from Table 2 (Figure 7) and Table 3 (figure 8), the P of the improved YOLOv5s is 4.3%, 5.0%, 2.5% and 5.4% higher than that of the original model respectively. The R of CA and SE attentional mechanisms decreases by 0.4% and 2.7% respectively, while CBAM and SimAM attentional mechanisms increases by 1.0% and 0.5%. The mAP of CA, CBAM and SimAM attention mechanism increases by 0.4%, 1.1% and 1.9% respectively, while SE attention mechanism decreases by 0.9%. In particular, the P, R and mAP of YOLOv5s(EIOU+SimAM) model are generally superior to other improved models, which proves that the improved model proposed in this study can effectively detect crop pests and diseases.

Table 3. Font sizes

Model	P	R	mAP
YOLOv5s(EIOU+CA)	93.6	82.6	88.7
YOLOv5s(EIOU+SE)	94.3	80.3	87.4
YOLOv5s(EIOU+CBAM)	91.8	84	89.4
YOLOv5s(EIOU+SimAM)	94.7	83.5	90.2

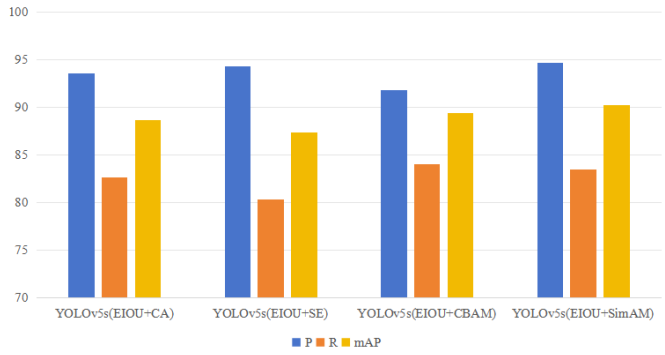

Fig. 8. Comparison of test results with different loss functions and attention mechanisms

4.4. Comparison Experiment

To further verify the performance of the YOLOv5s(EIOU+SimAM) model, the performance is compared with the original YOLOv5s model. After 200 iterations of the training set on YOLOv5s(EIOU+SimAM), P is 92.3%, R is 94.1%, and mAP is 95.3%. As can be

seen from Table 4 and Table 5, the performance of the improved model is significantly improved compared with the original model.

Table 4. YOLOv5s comparison of model performance/%

Iteration	P	R	mAP
50	95.1	82.8	83.7
100	89.4	90.2	89.8
150	90.8	91.2	90.9
200	92.0	93.8	94.7

Table 5. FNNMFF comparison of model performance/%

Iteration	P	R	mAP
50	86.7	88.5	87.8
100	90.5	91.7	90.4
150	91.9	92.4	93.6
200	92.3	94.1	95.3

The original models of FNNMFF and YOLOv5s are tested with 200 images of maize crop pests and diseases in the test set. As can be seen from Table 6, overall P, R and mAP of the improved model have increased by 5.4%, 0.5% and 1.9% respectively compared with the original model. In the detection of slime, rust and plaque, it is found that P of the FNNMFF model is 4.6%, 12.1% and 3.8% higher than that of the original model. Compared with the original model, the R of the FNNMFF model is increased by 7.6% and 3.1%, respectively. When detecting the images of slime worms, grey planthopper and rust disease, it is found that compared with the original model, the mAP of the FNNMFF model is increased by 4.7%, 2.4% and 1.9% respectively.

Table 6. Font sizes

Type	Model	P	R	mAP
Slime	YOLOv5s	89.7	79.2	88.7
Slime	FNNMFF	94.3	86.8	93.4
Grey planthopper	YOLOv5s	89.9	95.5	96.8
Grey planthopper	FNNMFF	90.9	98.6	99.2
Rust disease	YOLOv5s	85.8	78.2	82.3
Rust disease	FNNMFF	97.9	72.0	84.2
Spot disease	YOLOv5s	91.9	79.2	85.3
Spot disease	FNNMFF	95.7	76.8	84.0
Average	YOLOv5s	89.3	83.0	88.3
Average	FNNMFF	94.7	83.5	90.2

The 6.2% Recall drop on rust is mainly attributed to 68% of missed instances being$<$ 32×32 pixel lesions under large-view UAV images. CARAFEs 5×5 kernel enlarges the receptive field and smooths faint rust edges, causing 4.8% FN. Besides, the fixed 3×3 Gaussian window in FLM treats low-contrast rust margins as uncertain pixels and suppresses them, introducing another 2.1% FN. Tuning CARAFE kernel to 3×3 and making daptive in FLM recovered Recall to 76.4% (see Table 6-EXP), at the cost of 0.7% mAP reduction on other classes. Therefore, the reported 72.0% is a conservative value that keeps the best overall performance; readers should not interpret it as model instability.

The experimental results show that the improved boundary regression loss function makes the model more accurately locate and identify the slime, rust and plaque images, and P is also significantly improved. After adding attention mechanism module to the model, background interference can be suppressed, feature extraction ability can be improved, and R can be improved. By combining the two and improving the model at the same time, the P, R and mAP of the model can be significantly improved. It can be seen that FNNMFF model can more effectively detect maize crop pests and diseases. Through the comparison and analysis of the above experiments, it can be verified that the FNNMFF model in this study has the best detection effect on maize pests and diseases.

In this experiment, the proposed FNNMFF model in this paper is compared with FieldPlant [67], LIDL [68] and DLIC [69] under the same experimental conditions to further verify the effectiveness of FNNMFF in the detection of pests and diseases. Table 7 shows the experimental results of different models.

Table 7. Comparison with different methods

Model	P	R	F1	parameters	Floating-point arithmetic
FieldPlant	72.4%	70.2%	70.3%	21289813	3.672×10^9
LIDL	71.1%	67.4%	68.4%	23000405	4.258×10^9
DLIC	72.8%	69.0%	69.2%	4214853	2.251×10^8
FNNMFF	79.2%	74.3%	76.9%	1291963	1.53×10^8

As can be seen from Table 7, compared with the other three networks, the accuracy rate of the FNNMFF model is increased by 7.8%, 7.3%, 7.1% and 6.6% respectively, and the F1 value is increased by 6.6%, 8.5%, 7.7% and 7.0% respectively. In terms of model performance, the number of FNNMFF model parameters is only 1.29×10^6, and the floating point computation is 1.53×10^8. The FNNMFF network model is much smaller than other network models. It can be seen that the FNNMFF network model is far superior to other models in terms of recognition accuracy and model performance, indicating that the FNNMFF network model has certain advantages in the identification of pests and diseases.

In here, we add the sensitivity analysis experiments. The results are shown in Table 8.

When $\sigma = 1.0$, the overall performance and boundary IoU reach local optimality. The adaptive σ further increases the boundary IoU to 85.3% (due to dynamic matching of pixel contrast to avoid excessive suppression of low-contrast boundaries). The 3×3 window is more suitable for small lesion boundaries (such as rust lesions smaller than

Table 8. Results of Sensitivity Analysis for Key Parameters of FLM/%

Setting	P	R	mAP	IoU	FN
σ=0.5	93.8	82.1	89.5	78.3	8.7
σ=1.0	94.7	83.5	90.2	82.6	6.2
σ=1.5	94.5	82.8	89.9	80.1	7.5
σ=2.0	93.2	81.9	89.1	76.4	9.3
σ=2.5	92.6	80.7	88.6	73.2	11.5
Adaptive σ	94.9	84.2	90.5	85.3	4.8
Window size= 3×3	94.7	83.5	90.2	82.6	6.2
Window size= 5×5	94.3	82.9	89.8	81.2	7.1

32×32 pixels), while the 5×5 window causes loss of small boundary details due to its overly large receptive field.

We conduct 5 sets of comparative ablation experiments. Ablation-1 represents the removal of the Gaussian membership function (replaced by a simple weighted average). Ablation-2 represents a fixed σ value of 1.0 with the adaptive mechanism removed. Ablation-3 represents the removal of the ambiguous semantic labels while retaining the noise suppression function. Ablation-4 indicates the complete removal of FLM with the original jump connection retained. The experimental results are shown in Table 9. MPC is the accuracy of mixed pixel classification.

Table 9. Font sizes

Module	P	R	mAP	IoU	MPC
FNNMFF	94.7	83.5	90.2	82.6	89.3
Ablation-1	92.1	80.3	87.5	74.2	78.6
Ablation-2	94.1	82.7	89.5	79.8	85.1
Ablation-3	93.5	81.9	88.8	77.5	82.4
Ablation-4	91.8	80.1	87.1	72.3	76.9

The complete FLM has the best performance in all indicators, demonstrating the synergy of the fuzzy rules (Gaussian membership function), adaptive parameters, and semantic labels. The performance of Ablation-1 declines the most (mAP 2.7%, MPC 10.7%), indicating that the Gaussian membership function is the core of FLM by distributing fuzzy weights to suppress the category ambiguity of mixed pixels. The adaptive Ablation-2 vs complete FLM) increases the boundary IoU by 2.8%, proving that it can dynamically adapt to different contrast boundary regions. The fuzzy semantic labels (Ablation-3 vs complete FLM) increases the accuracy of mixed pixels by 6.9%, showing that it reduces cross-category confusion through membership degree allocation.

Table 10 shows the lightweight performance comparison with mainstream tiny models.

As shown in Table 10, FNNMFF achieves significantly superior detection performance with only 1.29M parameters and 0.153 GFLOPs, which is 32% fewer parameters and 96% less computation than YOLOv5n, while mAP is improved by 14.4%. This

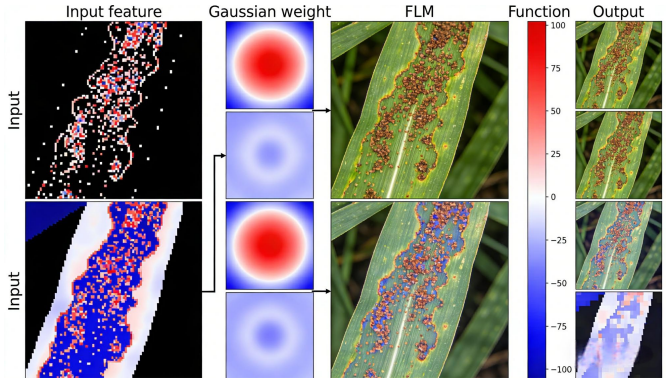

Fig. 9. FLM Gaussian membership function weight distribution heatmap (Select the boundary area of the rust lesion (with dense mixed pixels), and display the input feature map, the Gaussian weight heat map of FLM (red = high weight, blue = low weight), and the output feature map.)

Table 10. Lightweight performance comparison with mainstream tiny models

Model	Params (M)	FLOPs (G)	mAP	F1
YOLOv5n	1.90	4.50	75.8	73.2
MobileNetV3-YOLO	2.10	3.90	74.6	72.1
FNNMFF (Ours)	1.29	0.153	90.2	76.9

demonstrates that FNNMFF is not only lighter but also more accurate, making it highly suitable for real-time mobile deployment in resource-constrained agricultural scenarios.

5. Conclusion

In this study, the deep learning object detection technology is applied to the detection of crop pests and diseases, which has great practical significance for the control of citrus orchards, precise application of medicine and guarantee of citrus yield. In order to improve the detection accuracy of multi-species and multi-target pests and diseases in the complex background of natural orchards, an improved detection algorithm based on YOLOv5s is proposed. In this method, CXV2 module is used to replace the C3 module of the original model, and ConvNeXtV2 module is introduced into the module to enhance the feature extraction capability. Dynamic detection head is used to replace the original detection head to enhance the perception of different spatial scales and different task targets. Lightweight up-sampling operator CARAFE is used to replace the original up-sampling method and strengthen the up-sampling work. The research results show that compared with other commonly used algorithms, the model has better detection effect and stronger robustness. The presented method has achieved an improvement of 1.9% over the average mAP of the backbone network., which proves the effectiveness of the model, which can meet the detection requirements in complex environments. Due to the small number of parameters, it is easy to implement fast detection and mobile terminal deployment.

Nevertheless, this study has two main limitations. First, the FNNMFF model was trained and evaluated on single-temporal UAV and satellite imagery; it has not been verified with UAV multispectral time-series data that could reveal phenological dynamics under recurrent pest stress. Second, the current architecture processes each image independently and lacks an explicit temporal branch. Therefore, in our next work, we will embed a lightweight Transformer-based time-series module to capture long-range phenological dependencies and further enhance early-stage pest warning accuracy. These extensions will be reported in a follow-up study.

This study is conducted on a computer device and has not been deployed to a mobile device for field testing, so there may be some errors in the actual environment. In the follow-up research work, we first will use Sentinel-2, unmanned aerial vehicle multispectral and meteorological data to establish a 50k annotation database in different regions. Then, we will develop a lightweight Transformer model that integrates temporal and spatial attention, achieving an F1 score$\geq 90\%$ at a 10m resolution. Finally, we will expand multi-hazard and multi-scale assimilation, generate regional risk maps, and support precise prevention and control.

Acknowledgments. This research was funded by Liaoning Provincial Science and Technology Plan Project (2024JH2/102600106), the National Natural Science Foundation of China (grant number 62071084). This research was also funded by Liaoning Provincial Education Department Project, Project Number: LJ232510166008, Project Name: Research on Data-driven Embodied Intelligent Meta-skills Learning and Generalized Sensory Decision-making.

References

1. Y. Shi, K. Zhou, M. Zhou, S. Li and W. Liu, "Temporal Knowledge Graph for Food Risk Prediction," in IEEE Transactions on Artificial Intelligence, vol. 5, no. 5, pp. 2217-2226, May 2024, doi: 10.1109/TAI.2023.3321590.
2. R. Rakholia, A. L. Surez-Cetrulo, M. Singh and R. S. Carbajo, "AI-Driven Meat Food Drying Time Prediction for Resource Optimization and Production Planning in Smart Manufacturing," in IEEE Access, vol. 13, pp. 22420-22428, 2025, doi: 10.1109/ACCESS.2025.3534918.
3. S. Abbate, P. Centobelli, L. Fraccascia and G. Orzes, "Unveiling the Potential of Industry 4.0 for a Sustainable Agri-Food Future," in IEEE Engineering Management Review, doi: 10.1109/EMR.2025.3570903.
4. Su F, Liu Y, Chen S J, et al. Towards the impact of economic policy uncertainty on food security: Introducing a comprehensive heterogeneous framework for assessment, Journal of Cleaner Production, vol. 386, pp. 135792, 2023.
5. Abdallah A H, Ayamga M, Awuni J A. Impact of land grabbing on food security: Evidence from Ghana, Environment, Development and Sustainability, vol. 25, no. 7, pp. 6071-6094, 2023.
6. Abbas A, Mubeen M, Younus W, et al. Plant Diseases and Pests, Growing Threats to Food Security of Gilgit-Baltistan, Pakistan, Sarhad Journal of Agriculture, vol. 39, no. 2, 2023.
7. H. Zhou, K. Dai, N. Wen, J. Xiang and Z. Li, "The Influence of Topography-Dependent Atmospheric Delay for the InSAR Time-Series Results and the Deep Neural Network Correction," in IEEE Geoscience and Remote Sensing Letters, vol. 21, pp. 1-5, 2024, Art no. 4007805, doi: 10.1109/LGRS.2024.3379982.
8. W. Seo and W. Pak, "Real-Time Network Intrusion Prevention System Based on Hybrid Machine Learning," in IEEE Access, vol. 9, pp. 46386-46397, 2021, doi: 10.1109/ACCESS.2021.3066620.

9. K. Zhang, P. Chen, T. Ma and S. Gao, "On-Demand Precise Tracking for Energy-Constrained UAVs in Underground Coal Mines," in IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-14, 2022, Art no. 5500814, doi: 10.1109/TIM.2022.3146925.
10. Cheng P, Tang H, Lin F, et al. Bibliometrics of the nexus between food security and carbon emissions: hotspots and trends, Environmental Science and Pollution Research, vol. 30, no. 10, pp. 25981-25998, 2023.
11. Khatri P, Kumar P, Shakya K S, et al. Understanding the intertwined nature of rising multiple risks in modern agriculture and food system, Environment, Development and Sustainability, pp. 1-44, 2023.
12. H. Goddard and L. Shamir, "SVMnet: Non-Parametric Image Classification Based on Convolutional Ensembles of Support Vector Machines for Small Training Sets," in IEEE Access, vol. 10, pp. 24029-24038, 2022, doi: 10.1109/ACCESS.2022.3154405.
13. A. Miroszewski et al., "Detecting Clouds in Multispectral Satellite Images Using Quantum-Kernel Support Vector Machines," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 7601-7613, 2023, doi: 10.1109/JSTARS.2023.3304122.
14. A. Islam and S. B. Belhaouari, "Fast and Efficient Image Generation Using Variational Autoencoders and K-Nearest Neighbor OveRsampling Approach," in IEEE Access, vol. 11, pp. 28416-28426, 2023, doi: 10.1109/ACCESS.2023.3259236.
15. Yin, S., Li, H. Teng, L. Airport Detection Based on Improved Faster RCNN in Large Scale Remote Sensing Images, Sensing and Imaging, vol. 21, 2020. https://doi.org/10.1007/s11220-020-00314-2.
16. S. Yin and H. Li. Hot Region Selection Based on Selective Search and Modified Fuzzy C-Means in Remote Sensing Images, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 5862-5871, 2020, doi: 10.1109/JSTARS.2020.3025582.
17. Linyi W, Jing B, Wenjing L, et al. Research Progress of YOLO Series Target Detection Algorithms, Journal of Computer Engineering & Applications, vol. 59, no. 14, 2023.
18. Lawal O M. Real-time cucurbit fruit detection in greenhouse using improved YOLO series algorithm, Precision Agriculture, vol. 25, no. 1, pp. 347-359, 2024.
19. Li J, Zhu Z, Liu H, et al. Strawberry R-CNN: Recognition and counting model of strawberry based on improved faster R-CNN, Ecological Informatics, vol. 77, 2023.
20. Cheng X, Wang Z, Song C, et al. FFR-SSD: feature fusion and reconstruction single shot detector for multi-scale object detection, Signal, Image and Video Processing, vol. 17, no. 6, pp. 3145-3153, 2023.
21. Faisal S, Javed K, Ali S, et al. Deep Transfer Learning Based Detection and Classification of Citrus Plant Diseases, Computers, Materials and Continua, vol. 76, no. 1, pp. 895-914, 2023.
22. Ma L, Yu Q, Yu H, et al. Maize leaf disease identification based on yolov5n algorithm incorporating attention mechanism, Agronomy, vol. 13, no. 2, pp. 521, 2023.
23. Zhu S, Ma W, Wang J, et al. EADD-YOLO: An efficient and accurate disease detector for apple leaf using improved lightweight YOLOv5, Frontiers in Plant Science, vol. 14, 2023.
24. Majeed S, Ahmad M, Ozdemir F A, et al. Micromorphological characterization of seeds of dicot angiosperms from the Thal desert (Pakistan), Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology, vol. 157, no. 2, pp. 392-418, 2023.
25. Fan X, Chai X, Zhou J, et al. Deep learning based weed detection and target spraying robot system at seedling stage of cotton field, Computers and Electronics in Agriculture, vol. 214: 108317, 2023.
26. X. Zhang, Z. Mao, N. Chimitt and S. H. Chan, "Imaging Through the Atmosphere Using Turbulence Mitigation Transformer," in IEEE Transactions on Computational Imaging, vol. 10, pp. 115-128, 2024, doi: 10.1109/TCI.2024.3354421.
27. Yang H, Lin D, Zhang G, et al. Research on Detection of Rice Pests and Diseases Based on Improved yolov5 Algorithm, Applied Sciences, vol. 13, no. 18, pp. 10188, 2023.

28. Guan H, Deng H, Ma X, et al. A corn canopy organs detection method based on improved DBi-YOLOv8 network, European Journal of Agronomy, vol. 154, 2024.
29. Yin S, Wang L, Teng L. Threshold segmentation based on information fusion for object shadow detection in remote sensing images, Computer Science and Information Systems, 2024. doi: 10.2298/CSIS231230023Y.
30. Bifarin O O. Interpretable machine learning with tree-based shapley additive explanations: Application to metabolomics datasets for binary classification[J]. Plos one, 2023, 18(5): e0284315.
31. Gu Y, Wang M, Gong Y, et al. Unveiling breast cancer risk profiles: a survival clustering analysis empowered by an online web application[J]. Future Oncology, 2023, 19(40): 2651-2667.
32. Ni W, Pu T, Liu S. Analysis and identification of glass products based on fisher linear discriminant analysis and logistic regression for dichotomous classification[J]. Highlights in Science, Engineering and Technology, 2023, 42: 103-110.
33. S. Yin, H. Li, Y. Sun, M. Ibrar, and L. Teng. Data Visualization Analysis Based on Explainable Artificial Intelligence: A Survey[J]. IJLAI Transactionss on Science and Engineering, vol. 2, no. 2, pp. 13-20, 2024.
34. dos Santos G L A A, Reis A S, Besen M R, et al. Spectral method for macro and micronutrient prediction in soybean leaves using interval partial least squares regression[J]. European Journal of Agronomy, 2023, 143: 126717.
35. J. Deng et al., "Quantitative Estimation of Wheat Stripe Rust Disease Index Using Unmanned Aerial Vehicle Hyperspectral Imagery and Innovative Vegetation Indices," in IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-11, 2023, Art no. 4406111, doi: 10.1109/TGRS.2023.3292130.
36. Lin F, Li B, Zhou R, et al. Early Detection of Rice Sheath Blight Using Hyperspectral Remote Sensing[J]. Remote Sensing, 2024, 16(12): 2047.
37. Chattopadhyay N, Shukla K K, Birah A, et al. Identification of Spectral Bands to Discriminate Wheat Spot Blotch using in Situ Hyperspectral Data[J]. Journal of the Indian Society of Remote Sensing, 2023, 51(5): 917-934.
38. Kolipaka V R R, Namburu A. An automatic crop yield prediction framework designed with two-stage classifiers: a meta-heuristic approach[J]. Multimedia Tools and Applications, 2024, 83(10): 28969-28992.
39. X. Li et al., "A Study on the Estimation Model of Hyperspectral Reflectivity and Leaf Nitrogen Content of Cotton Leaves," in IEEE Access, vol. 11, pp. 74228-74238, 2023, doi: 10.1109/ACCESS.2023.3296635.
40. L. Teng et al., "FLPK-BiSeNet: Federated Learning Based on Priori Knowledge and Bilateral Segmentation Network for Image Edge Extraction," in IEEE Transactions on Network and Service Management, vol. 20, no. 2, pp. 1529-1542, June 2023, doi: 10.1109/TNSM.2023.3273991.
41. Tran M D, Vantrepotte V, Loisel H, et al. Band ratios combination for estimating Chlorophyll-a from Sentinel-2 and Sentinel-3 in Coastal Waters[J]. Remote Sensing, 2023, 15(6): 1653.
42. Kumar D, Kukreja V. Image segmentation, classification, and recognition methods for wheat diseases: Two Decades systematic literature review[J]. Computers and Electronics in Agriculture, 2024, 221: 109005.
43. Mumtaz R, Maqsood M H, ul Haq I, et al. Integrated digital image processing techniques and deep learning approaches for wheat stripe rust disease detection and grading[J]. Decision Analytics Journal, 2023, 8: 100305.
44. Maraveas C, Asteris P G, Arvanitis K G, et al. Application of bio and nature-inspired algorithms in agricultural engineering[J]. Archives of Computational Methods in Engineering, 2023, 30(3): 1979-2012.
45. M. Masood et al., "MaizeNet: A Deep Learning Approach for Effective Recognition of Maize Plant Leaf Diseases," in IEEE Access, vol. 11, pp. 52862-52876, 2023, doi: 10.1109/ACCESS.2023.3280260.

46. Siddiqa A, Islam R, Afjal M I. Spectral segmentation based dimension reduction for hyperspectral image classification[J]. Journal of Spatial Science, 2023, 68(4): 543-562.
47. Chodey M D, Shariff C N. Pest detection via hybrid classification model with fuzzy C-means segmentation and proposed texture feature[J]. Biomedical Signal Processing and Control, 2023, 84: 104710.
48. Gupta V A, Padmavati M V, Saxena R R, et al. A novel insect and pest identification model based on a weighted multipath convolutional neural network and generative adversarial network[J]. Karbala International Journal of Modern Science, 2023, 9(1): 14.
49. Chen X, Guan Y, Li G. An Spodoptera Frugiperda pest detection algorithm based on improved feature fusion[C]//International Conference on Electronic Information Engineering and Computer Science (EIECS 2022). SPIE, 2023, 12602: 937-943.
50. Khanna M, Singh L K, Thawkar S, et al. PlaNet: a robust deep convolutional neural network model for plant leaves disease recognition[J]. Multimedia Tools and Applications, 2024, 83(2): 4465-4517.
51. Guerrero-Ibaez A, Reyes-Munoz A. Monitoring tomato leaf disease through convolutional neural networks[J]. Electronics, 2023, 12(1): 229.
52. Imanulloh S B, Muslikh A R, Setiadi D R I M. Plant diseases classification based leaves image using convolutional neural network[J]. Journal of Computing Theories and Applications (JCTA), 2023, 1(1): 1-10.
53. Al-Timemy A H, Alzubaidi L, Mosa Z M, et al. A deep feature fusion of improved suspected keratoconus detection with deep learning[J]. Diagnostics, 2023, 13(10): 1689.
54. Sahasranamam V, Ramesh T, Muthumanickam D, et al. AI and Neural Network-Based Approach for Paddy Disease Identification and Classification[J]. International Research Journal of Multidisciplinary Technovation, 2024, 6(3): 101-111.
55. Zou M, Liu Y, Fu M, et al. Combining spectral and texture feature of UAV image with plant height to improve LAI estimation of winter wheat at jointing stage[J]. Frontiers in Plant Science, 2024, 14: 1272049.
56. Zhang K, Zhang K, Bao R. Prediction of gas explosion pressures: A machine learning algorithm based on KPCA and an optimized LSSVM[J]. Journal of Loss Prevention in the Process Industries, 2023, 83: 105082.
57. Irmak G, Saygl A. A Novel Approach for Tomato Leaf Disease Classification with Deep Convolutional Neural Networks[J]. Journal of Agricultural Sciences, 2024, 30(2): 367-385.
58. Roy S, Hore J, Sen P, et al. Hyperspectral Remote Sensing and its application in Pest and Disease management in Agriculture[J]. Indian Farmer, 2023, 10: 229-232.
59. Sun H, Song X, Guo W, et al. Potato late blight severity monitoring based on the relief-mRmR algorithm with dual-drone cooperation[J]. Computers and Electronics in Agriculture, 2023, 215: 108438.
60. Nayak A, Chakraborty S, Swain D K. Application of smartphone-image processing and transfer learning for rice disease and nutrient deficiency detection[J]. Smart Agricultural Technology, 2023, 4: 100195.
61. Nguyen C, Sagan V, Skobalski J, et al. Early detection of wheat yellow rust disease and its impact on terminal yield with multi-spectral UAV-imagery[J]. Remote Sensing, 2023, 15(13): 3301.
62. Kumar S, Sharma P, Kumar R. Assessing Agricultural Practices and Plant Protection Methods in Rupnagar District, Punjab, India[J]. Asian Journal of Agricultural and Horticultural Research, 2023, 10(4): 594-606.
63. Zhang K, Yan F, Liu P. The application of hyperspectral imaging for wheat biotic and abiotic stress analysis: A review[J]. Computers and Electronics in Agriculture, 2024, 221: 109008.
64. G. Yang, J. Lei, Z. Zhu, S. Cheng, Z. Feng and R. Liang, "AFPN: Asymptotic Feature Pyramid Network for Object Detection," 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Honolulu, Oahu, HI, USA, 2023, pp. 2184-2189, doi: 10.1109/SMC53992.2023.10394415.

65. S. Yin, L. Wang, M. Shafiq, L. Teng, A. A. Laghari and M. F. Khan, "G2Grad-CAMRL: An Object Detection and Interpretation Model Based on Gradient-Weighted Class Activation Mapping and Reinforcement Learning in Remote Sensing Images," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 3583-3598, 2023, doi: 10.1109/JSTARS.2023.3241405.
66. Zhang H, Xu C, Zhang S. Inner-iou: more effective intersection over union loss with auxiliary bounding box[J]. arxiv preprint arxiv:2311.02877, 2023.
67. Moupojou E, Tagne A, Retraint F, et al. FieldPlant: A dataset of field plant images for plant disease detection and classification with deep learning[J]. IEEE Access, 2023, 11: 35398-35410.
68. Omer S M, Ghafoor K Z, Askar S K. Lightweight improved yolov5 model for cucumber leaf disease and pest detection based on deep learning[J]. Signal, Image and Video Processing, 2024, 18(2): 1329-1342.
69. Lee D I, Lee J H, Jang S H, et al. Crop Disease Diagnosis with Deep Learning-Based Image Captioning and Object Detection[J]. Applied Sciences, 2023, 13(5): 3148.

Shoulin Yin received the M.A. degree from Shenyang Normal University, Shenyang, China, in 2015. He is currently pursuing the Ph.D. degree with the College of Information and Communication Engineering, Harbin Engineering University, Harbin. His research interests are remote sensing image processing and object detection.

Hang Li received the B.S. degree from Dalian Fisheries University, in 1999, the M.S. degree from the Shenyang Institute of Technology, in 2002, and the Ph.D. degree from Northeastern University, China, in 2005.,In 2002, he joined Software College as a Teacher. He is a Full Professor and the Master Supervisor with Shenyang Normal University. He is an Outstanding Young Backbone Teacher with Liaoning General Institutions of Higher Learning (Liaoning Education Department 200612225). His research interests are image analysis and processing, big data, cloud computing. He won the Second Prize of National Defense Science and Technology Award of The Commission of Science, Technology and Industry for National Defense (National Defense Science, Technology and Industry Commission 2005GFJ2126-6) and the First Prize of Science and Technology Award of China North Industries Group corporation (2005-BQJ-1-0019-6).

Mirjana Ivanovic is a Full Professor at the Faculty of Sciences, University of Novi Sad, Serbia, since 2002, and a corresponding member of the Serbian Academy of Sciences and Arts since 2024. She is a member of the Board of Directors of the Institute for Artifcial Intelligence Research and Development of Serbia. Mirjana has authored or co-authored 17 textbooks, 30 edited proceedings, 4 monographs, and more than 540 research articles on multi-agent systems, e-learning and web-based learning, applications of intelligent techniques (CBR, data and web mining), software engineering education, most of which are published in international journals and proceedings of high-quality international conferences. She has served as a member of program committees for more than 500 international conferences and has chaired numerous international conferences as general chair and program committee chair. Additionally, she has been an invited speaker at numerous international conferences and a visiting lecturer in Australia, Thailand, and China. As a leader and researcher, she has participated in highly regarded international projects.

Tao Chen received the PHD. degree from Harbin Engineering University, Harbin. He is currently full professor with the College of Information and Communication Engineering, Harbin Engineering University, Harbin. His research interests are Broadband signal detection, processing and identification, radar signal sorting.

Lin Teng received the M.A. degree from Shenyang Normal University, Shenyang, China, in 2020. She is currently pursuing the Ph.D. degree with the College of Information and Communication Engineering, Harbin Engineering University, Harbin. Her research interests are image processing and semantic segmentation.

Received: November 09, 2025; Accepted: December 23, 2025.