Fuzzy EfficientDet: An Approach for Precise Detection of Larch Infestation Severity in UAV Imagery Under Dynamic Environmental Conditions

Shuo Yang [ID], Jingbin Li [ID], Yang Li [ID], Jing Nie [ID], Yujie Qiao [ID], and Sezai Ercisli [ID]

Abstract—In this article, a novel deep learning framework, fuzzy EfficientDet, is proposed to address the challenge of accurately detecting larch infested by Coleophora laricella pests in UAV imagery, where the key innovation is the incorporation of the fuzzy spatial attention mechanism (FSAM), which can effectively deal with the problem of model uncertainty due to the complexity of environmental transformations and image features. First, this study designs and implements a global–local squeeze-and-excitation module, which profoundly integrates global and local feature information, realizes the dynamic adaptation of the importance of feature channels in the EfficientNet, and thus improves the overall feature expression efficiency of the network. Second, this study constructed a dense Bi-FPN architecture, which adds a dense connection structure to the original Bi-FPN to enhance the modeling accuracy for small targets and long-range spatial dependencies. Finally, this study develops the FSAM, which can effectively mitigate the unstable performance of the EfficientDet in the face of image feature fluctuations triggered by changes in lighting conditions and seasonal effects. Experiments demonstrate that the proposed fuzzy EfficientDet model shows superior performance compared to the traditional SSD, Faster R-CNN, YOLO V5, and the unimproved EfficientDet target detection method on the Swedish Forest Agency (2021) dataset, with its mAP as high as 94.29%. This result demonstrates that fuzzy EfficientDet provides an efficient and reliable solution when dealing with the task of target detection in UAV images, especially in dealing with environmental uncertainty and complex feature extraction.

Index Terms—Attention mechanism, efficientdet, fuzzy theory, pest inspection, UAV image.

I. INTRODUCTION

COLEOPHORA laricella [1] belongs to the Lepidoptera family, mainly distributed in Europe, America, Asia, the

northern temperate zone, and the subboreal zone. The main species of the host tree for the larch is one of the common larch pests. Larvae feeding on larch leaf flesh, forest trees suffering from pests can lead to foliage fading, tree weakness, serious damage will cause a number of forest area large-scale deaths, a yellow forest stand, seriously affecting the growth of forest trees and ecological health [2]. Therefore, timely detection and evaluation of Coleoptera laricella pests and diseases of larch are of great practical significance.

There are mainly artificial ground surveys, low-altitude UAV image detection, and remote sensing satellite monitoring methods for larch infestation detection. The artificial ground survey requires researchers to go deep into the forest area to contact the trees directly to observe and record the detailed pest situation at close range with high accuracy. Still, its detection efficiency is low, limited by the terrain and landscape, and the monitoring cost is high for the large forest area. Remote sensing satellite monitoring covers a wide range of areas and can realize macromonitoring of ultra-large areas. Still, the resolution of remote sensing satellites is relatively low, with limited ability to identify the details of pests and diseases in a small range or at the individual level, and is easily affected by cloud cover. The acquired data have limitations in terms of timeliness and continuity. In contrast, low-altitude UAV imagery can provide rich near-surface microscopic information [3] and is not limited by terrain complexity, which is one of the increasingly critical intelligent tools in modern forestry [4].

With the continuous development of deep learning technology, deep learning methods show some apparent advantages in the fields of image detection [5], [6], [7], [8], image semantic information processing [9], [10], [11], [12], and remote sensing image processing [13], [14], [15], [16], so the forest pest detection and assessment algorithms based on deep learning are getting more and more attention from scholars.

Widely used target detection algorithms in computer vision are mainly categorized into regression-based one-stage methods and classification-based two-stage methods [17], [18], [19], [20]. The YOLO series [21] and SSD [22] are the most representative algorithms for one-stage methods. The most representative algorithms for two-stage methods are the region-CNN series, such as R-CNN [23], fast R-CNN [24], and faster R-CNN [25]. Compared to two-stage models that need to generate candidate regions through RPN and subsequently refine classification and localization, one-stage models can predict categories directly

from images. Eliminates the candidate region generation and subsequent processing stages, so one-stage algorithms have faster inference speeds. Their end-to-end training approach is more concise and efficient, which can reduce the consumption of computational resources and model deployment complexity [26]. Therefore, researchers usually choose one-stage algorithms to perform the task of forest pest and disease detection. Chen et al. [27] utilized a random sampling scheme combined with UAV high-resolution images to improve YOLOv5 and enhance pine wilt detection accuracy. Liu et al. [28] proposed a YOLOv7-KCC model for tree species classification based on UAV RGB remote sensing images, which can effectively classify windbreak forest species. Lv et al. [29] proposed an improved YOLOv3 model combined with image enhancement techniques to effectively detect crop pests in real agricultural environments. Wu et al. [30] proposed a YOLOv3-based target detection algorithm for pine wilt disease, realizing early diagnosis of pine wilt disease in the field. Xue et al. [31] incorporated self-attention and convolutional integration as well as the convolutional block attention module into YOLOv5 and proposed the YOLO-Tea model, which has an improvement of 0.3% to 15.0% compared to YOLOv5. Yu et al. [32] proposed a shallow weighted feature enhancement network based on small target extension to detect pine wild diseased trees, which has significant advantages in detecting pine wild diseased trees, in response to the problem that the shallow feature layer of existing detection algorithms has insufficient feature extraction ability for small diseased trees. The above studies have demonstrated the effectiveness of one-stage algorithms in forest pest and disease detection, although such methods can satisfy the need for rapid screening of pests and diseases in large areas to a certain extent due to their fast response and real-time performance, however, in the complex and changing forest environments, these algorithms still need much improvement in terms of detection accuracy, especially the ability to recognize small targets and the efficiency of using information in context.

EfficientDet [33], as an innovator of a two-stage detection framework, combines the advantages of one-stage and two-stage detection frameworks in its design principle and structural characteristics, providing a new solution to overcome the above limitations. Compared with the traditional YOLO and SSD models, EfficientDet is favored by many scholars because of its innovative FPN structure and multilevel feature fusion mechanism, which optimizes the recognition performance of targets with significant scale differences, and its excellent network structure balances the speed and accuracy of detection well [10]. Jia et al. [34] proposed a new marine life object detection model based on EfficientDet to achieve fast and accurate detection of marine life in complex marine environments. Li et al. [35] proposed a multimodal EfficientDet with multiscale CapsNet to improve the detection performance of the model in real-world scenes with complex textures and occlusions. Zhu et al. [36] improved the detection accuracy of olive fruit ripeness in the orchard environment by reasonably embedding the CBAM module in EfficientDet. Zhuang et al. [37] proposed an image recognition method based on the combination of EfficientDet and generative adversarial network (GAN), which enabled the recognition model to obtain higher recognition speed and recognition accuracy in military object detection tasks. However, Efficient-Det also has some limitations in performing the task of detection and evaluation of Coleoptera laricella pests and diseases in larch:

1) Due to the use of depthwise convolution in its backbone network, EfficientNet reduces the number of parameters in the model and improves the computational efficiency of the model, but this will reduce a certain amount of feature representation ability when dealing with UAV images of larch Coleophora laricella infestation, the model is challenging to capture the complex texture details in the image, which makes it difficult for the model to evaluate the infestation level. In addition, depthwise convolution performs 1-D convolution for each input channel separately and independently, resulting in the loss of cross-channel interaction information.

2) Larch UAV image scenes are more complex. Their detection and evaluation tasks require the target detection model to have the ability to understand and deal with long-range spatial dependencies, as well as the ability to understand contextual information. In contrast, the spatial resolution of the feature maps decreases as the layers in the Bi-FPN increase in exchange for richer global contextual information [38]. However, the low resolution of high-level feature maps means that they are less able to capture fine details or precise positional relationships between distant objects.

3) Larch UAV images in outdoor environments exhibit significant visual differences due to factors such as seasonal changes and diversity of lighting conditions, and these uncertain feature variations pose a challenge to the detection robustness of the EfficientDet model [39], which in turn negatively affects its accuracy on the larch Coleophora laricella pest detection task, leading to suboptimal model performance.

The above shortcomings have inspired scholars to explore the possibilities of utilizing fuzzy theory. Yang et al. [40] designed a deep convolutional neural network incorporating a fuzzy attention mechanism, effectively improving the medical image segmentation accuracy. Nan et al. [41] proposed an airway segmentation method, which enhances the continuity of the segmentation utilizing a fuzzy attention neural network and a combined loss function. Xiao et al. [42], in a real driving scenario, combined fuzzy and attention mechanisms to deep neural networks to create a fuzzy deep attention network (FDAN), which improves the model's performance in an uncertain environment. Chong et al. [43] proposed a multiscale fuzzy bidomain attention network by combining fuzzy pattern recognition methods and deep learning methods, which improved the model's ability to distinguish similar objects in urban remote sensing images. In summary, fuzzy theory shows excellent potential in the field of computer vision and image processing due to its natural advantages in dealing with imprecise and ambiguous information as well as multiscale feature fusion, and the theory can provide a good and feasible idea for the detection of insect-infested larch under UAV images.
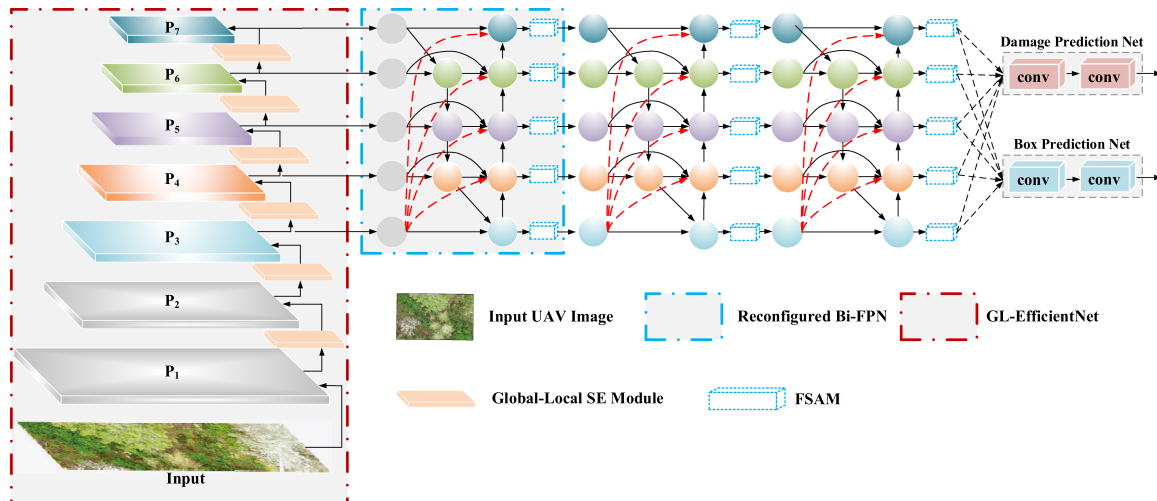
Fig. 1. Proposed fuzzy EfficientDet network architecture.

Given this, this study combines fuzzy theory and EfficientDet to propose a larch Coleophora laricella pest detection and assessment model called fuzzy EfficientDet. The main contributions of this study are as follows:

1) To improve the feature representation capability of EfficientNet as well as the problem of loss of cross-channel interaction information in depthwise convolution, this study combines the concept of global-local and proposes a new squeeze-and excitation module (SE module) [44], called global-local SE module, and implanted the module into EfficientNet. The module first applies pooling operations at different scales to the input feature maps. Then, the SE module generates the weight coefficients of the global and local features, which enhance the important features at specific scales.

2) To improve EfficientDet's modeling of long-range spatial dependencies, this study reconfigures the original Bi-FPN and introduces a densely connected version of the dense Bi-FPN structure. This improvement allows the model to more efficiently integrate contextual information and fully exploit the subtle local features embedded in the highest resolution layer, thus significantly improving the model's effectiveness in understanding and processing complex spatial layouts.

3) To further improve the detection performance of the model in the face of dynamic environmental changes such as seasonal changes and light conditions, this study constructed a fuzzy spatial attention mechanism (FSAM) based on fuzzy logic and embedded it into the dense Bi-FPN to build the dense fuzzy Bi-FPN structure. This design helps the model to maintain higher detection robustness and accuracy under uncertain environments, thus optimizing the detection of Coleophora laricella infestation in larch.

II. METHOD

EfficientDet, as an efficient target detection scheme, faced two main challenges when dealing with the task of detecting and evaluating larch infestation by Coleophora laricella: first, the images of larch forests captured by UAVs have highly complex textural features; second, uncertainties such as seasonal variations and differences in lighting conditions affect the detection accuracy. This study developed an improved model, fuzzy EfficientDet (see Fig. 1), to address the above problems.

This study first innovatively integrates the global–local SE module into EfficientNet to form the GL-EfficientNet architecture, in order to improve the model's deep representation of the input features and recover the lack of cross-channel information interactions caused by the depthwise convolution. The global–local SE module aims to improve the detection of key features through the combined weighting of global and local information. Furthermore, this study proposes a novel feature fusion structure called dense fuzzy Bi-FPN, which is initially designed to optimize the detection performance of tiny targets and improve the model's adaptability and detection accuracy under diverse and dynamic environmental conditions (e.g., seasonal shifts and light variations), especially in the detection and assessment of larch Coleophora laricella infestation showed significant robustness and accuracy gains.

In the constructed detection process, the UAV images are first extracted by multiscale feature extraction through GL-EfficientNet. Then, the dense fuzzy Bi-FPN structure is used to aggregate and refine these multiscale features. After that, the model performs synchronous category classification and bounding box localization regression operations. Finally, the NMS algorithm is used to eliminate duplicate detection results to ensure that only the best detection boxes are retained for each target category output.

A. Global–Local SE Module

To achieve optimal model performance with limited computing resources, researchers typically use EfficientNet as the backbone network for EfficientDet [45], [46]. However, given the inherently shallow properties of EfficientNet, the architecture has limitations in capturing the global contextual information of
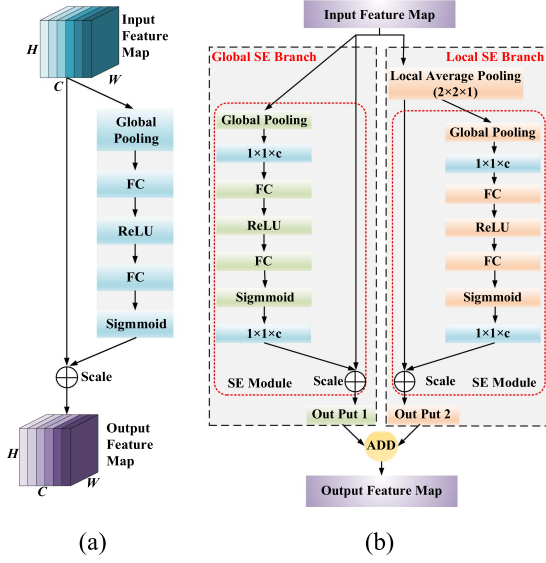
Fig. 2. Structure of the: (a) SE module; (b) global–local SE module.

the feature map. Moreover, since the model employs depthwise convolution to reduce the number of parameters and improve computational efficiency, this convolutional operation loses a certain amount of feature representation. Suppose the input feature map size is $H \times W \times C$, where H and W denote the height and width, respectively, and C represents the number of channels. For each channel c, depthwise convolution performs a convolution operation using a single-channel convolution kernel K_c of size $k_w \times k_h$ to generate a new feature map D_c:

$$D_c = \sum_{i=0}^{k_h-1} \sum_{j=0}^{k_w-1} K_c[i,j] * I_c[i:i+H-1, j:j+W-1] \tag{1}$$

where I_c is the 2-D matrix of the input feature map at the c channel, K_c is the corresponding convolution kernel, and $*$ denotes the convolution operation. Because each convolution kernel focuses on only a single channel during deep convolution, the information between different channels is not interacted with during the process. This results in the loss of cross-channel feature interaction information, reducing the ability of the model to capture complex texture details within the image.

To improve the EfficientNet feature representation capability as well as to ameliorate the problem of loss of cross-channel interaction information in depthwise convolution, this study adopts the concept of SE module as shown in Fig. 2(b). This module recovers and optimizes the correlation and importance allocation among feature channels, thus improving the quality of feature representation. However, the standard SE module regulates channel attention mainly based on the statistical information of global average pooling, ignoring the feature importance changes within the local region. To address this limitation, this study proposes a novel SE module called the global–local SE module, combining the global–local concept, whose structure is shown in Fig. 2(b).

The global–local SE module consists of two branches: the global SE branch and the Local SE branch, which are connected in parallel. The following describes the implementation details and computation of the global–local SE module. Suppose the input feature map is $F \in R^{W \times H \times C}$, where $W \times H$ is the spatial dimension size of the feature map and C is the number of feature data channels.

Global SE branch: For the global SE branch, the original input feature map F is directly fed into the standard SE module for processing. In this process, a global average pooling operation is first performed to obtain the global statistical information of each channel, which is calculated as follows:

$$Z_G = \text{GAP}(F) \tag{2}$$

where Z_G denotes the output feature map after global average pooling.

Then, the learning of channel attention weights is realized by two fully connected (FC) layers and activation function, which is calculated as follows:

$$S_G = \sigma(W_{12}(\delta(W_{11}(Z_G)))). \tag{3}$$

In (3), W_{11} and W_{12} denote the output feature maps after two FC layers, δ represent the ReLU function, σ represent the sigmoid function, and the above functions are used to generate the normalized channel weight vector S_G.

Applying these weights to the input feature map F yields the output of the global SE branch:

$$\text{Output}_G = F \odot S_G. \tag{4}$$

Local SE branch: In the Local SE branch, to capture the local spatial context information, the input feature map F is first processed using a local average pooling operation, and to ensure that the correlation between neighboring pixels can be preserved, the pooling window size is set to 2×2 and the step size to 1:

$$Z_L = LocalAvgPool_{2 \times 2, stride=1}(F). \tag{5}$$

Next, similar to the global SE branch, the local statistical features Z_L are input into the standard SE module to compute the channel attention weight vector S_L reflecting the local spatial dependencies:

$$S_L = \sigma(W_{22}(\delta(W_{21}(Z_L)))). \tag{6}$$

Based on the local channel attention weights S_L, the input feature maps are adjusted accordingly

$$\text{Output}_L = F \odot S_L. \tag{7}$$

Finally, the feature maps adapted by the global and local attention mechanisms are subjected to an element-by-element summation operation to integrate attentional information at different spatial scales:

$$\text{Output}_{\text{fuse}} = \text{Output}_G + \text{Output}_L. \tag{8}$$

B. GL-EfficientNet

In this study, the previously proposed global–local SE module is integrated into the back-end of the MBConv module in EfficientNet, and GL-EfficientNet is constructed to strengthen the

TABLE I
GL-EFFICIENTNET PARAMETERS

| Stage	Structure layer	Kernel size	Stride	Output size
Stem Convolution	Conv+BN+Swish	3×3	2	256×256
Stage 1	MBConvBlock	3×3	1	256×256
Global–Local SE 1	Global–Local SE module	/	/	256×256
Stage 2	MBConvBlock	3×3	2	128×128
Global–Local SE 2	Global–Local SE module	/	/	128×128
Stage 3	MBConvBlock	3×3	2	64×64
Global–Local SE 3	Global–Local SE module	/	/	64×64
Stage 4	MBConvBlock	3×3	2	32×32
Global–Local SE 4	Global–Local SE module	/	/	32×32
Stage 5	MBConvBlock	3×3	2	16×16
Head	Conv+ BN+GAP	1×1	2	/
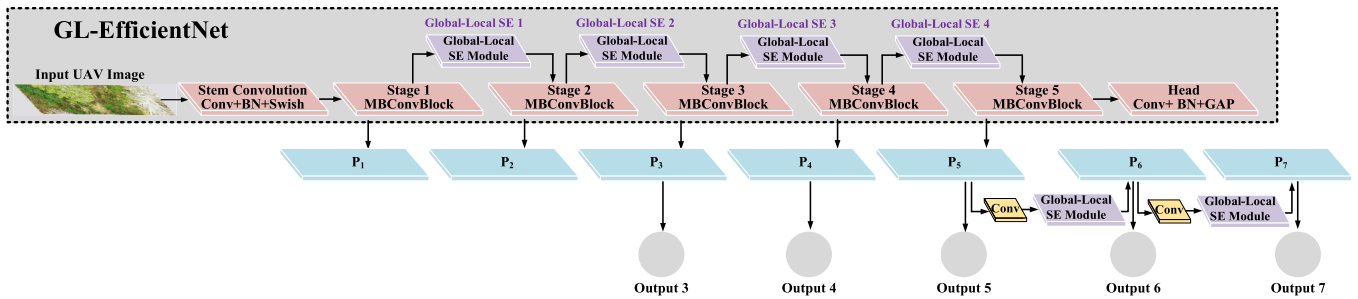
Fig. 3. Structure of the GL-EfficientNet.

model's ability to jointly model and express the spatial global and local information of the complex visual features. The structure diagram of the GL-EfficientNet network and the various layer parameters are shown in Table I and Fig. 3, respectively.

As shown in Table I and Fig. 3, in GL-EfficientNet, P1 to P5 represent the output feature map layers after five stages (stage1–stage5) of processing, respectively. A series of MBConv and global–local SE modules gradually generate these feature maps from the input images. P_6 and P_7 are the high-level feature representations obtained by downsampling P_5 successively and then processing them again by the global–local SE module.
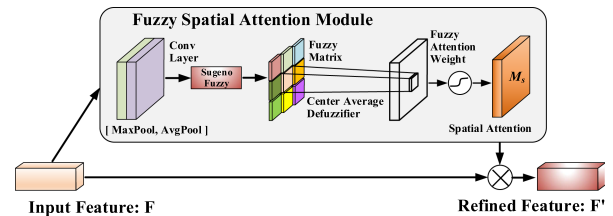
C. Dense Fuzzy Bi-FPN

Given that the UAV image acquisition process is subject to significant changes in environmental factors such as seasonal changes and lighting conditions [47], [48], resulting in high variability of color and morphological features of larch in different environments, which makes it difficult for the model to identify and detect larch infected by Coleophora laricella infestation. To address this problem, this study combines fuzzy logic to construct a FSAM, the structure of which is shown in Fig. 4.

The following illustrates the computational process of FSAM using the input feature mapping $X \in R^{W \times H \times K}$ as an example.

Fig. 4. Structure of FSAM.

First, average and maximum pooling are computed for X to obtain two new feature mappings, Z_{avg} and $Z_{\max}$, respectively:

$$Z_{\mathrm{avg}} = \frac{1}{WH} \sum_{i=1}^{W} \sum_{j=1}^{H} X_{i,j,k} \tag{9}$$

$$Z_{\max} = \max_{i \in [1,W], j \in [1,H]} X_{i,j,k}. \tag{10}$$

Then, Z_{avg} and $Z_{\max}$ are used as input variables to the fuzzy system, and a Sugeno fuzzy model [49], [50], [51] is applied to determine their affiliation with predefined fuzzy sets. In FSAM, four fuzzy sets are set in this study, which are light damage (A_l), high damage (A_h), healthy (A_hl), and other trees (A_o). For each pixel location (i, j) on the feature map, its corresponding average pooling and maximum pooling eigenvalues (x, y) are
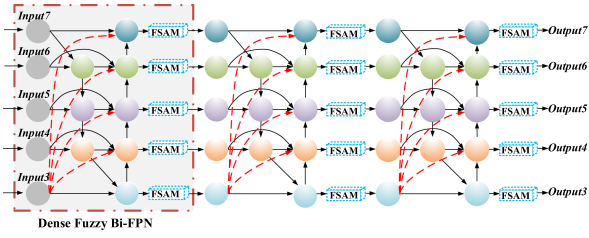
Fig. 5. Structure of dense fuzzy Bi-FPN network.

used as fuzzy inputs, and the triangular membership function in the Sugeno fuzzy model is applied to compute the degree of affiliation on each fuzzy set:

$$\begin{cases} \mu_{Al}(x,y) = \frac{1}{1+(x-c_l^x)^2+(y-c_l^y)^2} \\ \mu_{Ah}(x,y) = \frac{1}{1+(x-c_h^x)^2+(y-c_h^y)^2} \\ \mu_{Ahl}(x,y) = \frac{1}{1+(x-c_{hl}^x)^2+(y-c_{hl}^y)^2} \\ \mu_{Ao}(x,y) = \frac{1}{1+(x-c_o^x)^2+(y-c_o^y)^2} \end{cases} \quad (11)$$

where c_l^x, c_l^y, c_h^x, c_h^y, c_{hl}^x, c_{hl}^y, c_o^x, and c_o^y are the coordinates of the center point of each fuzzy set.

Generate spatial attention fuzzy matrix A by combining the affiliation of each fuzzy set:

$$A_{ij} = \sum_{m \in \{l,h,hl,o\}} w_m \cdot \mu_{Am}(x_i, y_j) \quad (12)$$

where w_m denotes the weight factor for category m, reflecting the relative importance of each category in the final attention weighting.

Subsequently, a central average deblurring method was used to convert the fuzzy attention matrix into a precise mapping of attention weights:

$$AttMap_{ij} = \frac{\sum_{m \in \{l,h,hl,o\}} w_m \cdot \mu_{A_m}(x_i, y_j) \cdot c_m}{\sum_{m \in \{l,h,hl,o\}} w_m \cdot \mu_{A_m}(x_i, y_j)} \quad (13)$$

where c_m denotes the actual attentional intensity value corresponding to category m.

Finally, the resulting precise attention weight mapping is multiplied elementwise with the original feature map X to obtain the FSAM-processed feature map X_{att}:

$$X_{att} = AttMap \odot X \quad (14)$$

where X_{att} will pay more attention to the features related to specific pest and disease categories, thus improving the robustness of the model for pest and disease detection under complex environmental and lighting conditions.

In addition, for the model to fully incorporate contextual information, this study constructs the dense Bi-FPN, which aims to fuse multiscale contextual information more efficiently, its structure is shown in Fig. 5. Different from the traditional Bi-FPN limited to feature interactions between local layers, the dense Bi-FPN constructs a global dense connection model, ensuring that any layer's output feature maps can deeply fuse the detailed spatial information from the highest resolution input feature maps. Specifically, the dense Bi-FPN first propagates the

input feature map with the highest resolution upward layer by layer and performs the splicing operation with the output feature maps of each layer, which breaks through the bottleneck of information between a single layer and realizes the flow and sharing of detailed information across layers. By realizing dense interconnection and multilevel fusion of feature maps, dense Bi-FPN significantly improves the model's detection performance of multiscale targets, especially the accuracy of recognizing small targets and the depth of contextual understanding in complex scenes, thus demonstrating higher accuracy and robustness on target detection tasks.

III. DATASETS AND EVALUATION METRICS

A. Datasets

The data used in this study were obtained from a public dataset called "Swedish Forest Agency (2021)" [52]. The dataset contains 840 high-resolution images collected by drone aerial photography from five specific study areas in Sweden - Bebehojd, Ekbacka, Jallasvag, Kampe, and Nordkap. Table II details the quantitative statistics of each category in the dataset.

The Swedish Forest Agency (2021) dataset was carefully categorized according to tree damage status into four main categories: Light Damage Larches (labeled as Larch_LD), High Damage Larches (labeled as Larch_HD), Healthy Larches (labeled Larch_other), and Other Trees (labeled Other_other), and example images of trees in each category are shown in Fig. 6. To efficiently develop and evaluate the model, the dataset was partitioned into three subsets according to standard practice: 80% of the data were used for the training set to train the model parameters, 10% of the data comprised the test set to evaluate the model's performance on unseen samples, and the other 10% constituted the validation set, which was used to tune the model hyperparameters as well as to monitor the model's performance during the training process to avoid overfitting.

B. Evaluation Metrics

To evaluate the experimental results, this study uses five core evaluation metrics: Recall, Precision, F1-Score, AP, and mAP. Among them, *AP* is a comprehensive measure of the detection model's localization and classification performance on specific categories, *mAP* is used to reflect the average precision (AP) performance of the model under each recall rate, while *F1-Score* integrates precision and recall and effectively measures the balance between precision and recall of the model. The above evaluation metrics are calculated as follows:

$$\text{Recall} = \frac{T_P}{T_P + F_N} \quad (15)$$

$$\text{Precision} = \frac{T_P}{T_P + F_P} \quad (16)$$

$$F1 - Score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (17)$$

$$AP = \int_0^1 P(r) \, dr \quad (18)$$

TABLE II
SWEDISH FOREST AGENCY (2021) DATASET LABELING STATISTICS

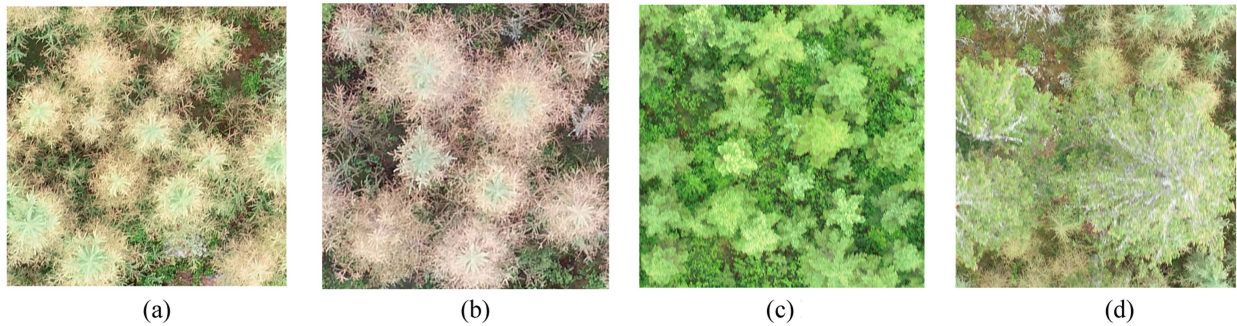| Damaged area	Images	Data label			
		Healthy larches	Light damage larches	High damage larches	Other trees
Bebehojd	172	142	6279	2328	2570
Ekbacka	215	327	5126	970	2702
Jallasvag	50	625	1743	212	475
Kampe	179	1905	11 318	2285	3219
Nordkap	224	488	8668	2106	2987
Total	840	3487	33 143	7901	11 953

| (a) | (b) | (c) | (d) |

Fig. 6. Swedish forest agency (2021) dataset example images of each type of data labeling. (a) Light damage larches. (b) High damage larches. (c) Healthy larches. (d) Other trees.

$$mAP = \frac{1}{|C|} \sum_{k=1}^{|C|} AP. \qquad (19)$$

In the above evaluation metric system, $P(r)$ denotes the Precision value when the recall rate is r. T_P (True Positives) refers to the number of target instances successfully identified and accurately located by the model, that is, the number of target individuals predicted by the model as a certain category and the actual situation is indeed in that category. In contrast, F_P (False Positives) measures the number of instances the model misjudges as a certain category, expressed as the number of objects that the model judges as a target category but do not belong to that category. In addition, F_N (False Negatives) is defined as the number of instances missed or not accurately identified by the model, that is, the number of targets that exist and should be attributed to a certain category, but the model fails to identify successfully. In the multicategory scenario, $|C|$ denotes the number of categories.

IV. EXPERIMENT

A. Experimental Setup

In this study, the convolutional neural network model was built, trained, and validated on the Windows 10 operating system platform using the Keras 2.2.5 deep learning framework. The development environment uses the Python 3.6 programming language. The experimental hardware configuration includes the NVIDIA RTX 2080 Ti GPU, which takes full advantage of its embedded CUDA computing architecture, and the integrated CuDNN acceleration library to improve the parallel computing efficiency of the model training phase. At the same time, the Intel(R) Core(TM) i7-10700 model was selected as the CPU. In the experimental phase, the sample images to be detected were uniformly preprocessed to a resolution of 480 pixels × 480 pixels, and the number of training samples per batch was set to 2. The initial learning rate was set to 0.01, and the cosine annealing strategy was used to gradually adjust the learning rate. When designing the network architecture, taking into account the need to discriminate larch species with different degrees of erosion and their nonlarch reference categories, in this study, the output classification number of the network was set to NCLASSES = 4 in order to carefully evaluate the different damage levels of larch and achieve effective discrimination from nonlarch samples.

To scientifically evaluate the performance of fuzzy Efficient-Det, this study conducts detailed comparison experiments between it and the current mainstream methods in the field of target detection, which include SSD, Faster R-CNN, YOLO V5, and the original version of EfficientDet. The systematic performance of multiple methods under the same conditions is tested and analyzed to quantitatively and qualitatively evaluate the performance advantages of fuzzy EfficientDet and its scope of applicability in the target detection task.

B. Performance of Global–Local SE Module

To systematically explore the impact of the global–local SE module on network performance and its effectiveness, this study implements a controlled experiment. This experiment compares

TABLE III
COMPARATIVE ANALYSIS OF THE IMPACT OF DIVERSE ATTENTION MECHANISMS ON THE PERFORMANCE METRICS OF EFFICIENTDET MODELS

Attention mechanism	Damage degree	Recall(%)	Precision(%)	F1-Score	AP(%)	mAP(%)
Global–Local SE	Healthy larches	88.30	88.64	88.47	90.31	
	Light damage larches	90.38	91.25	90.81	92.17	91.37
	High damage larches	89.11	89.83	89.47	91.52	
	Other trees	89.04	90.69	89.86	91.48	
Squeeze-and-Excitation	Healthy larches	86.32	87.72	87.01	88.74	
	Light damage larches	89.79	89.48	89.63	90.96	89.73
	High damage larches	86.95	87.70	87.32	89.27	
	Other trees	88.98	88.95	88.97	89.93	
None	Healthy larches	85.43	87.04	86.23	87.94	
	Light damage larches	88.27	88.91	88.59	90.03	88.87
	High damage larches	85.51	86.52	86.01	88.33	
	Other trees	85.38	87.35	86.35	89.16	

the performance of EfficientDet with the global–local SE module with that of the original EfficientDet model, which only includes the basic SE module and the original EfficientDet model without any SE module enhancement. The results of this comparison experiment are shown in Table III.

From the analysis of the experimental data in Table III, it can be seen that EfficientDet incorporating the global–local SE module shows significant advantages in target detection performance. Taking the lightly damaged larch as an example, the performance metrics of the original EfficientDet model, including mAP, recall, precision, F1 score, and category AP, were recorded as 88.87%, 88.27%, 88.91%, 88.59%, and 90.03%, respectively, with the respective values at the low end of the comparison models. This is mainly attributed to the original EfficientDet architecture, which utilized a lot of depthwise convolution. Although this structure helps to reduce the model size and computational burden, it splits the cross-channel information interaction process into two steps. It lacks the ability to directly integrate multichannel information in channel-by-channel convolution, which leads to a partial loss of cross-channel information in the feature extraction process, limiting the model's ability to extract cross-channel information.

In contrast, the corresponding performance metrics of the variant version of EfficientDet with the addition of the standard SE module are improved to 89.73% (recall), 89.79% (precision), 89.48% (F1 score), and 90.96% (AP), which are significantly better than the original model. This is because the SE module enhances the channel interactions within the model by learning and adjusting the interchannel weights, thus improving the overall feature representation.

After further introduction of the global–local SE module, the experimental results show that its detection performance in the light damage larches category reaches an optimal state with an mAP of 91.37%, which is a 1.64% improvement compared to the variant equipped with the SE module only. This is because the SE module emphasizes the global significance while ignoring the locally important features, while the global–local SE module proposed in this study can better fuse the global and local features extracted from the backbone network by introducing local average pooling, thus improving the ability of the EfficientDet model to detect insect-infested larch.

C. Performance of the Dense Fuzzy Bi-FPN

To deeply analyze the effectiveness of each innovative component in the dense fuzzy Bi-FPN proposed in this study, the ablation experiments in this study were designed by taking a step-by-step approach to add the CBAM attention mechanism, the FSAM attention mechanism, and the dense Bi-FPN structure to EfficientDet to clarify the specific contribution of each component to the performance of the model.

The results of the ablation experiments are shown in Table IV. By comparing the evaluation metrics under different combinations, it is possible to intuitively assess the role of the CBAM, the FSAM, and the dense Bi-FPN structure in improving the target detection accuracy as well as the overall effectiveness of the model.

The data in Table IV show that the detection performance of the damaged larch model was significantly improved by introducing the dense connection Bi-FPN structure. Taking the light damage larches category as an example, after only incorporating the dense connection structure into the detection network, the corresponding recall, precision, F1 score, and category AP increased to 89.53%, 90.76%, 90.14%, and 92. 26%, which is comparable to that of the original EfficientDet model by 1.26%, 1.85%, 1.55%, and 2.23%, respectively. It confirms that the construction of dense connections in Bi-FPN can enrich the model's ability to capture contextual information and improve its ability to handle long-range spatial dependencies.

TABLE IV
EXPERIMENTAL EVALUATION OF ABLATION OF DENSE FUZZY BI-FPN

Damage degree	Densely Bi-FPN	FSAM	CBAM	Recall(%)	Precision(%)	F1-Score	AP(%)
Healthy larchs	✗	✗	✗	85.43	87.04	86.23	87.94
	✓	✗	✗	86.95	88.91	87.92	89.59
	✗	✗	✓	86.32	87.72	87.01	87.65
	✗	✓	✗	88.17	89.52	88.84	90.98
	✓	✓	✗	89.25	90.61	89.92	92.07
Light damage larchs	✗	✗	✗	88.27	88.91	88.59	90.03
	✓	✗	✗	89.53	90.76	90.14	92.26
	✗	✗	✓	89.79	89.48	89.63	91.46
	✗	✓	✗	90.72	92.07	91.39	92.57
	✓	✓	✗	91.36	92.64	91.99	93.23
High damage larchs	✗	✗	✗	85.51	86.52	86.01	88.33
	✓	✗	✗	87.82	88.07	87.94	89.34
	✗	✗	✓	86.95	87.70	87.32	88.83
	✗	✓	✗	86.69	88.36	87.52	90.82
	✓	✓	✗	87.73	89.47	88.59	91.92
Other trees	✗	✗	✗	85.38	87.35	86.35	89.16
	✓	✗	✗	87.03	90.01	88.49	90.08
	✗	✗	✓	88.98	88.95	88.97	90.57
	✗	✓	✗	87.80	89.34	88.56	91.28
	✓	✓	✗	88.47	90.54	89.49	92.14

In addition, in order to systematically evaluate the performance advantages of the FSAM proposed in this study on the object detection task of larch Coleophora laricella insect pests, the FSAM and the currently widely used comparative benchmark attention module (CBAM) were respectively embedded into the basic EfficientDet model architecture. According to the analytical data in Table IV, the EfficientDet variant embedded in CBAM has outperformed the original EfficientDet network. CBAM adjusts the weight distribution of the feature map by jointly applying channel attention and spatial attention, strengthening the model's attention to the salient areas of the input image, which helps to mitigate the impact of visual differences caused by seasonal changes and diversity of lighting conditions on object detection accuracy.

Further, when FSAM is introduced, the experimental results show that its performance is better than that of the EfficientDet variant using CBAM, with a 2.14% improvement in mAP. Especially in the detection of light damage larchs, the indicators of FSAM-EfficientDet exceed those of the CBAM variant. This shows that the FSAM attention mechanism proposed in this study makes use of the continuity property of Sugeno-type fuzzy logic to dynamically allocate spatial attention according to the continuous change of the image feature space. Compared with the traditional discrete hard attention mechanism, FSAM exhibits stronger adaptability and stability when dealing with image feature perturbations caused by lighting changes and seasonal changes.

After further discussion, after integrating the dense connection structure and FSAM in the EfficientDet, the model performance reached the optimal state, and its indicators were Recall 91.36%, Precision 92.64%, F1-Score 91.99%, and AP 93.23%, which were significantly improved compared with the initial EfficientDet model. The experimental results verify that the dense Bi-FPN structure combined with the FSAM attention mechanism has a significant effect on improving the performance of detection of larch tree species Coleophora laricella pests, thereby enhancing the effectiveness and accuracy of the entire detection network.

D. Fuzzy EfficientDet Performance

The experiments in this section integrate the global–local SE module, the FSAM attention mechanism, and the dense Bi-FPN to construct a target detection model, Fuzzy EfficientDet, for detecting pests and diseases on larch images captured by UAVs. Fig. 7 shows the detection examples of fuzzy EfficientDet for larch with different degrees of damage under different lighting conditions in practical applications, which fully demonstrate the detection effect of the model in complex scenes.

Fig. 7(a) shows the detection of healthy larch under strong direct sunlight, showing that the fuzzy EfficientDet model can still accurately identify trees in extreme lighting environments. In conventional lighting scenarios, as shown in Fig. 7(b) and (c), the fuzzy EfficientDet model also performed well, successfully
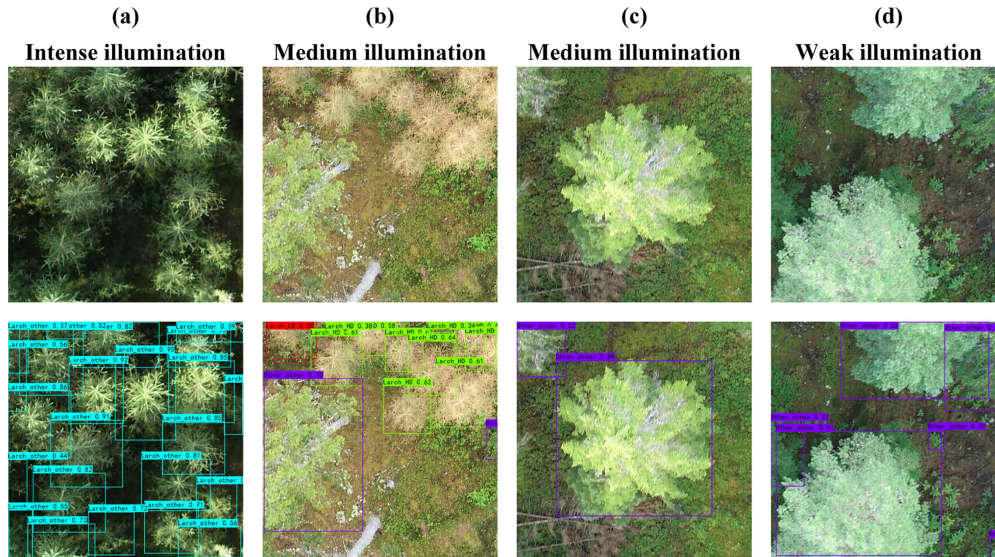
Fig. 7. Example of fuzzy EfficientDet's detection of larch under diverse light conditions.

achieving accurate classification of all tree species in the picture. Fig. 7(d) shows the detection ability of the fuzzy EfficientDet model in the background of poor lighting conditions. It can be observed that even in the case of low light, the model can still accurately identify nonlarch "Other trees," which once again verifies the robustness and adaptability of the fuzzy EfficientDet model under various lighting conditions, ensuring its ability to stably and accurately detect targets in complex environmental changes.

E. Performance Comparison of Different CNN Methods

To evaluate the performance advantages of the fuzzy EfficientDet proposed in this study in the task of larch Coleophora laricella pest detection and evaluation, a comparison experiment was designed in this study, in which mainstream target detection networks, such as SSD, Faster R-CNN, YOLO V5, and the original EfficientDet, were selected as control models. The above models and fuzzy EfficientDet are run under the same experimental conditions in the comparison experiment. Table V details all the data of the comparison experiments, including the evaluation metrics of mAP, Recall, Precision, and F1-Score.

The experimental data in Table V show that compared to SSD, Faster R-CNN, YOLO V5, and the original EfficientDet, the proposed fuzzy EfficientDet shows a significant improvement in all the primary evaluation metrics, especially in recall, precision, F1 score, and AP. The improvement of fuzzy EfficientDet on mAP is particularly outstanding, which is 5.37%, 1.76%, 3.82%, and 4.42% compared to SSD, Faster R-CNN, YOLO V5, and the original EfficientDet, respectively. This indicates that the fuzzy EfficientDet proposed in this study can better perform the larch Coleophora laricella pest detection and evaluation task.

V. DISCUSSION

The outstanding performance of fuzzy EfficientDet, an innovative deep learning framework for the challenges of Coleophora laricella larch pest detection in UAV images, stems from several

key design improvements. First, the model embeds the FSAM, specifically designed to mitigate the problem of uncertainty in image features due to environmental factors, lighting variations, and seasonal differences, and helps the model capture targets stably and accurately under complex conditions. Second, fuzzy EfficientDet introduces the global–local SE module, which is an advanced feature recalibration component that can adjust the feature channel weights in the EfficientNet by simultaneously considering both global and local feature information, thus enhancing the network's ability to key feature recognition and expression capabilities. Furthermore, the model constructs a dense Bi-FPN structure, which improves tiny targets' modeling accuracy and spatial dependencies by adding a dense connection layer on top of the original Bi-FPN. It solves the deficiencies of the traditional target detection network in dealing with such problems. Therefore, the algorithm exhibits unique advantages over one-stage and two-stage detection methods.

For the one-stage detection method, fuzzy EfficientDet can process complex environmental disturbances and lighting changes in real time without increasing too much computational cost, and significantly improve the positioning accuracy of potential targets, especially when dealing with small targets. The dense connection layer provided by the dense Bi-FPN structure enhances the sensitivity to target scale changes and the ability to fuse context information, thus improving the detection accuracy while ensuring speed. For two-stage detection methods, such as Faster R-CNN and EfficientDet, fuzzy EfficientDet further optimizes the feature extraction and selection process with the help of the global–local SE module. This module can dynamically adjust the feature channel weights according to the global and local features at different levels, enabling the model to target more accurately in both candidate region generation and fine position regression phases. In addition, the FSAM mechanism also plays an important role in the two-stage method, reducing the impact of feature uncertainty on detection performance, especially in the case of severe environmental changes or

TABLE V
DETECTION OF INSECT-DAMAGED LARCH BY DIFFERENT MODELS

Model	Damage degree	Recall(%)	Precision(%)	F1-Score	AP(%)	mAP(%)
SSD	Healthy larches	85.48	86.19	86.28	87.98	
	Light damage larches	88.45	88.9	88.67	90.09	88.92
	High damage larches	85.51	86.58	86.08	88.36	
	Other trees	85.45	87.43	86.39	89.23	
Faster RCNN	Healthy larches	88.79	91.62	90.11	91.24	
	Light damage larches	92.36	92.03	92.35	93.47	92.53
	High damage larches	89.12	90.75	89.27	92.19	
	Other trees	88.58	91.29	90.2	93.21	
YOLO V5	Healthy larches	87.04	88.62	87.9	89.58	
	Light damage larches	89.9	90.53	90.18	91.63	90.47
	High damage larches	87.05	88.12	87.63	89.94	
	Other trees	86.96	88.92	88.01	90.73	
EfficientDet	Healthy larches	85.43	87.04	86.23	88.94	
	Light damage larches	88.27	88.91	88.59	91.03	89.87
	High damage larches	85.51	86.52	86.01	89.33	
	Other trees	85.38	87.35	86.35	90.16	
Fuzzy EfficientDet	Healthy larches	90.59	92.34	91.42	93.67	
	Light damage larches	93.71	93.61	93.96	95.83	94.29
	High damage larches	89.81	90.72	90.25	93.21	
	Other trees	89.44	92.36	91.65	94.46	

difficult to distinguish target details, improving the robustness and generalization ability of the model.

In summary, whether it is in the one-stage direct prediction of the fast response of the target bounding box, or in the complex process of two-stage first roughly screening the candidate region and then accurately locating the target, the fuzzy EfficientDet has shown superior performance improvement through its unique design, not only significantly surpassing similar algorithms in key indicators such as mAP, but also achieving comprehensive growth in multiple evaluation dimensions such as Recall, Precision, and F1-Score.

VI. CONCLUSION

Real time and accurate detection and assessment of larch pests (Coleophora laricella) are of critical practical value for maintaining the health and stability of forest ecosystems. This study systematically sorted out the advantages and disadvantages of existing larch pest detection methods. This study focused on the challenges of larch pest detection using UAV imagery under complex natural environment conditions. To address these challenges, this study innovatively proposes the fuzzy EfficientDet model, which aims to improve the limitations of existing target detection techniques in dealing with such complex problems. First, the model incorporates the global–local SE module into the backbone network, which improves the problem of cross-channel interaction information loss caused by depthwise convolution in the backbone and thus significantly improves the feature expression capability of the backbone network. Second, this study introduces a dense connection structure in the Bi-FPN, strengthening the connection between features at different levels and significantly improving the model's ability to parse and model long-range spatial dependencies. Finally, the core innovation of fuzzy EfficientDet lies in designing a FSAM, which has more robust adaptability and stability in the face of image feature perturbations caused by light changes and seasonal transitions.

This study explores the application of target detection algorithms in the forestry field by designing and verifying the superior performance of the fuzzy EfficientDet model in the detection and assessment of larch infestation. Considering the needs and broad prospects of future practical applications, we shift our focus to improving the generalization ability and universality of the fuzzy EfficientDet model and plan to make the model capable of more forestry target detection tasks, such as identifying pests and diseases of different tree species, monitoring forest health, etc., to broaden its practical value in the field of forestry protection and management.

REFERENCES

[1] S. F. Ward et al., "The role of simulated spring water stress in interactions between eastern larch and larch casebearer," *Arthropod-Plant Interact.*, vol. 13, no. 4, pp. 621–633, 2019.

[2] J. Y. Huang et al., "An improved YOLOX algorithm for forest insect pest detection," *Comput. Intell. Neurosci.*, vol. 2022, 2022, Art. no. 5787554.

[3] Y. Zhang, C. Wu, T. Zhang, Y. Liu, and Y. Zheng, "Self-attention guidance and multiscale feature fusion-based UAV image object detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 6004305.

[4] H. Q. He, F. Zhou, Y. Xia, M. Chen, and T. Chen, "Parallel fusion neural network considering local and global semantic information for citrus tree canopy segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 1535–1549, 2024.

[5] Y. Fang, Y. Z. Cai, and L. Fan, "SDRCNN: A single-scale dense residual connected convolutional neural network for pansharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 6325–6338, 2023.

[6] H. F. Gu et al., "Satellite-detected contrasting responses of canopy structure and leaf physiology to drought," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2427–2436, 2023.

[7] G. B. Liang, J. Zulu, X. Xing, and N. Jacobs, "Unveiling roadway hazards: Enhancing fatal crash risk estimation through multiscale satellite imagery and self-supervised cross-matching," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 535–546, 2024.

[8] S. W. Liu et al., "An improved Kuhn Munkres algorithm for ship matching in optical satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 4724–4738, 2023.

[9] L. Sun, S. Cheng, Y. Zheng, Z. Wu, and J. Zhang, "SPANet: Successive pooling attention network for semantic segmentation of remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4045–4057, 2022.

[10] P. Qin, Y. Cai, J. Liu, P. Fan, and M. Sun, "Multilayer feature extraction network for military ship detection from high-resolution optical remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 11058–11069, 2021.

[11] J. Nie et al., "Data and domain knowledge dual-driven artificial intelligence: Survey, applications, and challenges," *Expert Syst.*, 2023, Art. no. e13425, doi: 10.1111/exsy.13425.

[12] Y. Li and S. Ercisli, "Explainable human-in-the-loop healthcare image information quality assessment and selection," *CAAI Trans. Intell. Technol.*, 2023, doi: 10.1049/cit2.12253.

[13] E. Z. Zhao, L. L. Dong, and J. Shi, "Infrared maritime target detection based on iterative corner and edge weights in tensor decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 7543–7558, 2023.

[14] W. X. Zhou, H. Guan, Z. Li, Z. Shao, and M. R. Delavar, "Remote sensing image retrieval in the past decade: Achievements, challenges, and future directions," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1447–1473, 2023.

[15] B. Zhang et al., "Progress and challenges in intelligent remote sensing satellite systems," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1814–1822, 2022.

[16] Y. Yang et al., "Dissimilarity-based active learning for embedded weed identification," *Turkish J. Agriculture Forestry*, vol. 46, no. 3, pp. 390–401, 2022.

[17] J. Nie et al., "Artificial intelligence and digital twins in sustainable agriculture and forestry: A survey," *Turkish J. Agriculture Forestry*, vol. 46, no. 5, pp. 642–661, 2022.

[18] K. Song et al., "Regional soil water content monitoring based on time-frequency spectrogram of low-frequency swept acoustic signal," *Geoderma*, vol. 441, 2024, Art. no. 116765.

[19] Y. Li et al., "Low-carbon jujube moisture content detection based on spectral selection and reconstruction," *IEEE Internet Things J.*, to be published, doi: 10.1109/JIOT.2024.3368140.

[20] X. Chao and Y. Li, "Semisupervised few-shot remote sensing image classification based on KNN distance entropy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 8798–8805, 2022.

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.

[22] W. Liu et al., "SSD: Single shot MultiBox detector," in *Proc. 14th Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.

[23] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. 27th IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 580–587.

[24] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448.

[25] S. Q. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. 29th Annu. Conf. Neural Inf. Process. Syst.*, 2015, pp. 1137–1149.

[26] J. Nie et al., "Sustainable computing in smart agriculture: Survey and challenges," *Turkish J. Agriculture Forestry*, vol. 46, no. 4, pp. 550–566, 2022.

[27] Y. Q. Chen et al., "An efficient approach to monitoring pine wilt disease severity based on random sampling plots and UAV imagery," *Ecological Indicators*, vol. 156, 2023, Art. no. 111215 .

[28] Y. Liu et al., "A tree species classification model based on improved YOLOv7 for shelterbelts," *Front. Plant Sci.*, vol. 14, 2023, Art. no. 1265025.

[29] J. W. Lv et al., "Detecting pests from light-trapping images based on improved YOLOv3 model and instance augmentation," *Front. Plant Sci.*, vol. 13, 2022, Art. no. 939498.

[30] B. Z. Wu et al., "Application of conventional UAV-based high-throughput object detection to the early diagnosis of pine wilt disease by deep learning," *Forest Ecol. Manage.*, vol. 486, 2021, Art. no. 118986.

[31] Z. Y. Xue et al., "YOLO-tea: A tea disease detection model improved by YOLOv5," *Forests*, vol. 14, no. 2, 2023, Art. no. 415.

[32] M. Yu et al., "A shallow pooled weighted feature enhancement network for small-sized pine wilt diseased tree detection," *Electronics*, vol. 12, no. 11, 2023, Art. no. 2463.

[33] M. Tan et al., "EfficientDet: Scalable and efficient object detection," 2020, *arXiv:1911.09070*.

[34] J. Jia et al., "Underwater object detection based on improved efficientdet," *Remote Sens.*, vol. 14, no. 18, 2022, Art. no. 4487.

[35] X. Li et al., "MEDMCN: A novel multi-modal EfficientDet with multi-scale CapsNet for object detection," *J. Supercomput.*, pp. 1–28, 2024, doi: 10.1007/s11227-024-05932-1.

[36] X. Zhu et al., "Detection the maturity of multi-cultivar olive fruit in orchard environments based on Olive-EfficientDet," *Scientia Horticulturae*, vol. 324, 2024, Art. no. 112607.

[37] X. Zhuang et al., "Military target detection method based on EfficientDet and generative adversarial network," *Eng. Appl. Artif. Intell.*, vol. 132, 2024, Art. no. 107896.

[38] Y. Q. Hu et al., "Detection of rice pests based on self-attention mechanism and multi-scale feature fusion," *Insects*, vol. 14, no. 3, 2023, Art. no. 280.

[39] L. Kang et al., "YOLO-FA: Type-1 fuzzy attention based YOLO detector for vehicle detection," *Expert Syst. Appl.*, vol. 237, 2024, Art. no. 121209.

[40] R. P. Yang et al., "An FA-SegNet image segmentation model based on fuzzy attention and its application in cardiac MRI segmentation," *Int. J. Comput. Intell. Syst.*, vol. 15, no. 1, 2022, Art. no. 24.

[41] Y. Nan et al., "Fuzzy attention neural network to tackle discontinuity in airway segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2023.3269223.

[42] W. C. Xiao et al., "FDAN: Fuzzy deep attention networks for driver behavior recognition," *J. Syst. Archit.*, vol. 147, 2024, Art. no. 103063.

[43] Q. P. Chong et al., "A multiscale fuzzy dual-domain attention network for urban remote sensing image segmentation," *Int. J. Remote Sens.*, vol. 43, no. 14, pp. 5480–5501, 2022.

[44] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[45] R. J. Xu et al., "A forest fire detection system based on ensemble learning," *Forests*, vol. 12, no. 2, 2021, Art. no. 217.

[46] N. AlDahoul et al., "Localization and classification of space objects using EfficientDet detector for space situational awareness," *Sci. Rep.*, vol. 12, no. 1, 2022, Art. no. 21896.

[47] T. B. Shahi et al., "Recent advances in crop disease detection using UAV and deep learning techniques," *Remote Sens.*, vol. 15, no. 9, 2023, Art. no. 2450.

[48] K. L. Zou et al., "A field weed density evaluation method based on UAV imaging and modified U-net," *Remote Sens.*, vol. 13, no. 2, 2021, Art. no. 310.

[49] X. P. Xie, D. Yue, and C. Peng, "Relaxed real-time scheduling stabilization of discrete-time Takagi–Sugeno Fuzzy systems via an alterable-weights-based ranking switching mechanism," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 6, pp. 3808–3819, Dec. 2018.

[50] J. Cheng, Y. Shan, J. Cao, and J. H. Park, "Nonstationary control for T–S fuzzy Markovian switching systems with variable quantization density," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 6, pp. 1375–1385, Jun. 2021.

[51] Y. L. Wang, Q.-L. Han, M.-R. Fei, and C. Peng, "Network-based T–S fuzzy dynamic positioning controller design for unmanned marine vehicles," *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2750–2763, Sep. 2018.

[52] Swedish Forest Agency, "Forest damages—Larch casebearer 1.0," version 880, National Forest Data Lab, Dataset, 2021.

Shuo Yang received the B.S. degree in mechatronic engineering from Tarim University, Xinjiang, China, in 2018. He is currently working toward the Ph.D. degree in agricultural engineering with the College of Mechanical and Electrical Engineering, Shihezi University, Shihezi, China.

His research interests include object detection, image segmentation, and pattern recognition.

Jing Nie received the M.S. degree in mechatronic engineering from Xinjiang University, Xinjiang, China, in 2007, and the Ph.D. degree in mechatronic engineering from Shihezi University, Shihezi, China, in 2023.

He is currently a Professor with the College of Mechanical and Electrical Engineering, Shihezi University, China. His research interests include deep learning, image processing, and agricultural IoT applications.

Jingbin Li received the M.S. degree in agricultural mechanization engineering from Shihezi University, Shihezi, China, in 2006, and the Ph.D. degree in mechatronic engineering from China Agricultural University, Beijing, China, in 2013.

He is currently a Professor with the College of Mechanical and Electrical Engineering, Shihezi University, China. His research interests include machine learning, image processing, and agricultural IoT applications.

Yujie Qiao is currently working toward the B.S. degree in human geography and urban and rural planning with the College of Sciences, Shihezi University, Shihezi, China.

Her research interests include smart agriculture, deep learning, and image processing.

Yang Li received the M.S. degree in electrical engineering from Dalian University of Technology, Dalian, China, in 2016, and the Ph.D. degree in information and communication engineering from Tianjin University, Tianjin, China, in 2023.

He is currently an Associate Professor with the College of Mechanical and Electrical Engineering, Shihezi University, China. His research interests include information processing, pattern recognition, and data quality assessment.
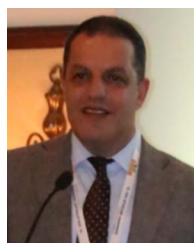
Sezai Ercisli received the Ph.D. degree in agricultural engineering from Ataturk University, Erzurum, Türkey, in 1996.

He is currently a Professor with the Department of Horticulture, Agricultural Faculty of Ataturk University, Türkey. His research interests include the modern technology in agricultural production, such as deep learning, image processing, and Internet of Things. He is the Editor-in-Chief of *Turkish Journal of Agriculture and Forestry*.