



Article

LWSDNet: A Lightweight Wheat Scab Detection Network Based on UAV Remote Sensing Images

Ning Yin ¹, Wenxia Bao ^{1,*}, Rongchao Yang ², Nian Wang ¹ and Wenqiang Liu ¹

¹ National Engineering Research Center for Agro-Ecological Big Data Analysis & Application, Anhui University, Hefei 230601, China; p21101009@stu.ahu.edu.cn (N.Y.); wn_xlb@ahu.edu.cn (N.W.); p20201092@stu.ahu.edu.cn (W.L.)

² Zhengzhou Tobacco Research Institute of CNTC, Zhengzhou 450001, China; yangrc@ztri.com.cn

* Correspondence: bwxia@ahu.edu.cn

Abstract: Wheat scab can reduce wheat yield and quality. Currently, unmanned aerial vehicles (UAVs) are widely used for monitoring field crops. However, UAV is constrained by limited computational resources on-board the platforms. In addition, compared to ground images, UAV images have complex backgrounds and smaller targets. Given the aforementioned challenges, this paper proposes a lightweight wheat scab detection network based on UAV. In addition, overlapping cropping and image contrast enhancement methods are designed to preprocess UAV remote-sensing images. Additionally, this work constructed a lightweight wheat scab detection network called LWSDNet using mixed deep convolution (MixConv) to monitor wheat scab in field environments. MixConv can significantly reduce the parameters of the LWSDNet network through depthwise convolution and pointwise convolution, and different sizes of kernels can extract rich scab features. In order to enable LWSDNet to extract more scab features, a scab feature enhancement module, which utilizes spatial attention and dilated convolution, is designed to improve the ability of the network to extract scab features. The MixConv adaptive feature fusion module is designed to accurately detect lesions of different sizes, fully utilizing the semantic and detailed information in the network to enable more accurate detection by LWSDNet. During the training process, a knowledge distillation strategy that integrates scab features and responses is employed to further improve the average precision of LWSDNet detection. Experimental results demonstrate that the average precision of LWSDNet in detecting wheat scab is 79.8%, which is higher than common object detection models and lightweight object detection models. The parameters of LWSDNet are only 3.2 million (M), generally lower than existing lightweight object detection networks.

Keywords: lightweight object detection; wheat scab; UAV remote sensing; feature fusion; knowledge distillation



Citation: Yin, N.; Bao, W.; Yang, R.; Wang, N.; Liu, W. LWSDNet: A Lightweight Wheat Scab Detection Network Based on UAV Remote Sensing Images. *Remote Sens.* **2024**, *16*, 2820. <https://doi.org/10.3390/rs16152820>

Academic Editor: Amin Beiranvand Pour

Received: 26 June 2024

Revised: 22 July 2024

Accepted: 29 July 2024

Published: 31 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Wheat, as a crucial global crop, is closely linked to human life and the development of the entire economy and society, playing a central role in ensuring food security. As a staple food source for many people worldwide, wheat is deeply processed into daily necessities such as noodles and bread, providing the human body with various nutrients such as carbohydrates and protein. However, during the growth cycle of wheat, various diseases such as rust, scab, powdery mildew, and root rot often pose a serious threat to its yield and quality [1]. Wheat scab is particularly prominent, as it can directly affect wheat grains, causing not only a reduction in grain yield but also mycotoxin pollution, thereby seriously affecting grain quality [2]. Mycotoxins not only cause poor development and shape damage of seeds but also pose a threat to consumer health when ingested into the body. There is a need to research efficient and accurate detection methods to enable the rapid and precise identification of the location of wheat scab. These methods would enable farmers to promptly identify the spread of the disease in the field. As a preventive measure, farmers

could then harvest wheat from infected areas early to interrupt the spread of scab. This would help ensure the maximum quality and safety of wheat products. Simultaneously, through the scientific detection and data analysis of wheat scab, agricultural researchers can more precisely assess the severity of the disease in the field. This enables the selection of varieties highly resistant to scab, promoting the effective conduct of disease-resistant breeding and fundamentally strengthening the wheat industry's resilience against disease.

The traditional wheat scab detection process involves agricultural experts conducting on-site inspections and collecting samples for analysis and discrimination, which is not only time-consuming but also requires a significant amount of labor costs. In recent years, with the rapid development of artificial intelligence, researchers in the field of agriculture have been using machine learning or deep learning methods to analyze and detect diseases. Before deep learning was widely studied, traditional machine learning methods were often used for disease detection [3–6]. Utilizing machine learning techniques for disease detection necessitates the meticulous design of the actual disease characteristics, followed by automated feature extraction during the testing phase and subsequent automated detection based on these features. Pallathadka et al. [7] introduced a machine learning-inspired framework for leaf disease classification and detection. Initially, the framework employs histogram equalization to preprocess the leaves. Subsequently, it utilizes the PCA algorithm to extract features from the leaves. Finally, the leaves are classified using the support vector machine (SVM). Kumar Nagothu et al. [8] developed a machine-learning algorithm for weed detection, which can differentiate between weeds and crops in each image based on predefined plants. After testing, the algorithm achieved an accuracy rate of over 90% in weed detection. Jaisakthi et al. [9] utilized image processing and machine learning techniques to develop an automated detection system for grapevine diseases. Initially, the system segments leaf from the background image, followed by a further segmentation of the diseased areas. By utilizing support vector machines and random forests for disease classification, the system achieved an accuracy of 93%. Bao et al. [10] proposed a wheat leaf disease and its severity discrimination method based on Elliptical Maximum Margin Criterion metric learning. The recognition accuracy of this algorithm for wheat powdery mildew and stripe rust was about 94.2%. Larijani et al. [11] utilized the K-means improved KNN (K-Nearest Neighbor) algorithm to classify images in the Lab color space for detecting disease spots on rice leaves. Although traditional machine learning methods can achieve good results, they rely more on manually selected features. In natural scenes, the image background is relatively complex, and manually selecting features is somewhat difficult.

Deep learning can automatically select feature information that is suitable for the current task. This capability reduces the labor cost associated with manually selecting features, as compared to machine learning approaches. As a result, deep learning is well-suited for agricultural disease detection tasks in complex environments. Scholars have utilized deep learning methods for disease detection and achieved abundant research results [12–18]. Li et al. [19] proposed a vegetable disease detection based on YOLOv5 and innovatively improved the cross-stage partial network (CSPNet), feature pyramid network (FPN), and non-maximum suppression (NMS) modules in YOLOv5s to enhance the multi-scale feature extraction ability of the network. Through experiments, it was proven that the average accuracy of vegetable disease detection reached 93.1%. Zhang et al. [20] proposed a detection algorithm for accurately detecting multi-scale apple leaf spots in unconstrained environments and designed a BoLe convolution module to extract more effective feature information. They also proposed a cross-attention module to make the network pay more attention to foreground content. Through experimental verification, the proposed network achieved an average accuracy of 85.23% on a self-built dataset. Kerkech, Hafiane and Canals [13] used UAV to collect image data of vineyards and used convolutional neural networks and color information to detect grape leaf diseases. Bao et al. [21] proposed a YOLO object detection algorithm with a two-dimensional mixed attention mechanism for monitoring tea wilt disease and used super-resolution and image contrast enhancement

technology to enhance image features. Experimental results showed that the proposed algorithm improved the baseline recall by 6.5%. However, high-precision detection models have a large number of parameters and calculations, which makes them unable to be deployed on devices with limited computing resources. Moreover, models with large parameters cannot achieve real-time detection and are not suitable in some efficient scenarios.

In practical applications of deep learning-based disease detection models, most model parameters and computational complexity are substantial, rendering it challenging to deploy the model on devices with limited computing resources.

To ensure the accurate and efficient performance of convolutional neural networks in making predictions, researchers have extensively investigated the utilization of model compression methods. These methods aim to reduce the number of network parameters and typically involve techniques such as model pruning [22,23] and knowledge distillation [24–27]. By employing these approaches, the networks can achieve compact yet effective representations, facilitating improved computational efficiency and resource utilization. Bao et al. [28] constructed a lightweight convolutional neural network called SimpleNet using convolutional and inverted residual blocks and enhanced its feature expression ability using CBAM and feature fusion. The network was used to automatically identify wheat ear diseases, with a recognition accuracy of up to 94.1 and a model parameter size of only 2.13 M.

Liu, Gao, Chen and Cheng [22] proposed a simple and effective model pruning method called discriminative perception channel pruning and recognition kernel pruning, which removes channels and channel kernels that are not conducive to discriminative ability. In image classification tasks, the accuracy is still 0.36% higher than the baseline model, even when the channel is reduced by 30%. Liu, Zhang and Wang [24] introduced a novel adaptive learning framework for multi-teacher, multi-level knowledge extraction, capable of adaptively gathering intermediate-level knowledge from multiple teachers and attaining superior performance compared to its competitors. Yang et al. [29] introduced a novel distillation method called focus and global distillation, which divides the knowledge distillation process into focus distillation and global distillation. Focus distillation separates foreground and background, allowing the student network to focus on the teacher's key pixels and channels. Global distillation identifies the connections between different pixels and transmits them to the student network.

In the field environment, disease detection commonly employs a method that integrates UAVs with artificial intelligence. UAV images are fed to the detection model to pinpoint the disease-affected areas. Yet, detection models tend to be voluminous, and the computational resources of UAVs fall short of enabling real-time on-site detection. Consequently, this study introduces a lightweight model for detecting wheat scab tailored to address challenges posed by limited UAV computational resources, intricate image backgrounds, and minimally visible scab targets. The main contributions and innovations of this study are as follows:

- The overlapping cropping and image contrast enhancement methods are designed to preprocess the collected UAV remote sensing images, and an object detection dataset of wheat scab is constructed.
- A lightweight wheat scab detection network, LWSDNet, was developed using MixConv. The designed scab feature enhancement and MixConv adaptive feature fusion modules in the network can enhance the feature representation of scab while fusing feature information from different contexts to detect minimal scab targets.
- A knowledge distillation strategy that integrates scab features and responses is proposed. During the training process of LWSDNet, the actual scab location information is combined with feature-based knowledge distillation to ensure that only the features relevant to the scab region are subjected to distillation. Concurrently, the knowledge of the teacher network is transferred to LWSDNet to enhance its detection accuracy for scab by using response-based knowledge distillation.

- The designed LWSDNet has only 3.2 M parameters, and the average precision of detection on the wheat scab dataset is higher than that of common object detection models and lightweight object detection models, reaching 79.8%.

2. Materials and Methods

2.1. Data Acquisition

The wheat dataset used in this study was collected in May 2021, and the crop was in the wheat-filling stage. On the day of collection, the weather was sunny and accompanied by a gentle breeze. The collection site is located in Ba Town, Chaohu City, Anhui Province, China. The variety of wheat planted locally is Wanmai 1648, which belongs to the half-winter variety.

The wheat image acquisition device is DJI Mavic Air 2, manufactured by DJ-Innovations in Shenzhen, China, with a maximum range of 18.5 kilometers. In addition, its horizontal and vertical hovering errors can be maintained within 0.1 m. The UAV can achieve a telemetry imaging distance of 10 km at a resolution of 1080 P and maintain a low delay of 130 milliseconds, which can meet the requirements of wheat data collection. The camera carried by the UAV has an effective pixel size of 48 million, an aperture of $f/2.8$, and an exposure time of $1/125$ s. Besides, the camera can obtain a maximum image size of 6000×8000 , and it can clearly record the wheat situation on the ground at a flight altitude of 4 m. At the same time, the UAV is equipped with a precise visual positioning system and global navigation satellite system (GNSS), which can record the longitude, latitude, and altitude information of the image shooting location in detail. The UAV also integrates an inertial measurement unit (IMU) module for easy control during data collection. During data collection, the flight altitude of the UAV is adjusted to 4 m and hovered with the camera lens perpendicular to the ground for shooting.

2.2. Overlapping Cropping

In order to meet the input requirements of the network proposed in this paper, cropping operations are performed on the wheat images collected by the UAV. This study designs an overlapping cropping method to improve the generalization performance of the network, in which the specific operation process is as follows:

As shown in Figure 1, the width and length of the cropped small image are defined as p_w and p_h , respectively, and use Equations (1) and (2) to crop the small image.

$$w_e - w_s = p_w \quad (1)$$

$$h_e - h_s = p_h \quad (2)$$

The w_s and h_s in Equations (1) and (2) represent the starting position of the small image to be cropped on the original image, while w_e and h_e represent the end position of the small image to be cropped on the original image.

A window of size $p_w \times p_h$ is used to slide on the original image and first crop along the width of the original image. After cropping a small image, update the starting position w_s and ending position w_e according to Equations (3) and (4) to obtain the next cropped small image, where $\gamma \in (0, 1)$ represents the overlap ratio of cropped small images, which is set to 0.15 in this article.

$$w'_s = w_e - \gamma \times p_w \quad (3)$$

$$w'_e = w'_s + p_w \quad (4)$$

when the starting width position of cropping w_s is greater than the original image width, it indicates that a row of horizontal small images has been cropped. It is necessary to update h_s and h_e according to Equations (5) and (6) to continue cropping.

$$h'_s = h_e - r \times p_h \quad (5)$$

$$h'_e = h'_s + p_h \quad (6)$$

when the starting height of cropping is greater than the original image height, it indicates that the UAV image has been captured. The cropped wheat image is shown in Figure 2.

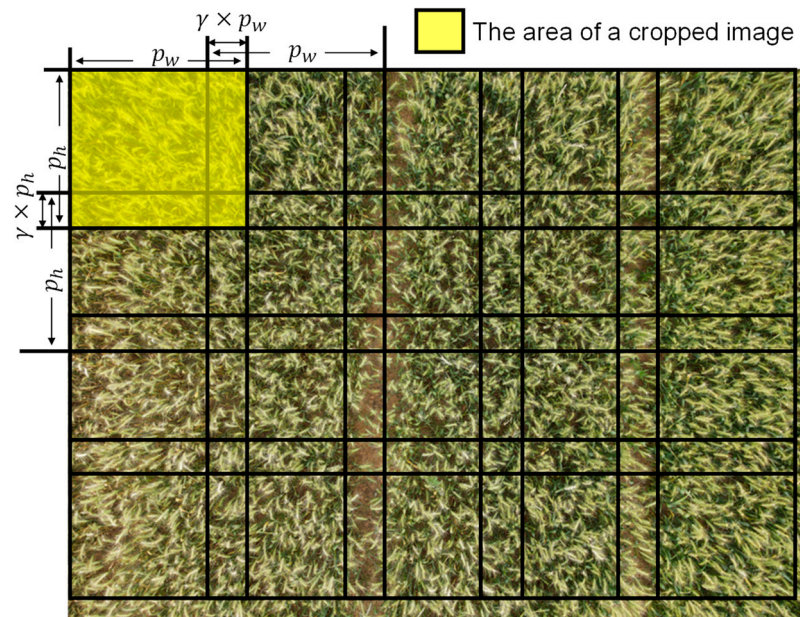


Figure 1. Overlapping cropping.

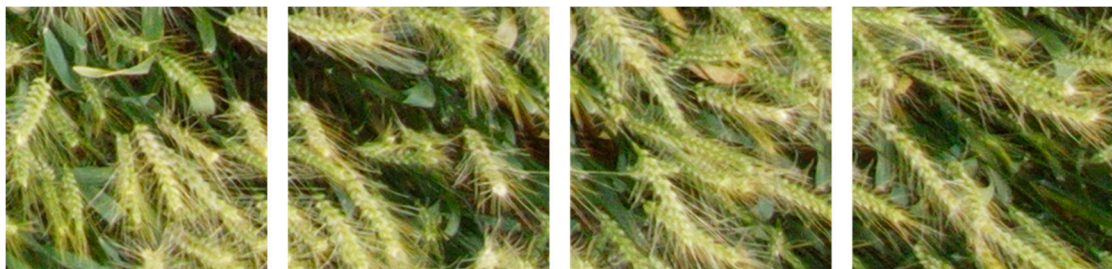


Figure 2. Cropped wheat image.

The utilization of the overlapping cropping technique serves to preserve the integrity of edge defect detection while concurrently facilitating data expansion. To enhance the overall diversity of the dataset, this research adopts the approach of overlapping cropping from both the top left and bottom right corners of the UAV imagery. Consequently, the original UAV images, originally sized at 8000×6000 , are uniformly transformed into smaller images measuring 416×416 .

2.3. Image Contrast Enhancement

The UAV-acquired wheat images show that color and tone represent the most apparent distinctions between the regions affected by the scab and the background. Given that the data used in this study was collected from wheat fields under natural lighting conditions, the captured images display irregular illumination and occasional instances of excessive exposure. As a result, the disease spots within the images are often represented indistinctly. To alleviate these concerns, this research employs an image contrast enhancement technology as a preliminary processing step for the wheat image data, with the purpose of highlighting the distinctive attributes of the disease-affected regions.

The color intensity of the three channels of the image is transformed separately according to Equation (7), followed by a normalization operation as per Equation (8). Equation (7) enhances the clarity of regions with higher color intensity associated with scab in the

image while suppressing the background regions with lower color intensity. By employing this image enhancement technique, the diseased regions in the image can be highlighted, facilitating the subsequent convolutional neural network to extract more disease-specific features and thereby achieve more accurate discrimination and detection, as described in this study.

$$I'(x, y) = I^3(x, y), x \in [0, W), y \in [0, H) \quad (7)$$

$$I''(x, y) = \frac{I'(x, y) - \min(I'(x, y))}{\max(I'(x, y)) - \min(I'(x, y))} \times 255 \quad (8)$$

In the given equations, $I(x, y)$ represents the color intensity at coordinates (x, y) in a specific channel of the wheat image. The variables W and H correspond to the image's width and height, respectively. The symbol $I'(x, y)$ indicates the transformed color intensity, while $I''(x, y)$ represents the normalized channel color intensity of $I'(x, y)$. To provide visual context, a portion of the image before and after contrast enhancement is depicted in Figure 3. The region enclosed by the red box signifies the area affected by wheat scab. Notably, the application of image contrast enhancement results in a more distinct appearance of the affected regions, thereby facilitating improved differentiation.

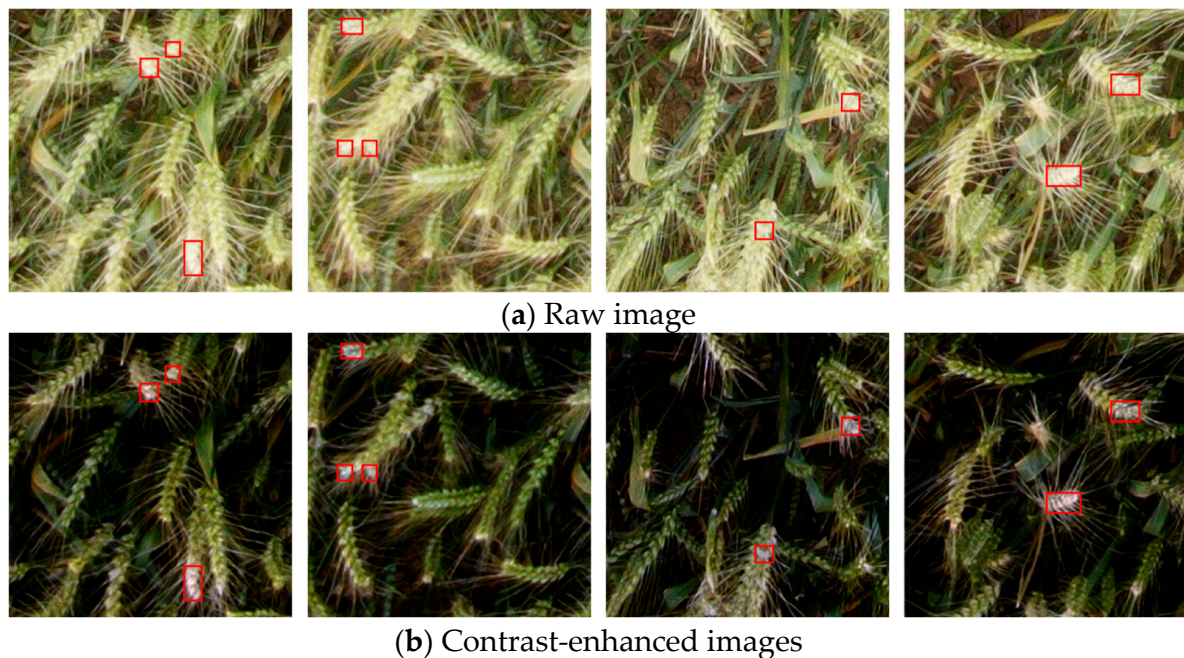


Figure 3. Wheat raw images and contrast-enhanced images. The red box represents the area of wheat scab.

2.4. Dataset Production

After applying the image contrast enhancement technology, the regions affected by scab exhibit significantly improved distinguishability, thereby facilitating their recognition during the annotation process. The wheat scab dataset was annotated using the Labeling commercial software version 1.8.1 in this study. This software enables the generation of label files for object detection by creating annotation bounding boxes. The graphical user interface of Labeling during the wheat image annotation process is depicted in Figure 4. Following the completion of annotation, the dataset was divided into training and testing sets in a 2:1 ratio. As a result, the final dataset comprised 1344 images for training and 672 images for testing.

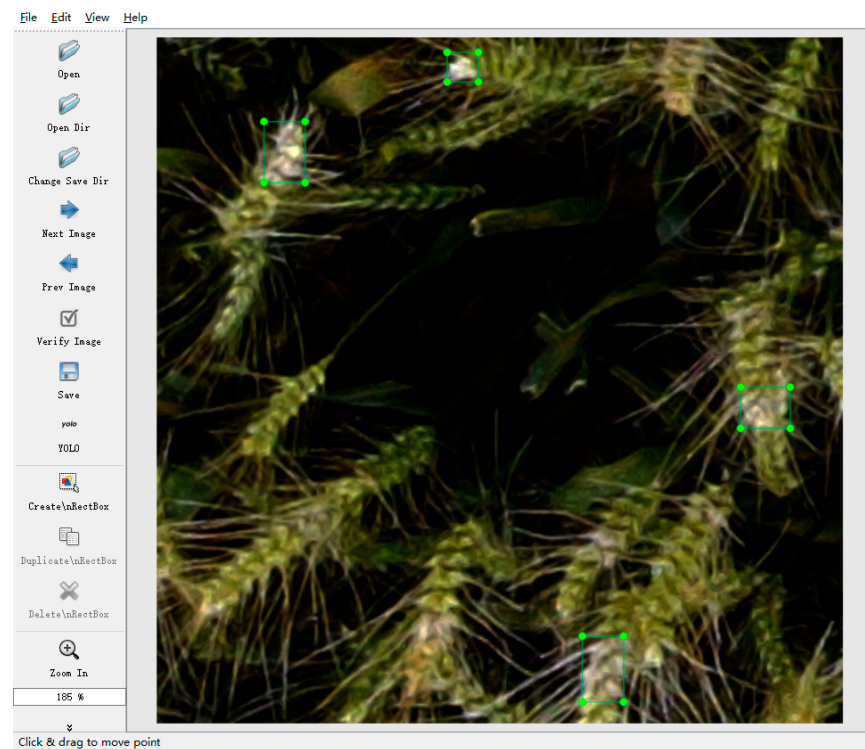


Figure 4. Labeling 1.8.1 software interface. The green area is the labeled area of wheat scab.

2.5. LWSDNet

Given the issue of insufficient computational resources in UAV, as well as the problems of complex backgrounds and relatively small targets in UAV imagery compared to ground-based images. To overcome these challenges, a lightweight wheat scab disease detection model specifically designed for UAV images was proposed. The method introduced LWSDNet, a lightweight wheat scab detection network constructed using MixConv [30]. MixConv effectively captures features with diverse receptive fields by combining different kernel sizes within a single convolutional operation. Moreover, the parameters and computational complexity of the network were reduced through the utilization of deep convolution and pointwise convolution operations in MixConv. Furthermore, the proposed method incorporated a scab feature enhancement module within LWSDNet to emphasize scab-specific features. In addition, this method designs a MixConv feature fusion module, which facilitates the fusion of features from different layers to detect scab of varying scales. To further enhance the precision of LWSDNet, this approach is designed to improve the overall performance of the network.

The proposed LWSDNet adopted the network framework of YOLOv5, as depicted in Figure 5. The network architecture consists of three key components: Backbone, Neck, and Prediction Head. To achieve lightweightness, the MixConv was employed throughout the network for efficient feature extraction. Within the Backbone module, a scab feature enhancement module is specifically designed to enhance the network's ability to extract relevant features related to wheat scab. Moreover, in the neck module, a novel MixConv adaptive feature fusion module, introduced in this study, is utilized to fuse features from different layers. By leveraging both shallow-level detailed features and deep-level semantic features, the precision of the detection process can be enhanced.

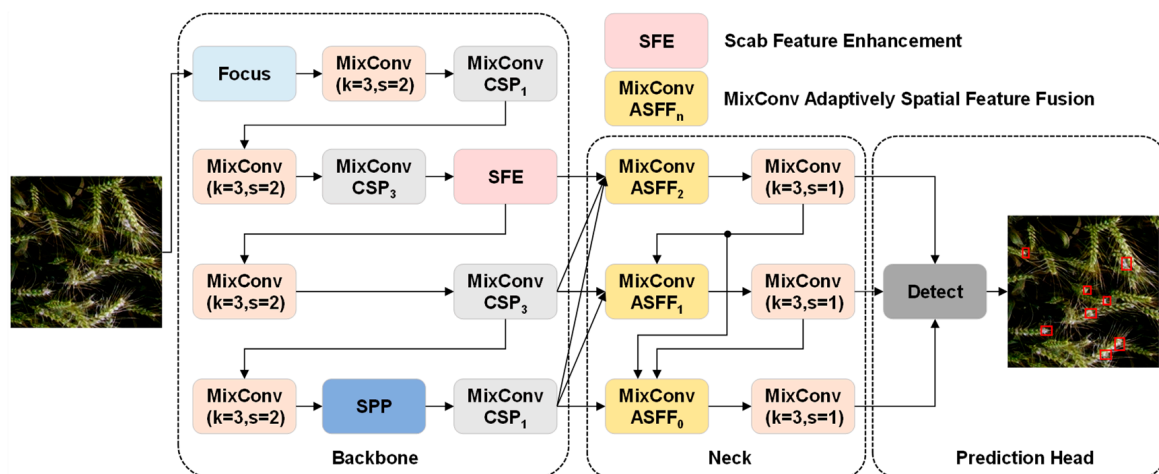


Figure 5. The LWSDNet architecture. The red boxed represent the wheat scab detected by the network.

2.5.1. MixConv

In order to minimize the parameters and computational complexity of the network, LWSDNet utilized a MixConv for feature extraction. Typically, convolutional operations employ a fixed kernel size of 3×3 . However, research has demonstrated that different kernel sizes can produce varying detection outcomes [31]. Given that distinct kernel sizes can capture features at different scales, in object detection tasks, the network requires both large kernels to capture abstract features and small kernels to extract detailed features.

MixConv efficiently reduces the parameters and computational complexity of the network through the integration of depthwise convolution and pointwise convolution operations. As depicted in Figure 6, during the process of depthwise convolution, the input features are partitioned into multiple groups and diverse-sized convolutional kernels are applied to each group for convolution, thereby enriching the diversity of feature extraction. In MixConv, the input features are partitioned into three groups with a ratio of 2:1:1, and convolutional operations are conducted using 3×3 , 5×5 , and 7×7 kernel sizes, represented as k_1 , k_2 , and k_3 , respectively, in Figure 6. In comparison with conventional convolutions, MixConv substantially reduces the parameters of the network. Although MixConv introduces a slight increase in parameters compared to depthwise separable convolution that solely employs 3×3 kernels, it significantly enhances the feature extraction capability of the network.

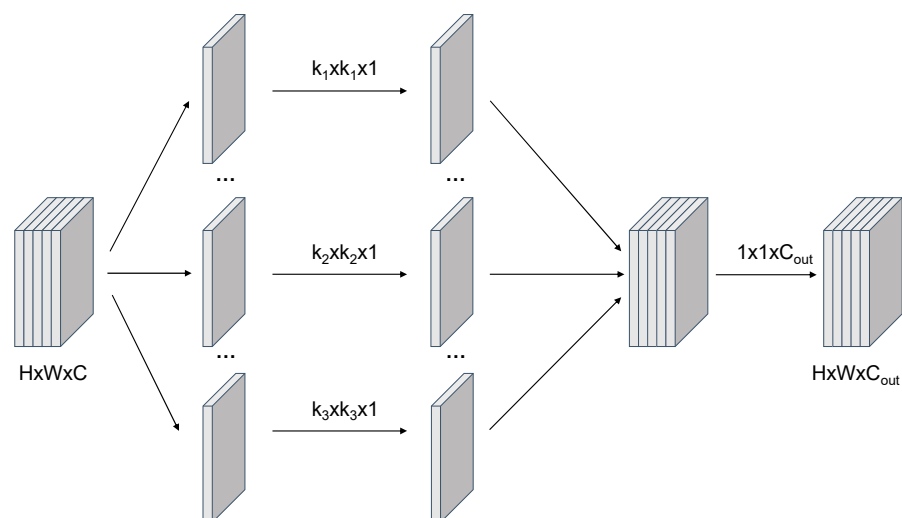


Figure 6. MixConv.

2.5.2. Scab Feature Enhancement

Although the MixConvCSP module effectively captures features from wheat images at different resolutions, it may introduce a considerable amount of background information into the extracted features. This can potentially have a detrimental impact on the detection process, particularly given the small size of the wheat scab in the images. To address this issue, this study introduced a scab feature enhancement module (SFE) in LWSDNet. This module emphasizes the regions of interest by leveraging the correlation between semantic and detailed information. As depicted in Figure 7, the SFE module processes the input features through two distinct branches, employing different strategies for feature enhancement. In the first branch, the input features are subjected to three separate branches, each utilizing dilated convolutions with varying dilation rates to extract features at different scales. By carefully configuring the parameters of the dilated convolutions, output features of the same size are achieved. Subsequently, the features obtained from the three branches are merged through element-wise summation to obtain enriched wheat scab disease features in the wheat images.

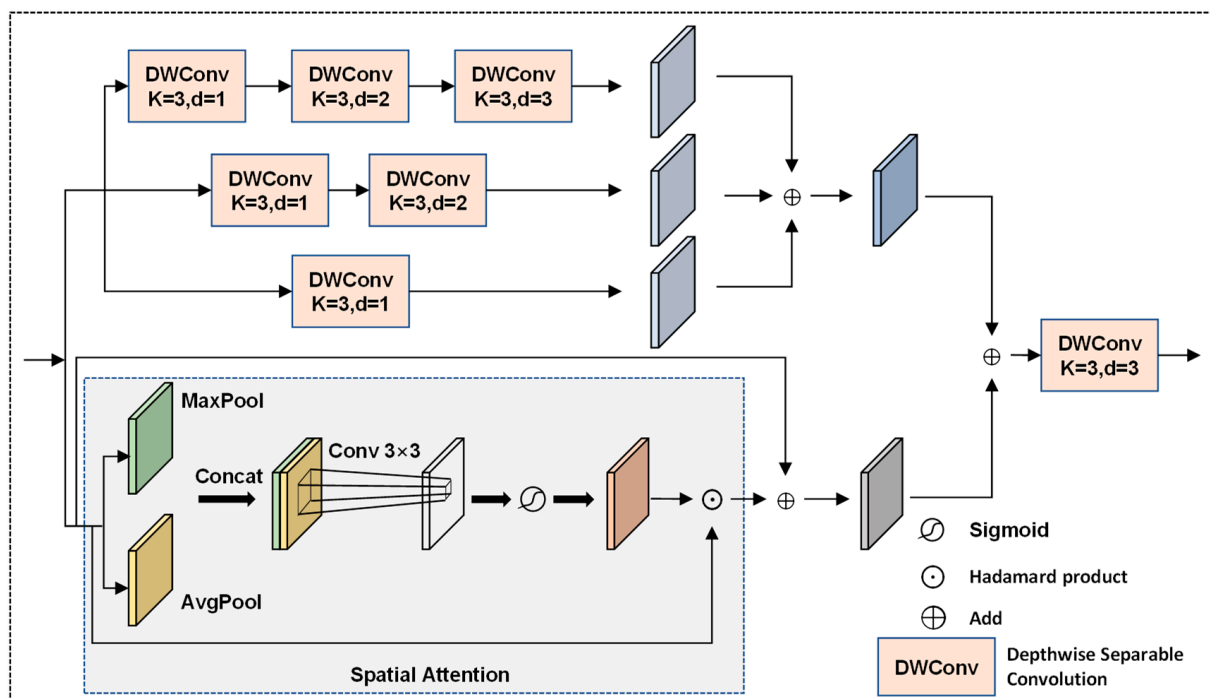


Figure 7. Scab feature enhancement.

Within another branch of the SFE module, a spatial attention mechanism is employed to amplify the information pertaining to scab in the input features. As depicted in Figure 7, the input features undergo separate operations of max pooling and average pooling. Subsequently, the resulting features are concatenated along the channel dimension and subjected to a depthwise separable convolution utilizing a 3×3 kernel. Ultimately, the obtained output is normalized through Sigmoid activation and Hadamard product with the input features. The features derived from the spatial attention module are then element-wise added to the input features, thereby achieving the desired effect of feature enhancement.

2.5.3. MixConv Adaptive Feature Fusion

In the UAV wheat scab dataset constructed in this study, the wheat scab exhibit varying sizes, with a prevalence of small targets. Within convolutional neural networks, different layers yield distinct information in their output features. Deep layers capture features related to larger targets or semantic information, while shallow layers encompass features associated with smaller targets or fine-grained details. To effectively leverage these

two types of features for the detection of wheat scab of diverse sizes, a MixConv adaptive feature fusion module is employed in this research to merge features from different layers. As depicted in Figure 8, features from different layers are initially resized to a consistent dimension, followed by convolution and channel-wise concatenation. The concatenated features then undergo further convolution and softmax operations to derive normalized weights corresponding to each feature. Finally, the obtained weights and features are processed using Hadamard product and MixConv to generate fused features.

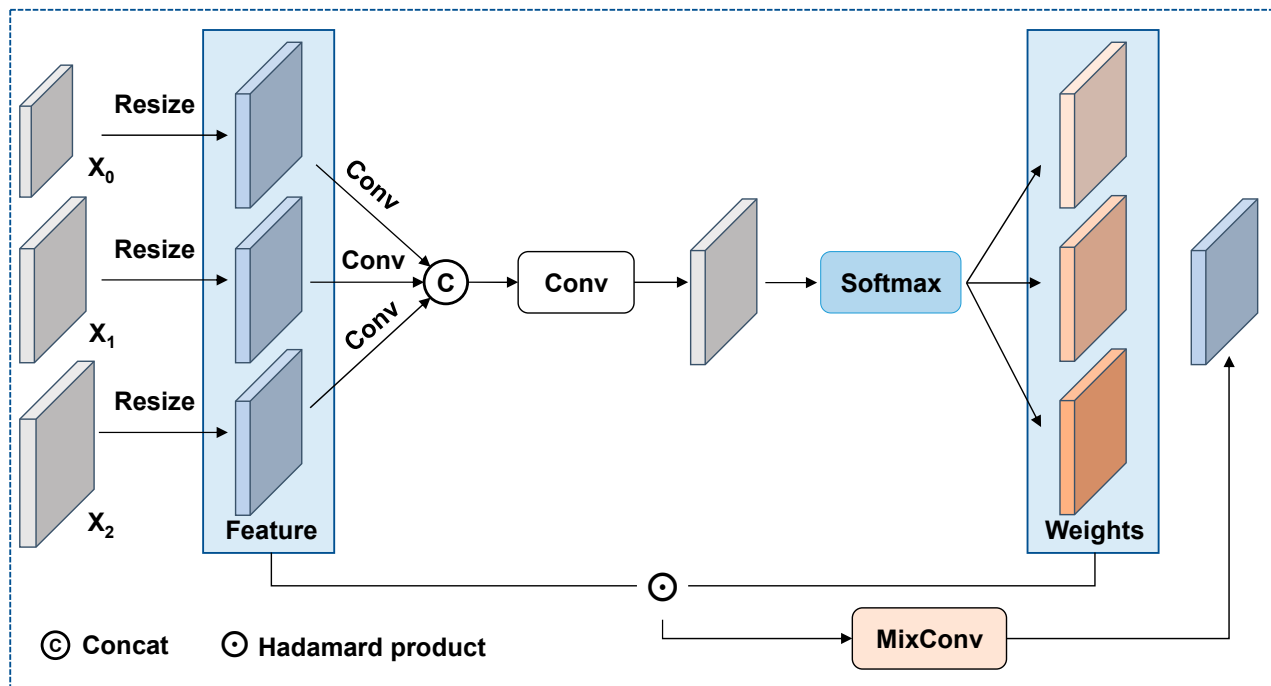


Figure 8. MixConv adaptive feature fusion.

2.6. Knowledge Distillation Strategy That Integrates Scab Features and Responses

In order to further enhance the average precision of LWSDNet predictions, this study employs the technique of knowledge distillation during its training process. The teacher network used for knowledge distillation is obtained by replacing lightweight convolutions with regular convolutions in LWSDNet. As illustrated in Figure 9, the training procedure incorporates feature-based knowledge distillation to enrich the scab features extracted by LWSDNet. The output scab features from the teacher network are utilized to guide the training of LWSDNet. Given that the SFE module integrates both local and global features, thus demonstrating strong feature representation capabilities, the output of the SFE module in the teacher network is employed as the learning objective for LWSDNet. To mitigate the interference from background features, this study combines the actual scab location information with feature-based knowledge distillation and subsequently locates the area of scab in the features through bounding boxes, ensuring that only the features relevant to the scab region are subjected to distillation. During the training process, preprocessed wheat images are simultaneously fed into both the teacher network and LWSDNet. The output features of the SFE module are element-wise multiplied with the scab label mask to obtain the scab features, which are then utilized as the learning target for LWSDNet.

In order to facilitate LWSDNet in acquiring knowledge from the output responses of the teacher network, this study further employs a response-based knowledge distillation approach. This methodology enables LWSDNet to learn the detection boxes and confidence of scab from the teacher network. In the feature-based knowledge distillation process, a loss function is utilized to constrain the scab features extracted by both the teacher network and the student network (LWSDNet). Through iterative backpropagation of the

loss, the parameters of both networks are updated, progressively minimizing the disparity in the extracted scab features. This iterative process allows the student network to capture increasingly comprehensive and refined feature information that closely aligns with the teacher network. Concurrently, the response-based knowledge distillation utilizes the same loss function to enforce similarity between the predictions of scab boxes and confidence made by the teacher network and the student network. During the testing phase, the trained student network (LWSDNet) is employed to obtain the final test results by inputting the test data.

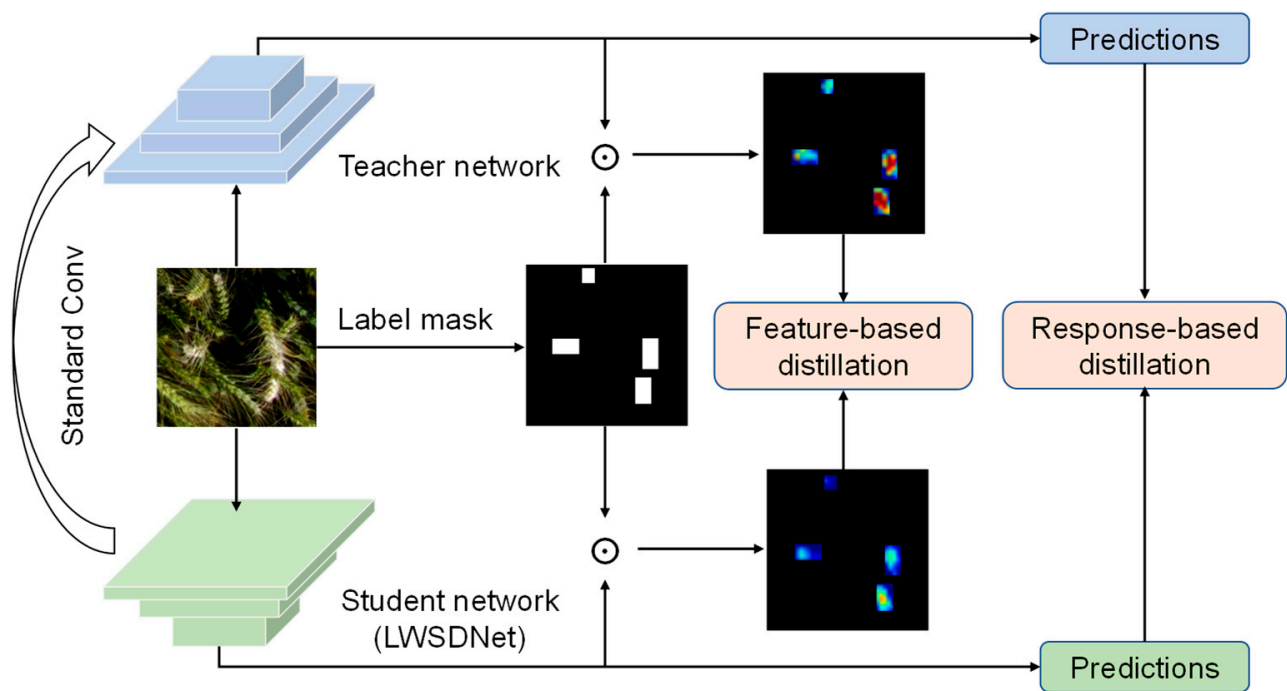


Figure 9. Knowledge distillation training framework. The boxes in the figure represent the area of wheat scab, and the redder the color, the more attention the network is paying to this area.

2.7. Loss Function

This study employs the complete intersection over union (CIOU) loss as the localization loss for LWSDNet, which is computed according to Equation (9).

$$L_{loc} = L_{CIOU} = 1 - \left(IOU - \frac{Distance_2^2}{Distance_C^2} - \frac{v^2}{(1 - IOU) + v} \right) \quad (9)$$

In Equation (9), IOU represents the ratio between the intersection and union of the ground truth and predicted bounding boxes, quantifying their overlap. $Distance_2$ denotes the distance between the center points of the ground truth and predicted bounding boxes, as illustrated by the brown line in Figure 10. $Distance_C$ refers to the diagonal distance between the ground truth and the minimum enclosing rectangle of the predicted bounding box, depicted by the blue line in Figure 10. Furthermore, in Equation (9), the parameter v serves as a measure of aspect ratio consistency, evaluating the relationship between the width and height of the bounding boxes. Its computation is defined by Equation (10).

$$v = \frac{4}{\pi^2} \left(\tan^{-1} \frac{w^{gt}}{h^{gt}} - \tan^{-1} \frac{w^p}{h^p} \right)^2 \quad (10)$$

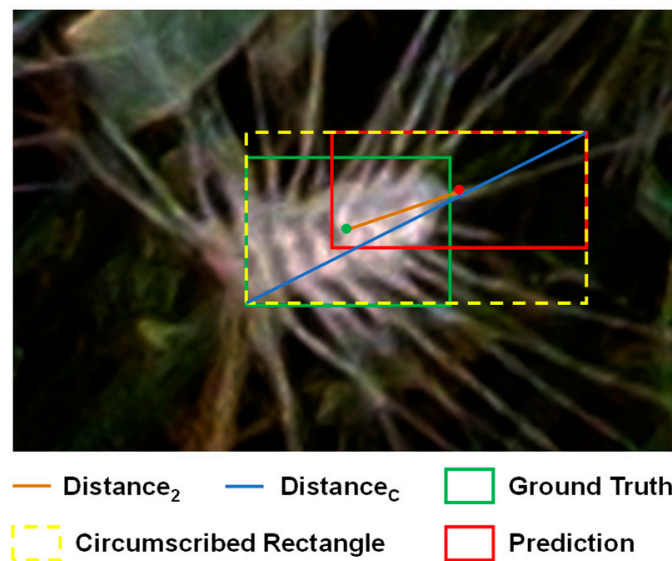


Figure 10. CIOU loss parameters.

In Equation (10), w^{gt} and h^{gt} represent the width and height of the ground truth, respectively, while w^p and h^p represent the width and height of the predicted bounding box, respectively.

Based on the number of classes in the wheat scab detection task, this study selects the binary cross-entropy loss as the confidence loss and classification loss for LWSDNet, which can be represented by Equation (11).

$$L_{cls} = L_{conf} = -\sum y \log y' \quad (11)$$

In Equation (11), y represents the ground truth and y' represents the predicted results of the network. After computing the loss, the network parameters are updated using the gradient descent method within the network.

In the feature-based knowledge distillation based on scab features, Mean Squared Error (MSE) is employed to constrain the scab features extracted by the teacher network and LWSDNet. The calculation is defined by Equation (12), where t and s represent the output features of the SFE module in the teacher network and LWSDNet, respectively. During the training process, the network parameters are updated through backpropagation to gradually reduce the value of MSE loss, thereby enabling the student network to extract scab features that closely resemble those of the teacher network. This facilitates knowledge transfer at the feature level.

$$L_{Feature} = L_{MSE}(s, t) = \frac{1}{n} \sum_{i=1}^n (t_i - s_i)^2 \quad (12)$$

This study adopts the Kullback–Leibler (KL) divergence as the loss function for response-based knowledge distillation, and its calculation is expressed by Equation (13). During the training process, both the teacher network and LWSDNet use the same loss, which comprises localization loss, classification loss, and confidence loss. Therefore, the total loss is represented by Equation (14). Here, $L_{Feature}$ and $L_{Response}$ denote the knowledge distillation losses based on scab features and response, respectively.

$$L_{Response} = L_{KL}(s, t) = \sum p_t \left(\log_2 \frac{p_t}{p_s} \right) \quad (13)$$

$$L_{LWSDNet} = L_{cls} + L_{loc} + L_{conf} + L_{Response} + L_{Feature} \quad (14)$$

3. Results

In this section, the performance of the proposed model in the wheat scab detection task is compared with commonly used object detection models, and the experimental results are analyzed. The experimental setup is presented in Table 1. During the training process, the parameters of the network are updated using Stochastic Gradient Descent (SGD), with an initial learning rate of 0.01. The training is conducted for 300 epochs with a batch size of 8.

Table 1. Experimental environment.

Platform	Configuration
CPU model	Intel(R) Xeon(R) Silver 4210
GPU model	NVIDIA GeForce RTX 3090
Operating System	Linux Ubuntu 22.04
CUDA version	12.0
Framework	Pytorch 2.2.1

To validate the predictive performance and accuracy of the proposed model, it is compared with commonly used object detection models in terms of average precision (AP), network parameters (Params), and computational complexity. Average precision is a widely used evaluation metric for object detection models, where a higher value indicates more accurate predictions by the object detection network. The calculation formula for average precision is given by Equation (15).

$$AP = \int_0^1 Precision(Recall) d(Recall) \quad (15)$$

where *Precision* and *Recall* are calculated as shown in Equations (16) and (17), respectively. In the equations, *TP* represents the number of targets where both the ground truth and the predictions indicate wheat scab, *FP* represents the number of targets where the network incorrectly predicts background as scab, and *FN* represents the number of targets where the network incorrectly predicts scab as background.

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

The network parameters represent the number of learnable parameters in the convolutional neural network, including parameters in convolutional layers, fully connected layers, Batch Normalization layers, etc. It can be used to measure the complexity of the network. Floating-point operations (FLOPs) refer to the number of floating-point operations performed by the network and can be used to measure the computational cost of the network.

3.1. Ablation Experiment

To validate the effectiveness of the proposed method, YOLOv5s is used as the baseline model, and ablation experiments are conducted by incorporating the employed image contrast enhancement, MixConv, scab feature enhancement module (SFE), and MixConv adaptive feature fusion module (MixConv AFF). The dataset utilized in the experiment was collected via UAV, and the wheat images were subsequently processed through overlapping cropping and contrast enhancement. The label overview of the dataset is presented in Figure 11. The top-left panel of the figure indicates that the dataset contains instances belonging to only the “scab” class. The top-right panel depicts the number and size distribution of the bounding boxes, while the bottom-left panel represents the spatial positioning of the bounding boxes within the images. The bottom-right panel of the figure shows the ratio of the target bounding box size to the image width and height. It can

be observed that the ratio for the scab labels is predominantly concentrated within 0.15. This observation aligns with the previous selection of a 0.15 overlap coefficient for the overlapping cropping step, as this ensures the integrity of the bounding box delineation around edge-affected regions during the detection of the scab.

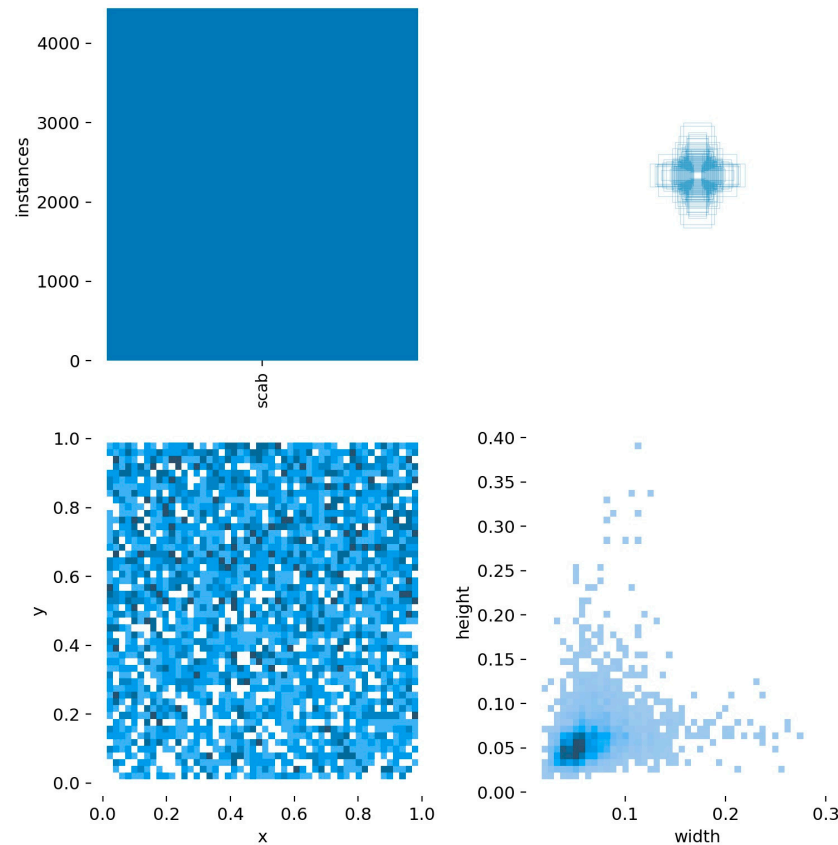


Figure 11. Overview of Dataset Labels. The darker the color, the more data there is at that point.

The results are presented in Table 2. Additionally, the ablation experiments also investigated the impact of implementing a knowledge distillation strategy during the training phase. The results in Table 2 reveal several key findings. When the image contrast enhancement algorithm proposed in this paper was applied to the dataset, the AP metric improved by 6%. The improvement in AP demonstrates the effectiveness of the image contrast enhancement algorithm proposed in this work. The baseline model incorporates MixConv for feature extraction; there is a slight decrease in the AP. Simultaneously, the parameters of the network decreased by 54.2%. By utilizing MixConv for feature extraction and replacing the PANet in the baseline model with MixConv AFF, the AP improves by 2.5% compared to using MixConv alone. Additionally, the parameters decrease due to the replacement of MixConv CSP with MixConv after MixConv AFF. Furthermore, the integration of the SFE module further enhances AP, reaching a value close to the baseline of 77.1%, with only an increase of 0.1 M parameters. Through the comprehensive utilization of the MixConv, SFE, MixConv AFF modules, and knowledge distillation strategy, the AP rises to 79.8%, surpassing the unoptimized baseline by 1.9%. Additionally, the parameters significantly decrease to 3.2 M, representing only 44.4% of the baseline. The ablation experiments validate the effectiveness of the proposed modules and methods, which not only achieve efficient parameter compression but also improve the ability of the model to detect wheat scab with higher AP.

Table 2. Ablation experiment.

Image Contrast Enhancement	MixConv	MixConv AFF	SFE	Knowledge Distillation	AP (%)	Params (M)
					71.9	7.2
✓					77.9	7.2
✓	✓				74.3	3.3
✓	✓	✓			76.8	3.1
✓	✓	✓	✓		77.1	3.2
✓	✓	✓	✓	✓	79.8	3.2

✓ indicates that the mechanism was used in the experiment.

3.2. Comparison Experiment of Lightweight Convolutions

The utilization of lightweight convolution can effectively decrease the model's parameters and computational cost. Besides MixConv, common lightweight convolutions primarily include Ghost convolution [32] and depthwise separable convolution [33]. Comparative experiments on wheat scab detection are conducted using LWSDNet, evaluating the performance of commonly used ordinary convolution and lightweight convolutions. The experimental results, as presented in Table 3, reveal some findings. When the network employed ordinary convolution for feature extraction, it achieved the highest AP. However, this came at the cost of the largest parameter size. Ghost convolution, which leverages conventional convolutions for initial feature extraction followed by linear transformations using depthwise convolution, demonstrated a 2.5% decrease in AP compared to standard convolutions. Notably, Ghost convolution exhibited a nearly 50% reduction in both parameters and computational cost. When the network utilized depthwise separable convolution for feature extraction, there was a 3.4% precision decrease compared to ordinary convolution. However, MixConv resulted in a remarkable 76.8% reduction in parameters. Compared to ordinary convolution, LWSDNet uses MixConv to reduce the total number of parameters by 11.6 M and the computational complexity by 15 gigaFLOPs, but the AP only decreases by 1%.

Table 3. Comparative experimental results of different lightweight convolutions.

Convolution	AP/%	Params/10 ⁶	FLOPs/10 ⁹
Ordinary convolution	80.8	13.8	19.2
Ghost convolution	78.3	7.2	9.7
Depthwise separable convolution	77.4	3.2	4.0
MixConv	79.8	3.2	4.2

3.3. Comparison Experiment of Different Knowledge Distillation Methods

To validate the superiority of the knowledge distillation method, a comparative analysis of five different knowledge distillation schemes for LWSDNet is conducted. The schemes examined were as follows: (1) Without knowledge distillation, representing the original performance of the LWSDNet network, (2) response-based knowledge distillation, (3) global feature-based knowledge distillation, (4) scab-features-based knowledge distillation, and (5) knowledge distillation that integrates scab features and responses. Table 4 presents the detection AP of LWSDNet under each knowledge distillation scheme.

Table 4 provides valuable insights into the impact of various knowledge distillation schemes on the detection performance of LWSDNet. Notable observations can be derived from the results presented in this table. Without knowledge distillation, the original network achieves an AP of 77.1%. However, employing response-based knowledge distillation, where the outputs of the teacher network serve as supervision for training LWSDNet, leads to a modest improvement of 0.6% in AP. Conversely, knowledge distillation based on global features fails to enhance the AP and instead results in a decline to 72.0%. Further analysis indicates that the degradation mentioned above can be attributed to the abundance of

background information present in the feature maps of the teacher network. Specifically, due to the small size of the scab in wheat images, this surplus of background information misguides the learning process of scab features in LWSDNet. In contrast, knowledge distillation based on the scab feature yields an enhancement of 1.1% in AP of LWSDNet. The proposed knowledge distillation scheme, which combines the scab feature and response, achieves the most pronounced improvement, elevating the AP of LWSDNet by 2.7%. These experimental findings convincingly demonstrate the superior efficacy of the knowledge distillation approach proposed in this study, effectively augmenting the detection performance of the network and improving its AP.

Table 4. Comparative experimental results of different knowledge distillation schemes.

Knowledge Distillation Scheme	Precision/%	Recall/%	AP/%
/	69.9	73.0	77.1
Response-based	72.5	72.6	77.7
Global feature-based	67.1	69.1	72.0
Scab feature-based	74.4	72.6	78.2
Scab feature-based + Response-based	74.5	75.2	79.8

3.4. Comparison Experiment of Different Object Detection Networks

In this section, a comparative analysis is conducted between LWSDNet and commonly used object detection networks. The experimental results are presented in Table 5. From Table 5, it is evident that YOLOv3, RetinaNet, YOLOv4, and SSD networks exhibit relatively subpar performance in wheat scab detection, with an AP of approximately 70%. The detection results of YOLOv5, YOLOv8, and YOLOv9 are better, with AP values of 77.9%, 77.4%, and 76.0%, respectively. Moreover, the table shows that among several YOLO versions, YOLOv5 has the lowest parameter and computational complexity. Faster R-CNN demonstrates a higher average precision of 79.7% but at the cost of increased parameters and computational complexity. Moreover, a comparison was made between the method used by Yang et al. [34] to detect rice leaf diseases and the method used by Yang et al. [35] to detect wheat ears. The above two methods were applied to the wheat scab detection task, achieving AP values of 66.6% and 79.4%, respectively. Compared to LWSDNet, they have a larger number of parameters and computational complexity.

Table 5. Comparative experimental results of common object detection network.

Model	AP/%	Params/ 10^6	FLOPs/ 10^9
YOLOv3	65.4	62.0	33.1
RetinaNet (ResNet34)	68.8	29.9	32.5
YOLOv4	70.1	64.0	30.0
SSD	70.5	23.8	30.4
YOLOv9s	76.0	7.2	26.7
YOLOv8s	77.4	11.1	28.4
YOLOv5s	77.9	7.2	16.6
Faster R-CNN (ResNet50)	79.7	41.5	134.6
Yang et al. [34]	66.6	36.7	105.4
Yang et al. [35]	79.4	20.0	57.3
LWSDNet	79.8	3.2	4.2

Notably, LWSDNet, the novel architecture proposed in this study, achieves the highest average precision of 79.8% in wheat scab detection. Impressively, LWSDNet accomplishes this while utilizing only 3.2 M parameters and 4.2 GFLOPs of computational cost. As a single-stage object detector, LWSDNet achieves performance on par with the two-stage object detector Faster R-CNN, with parameters and computational complexity of only 7.8% and 3.1% of Faster R-CNN. The experimental results demonstrate the superior performance

of LWSDNet in wheat scab detection. It not only achieves accurate detection but also exhibits notable advantages in terms of parameters and computational cost.

3.5. Comparison Experiment of Lightweight Object Detection Network

In order to further validate the detection performance of the proposed LWSDNet, a comprehensive comparison is conducted with commonly used lightweight object detection networks. The experimental results are presented in Table 6. Among the networks examined, YOLOv3-Tiny and YOLOv4-Tiny are designed to reduce network depth and employ only two output layers of different scales, aiming to decrease parameters and computational costs. However, these modifications result in a compromised performance in detecting small objects, with AP of 61.5% and 67.7%, respectively. On the other hand, YOLOv5n achieves the smallest parameter size and computational complexity by reducing both network depth and width, but at the expense of a noticeable drop in AP, with an AP of 73.1%.

Table 6. Comparison of experimental results of lightweight object detection networks.

Model	AP/%	Params/ 10^6	FLOPs/ 10^9
YOLOv3-Tiny	61.5	8.7	6.5
YOLOv4-Tiny	67.7	5.9	8.1
YOLOv5n	73.1	1.8	2.1
LWSDNet	79.8	3.2	4.2

In striking contrast, LWSDNet achieves the highest AP of 79.8%, surpassing YOLOv5n by a margin of 6.7%. When compared to widely adopted lightweight object detection models, LWSDNet demonstrates exceptional superiority in the wheat scab detection task, delivering the highest AP while maintaining a lower parameter size and computational complexity. The comparative analysis with lightweight object detection models unequivocally underscores the outstanding performance of LWSDNet in wheat scab detection. It not only achieves the highest AP but also exhibits a notable advantage in terms of parameter size and computational complexity.

To provide a comprehensive evaluation of LWSDNet, a visual analysis of the detection results obtained from various lightweight object detection networks is conducted. The visualization results are presented in Figure 12. In the case of YOLOv3-Tiny, the detection performance is notably subpar, with a significant number of missed detection areas. It fails to capture the regions affected by wheat scab fully and produces several false-positive detections. Similarly, YOLOv4-Tiny exhibits detection results comparable to YOLOv3-Tiny, where it manages to detect most of the affected areas. However, there are noticeable discrepancies in the predicted object positions and sizes compared to the ground truth, and false-positive detections are also present. On the other hand, YOLOv5n successfully detects all the areas affected by wheat scab, with the predicted object positions and sizes closely resembling the ground truth. However, it does suffer from a relatively higher rate of false detections.

In contrast, LWSDNet, the novel architecture proposed in this study, achieves comprehensive detection of all the scab areas while closely aligning the detection boxes with the ground truth in terms of position and size. Furthermore, it demonstrates a comparatively lower rate of false detections. The visual assessment of the detection results unequivocally highlights the superior performance of LWSDNet in accurately detecting wheat scab. It successfully captures all the affected areas, maintains consistency with the ground truth in terms of detection box position and size, and exhibits a lower rate of false detections.

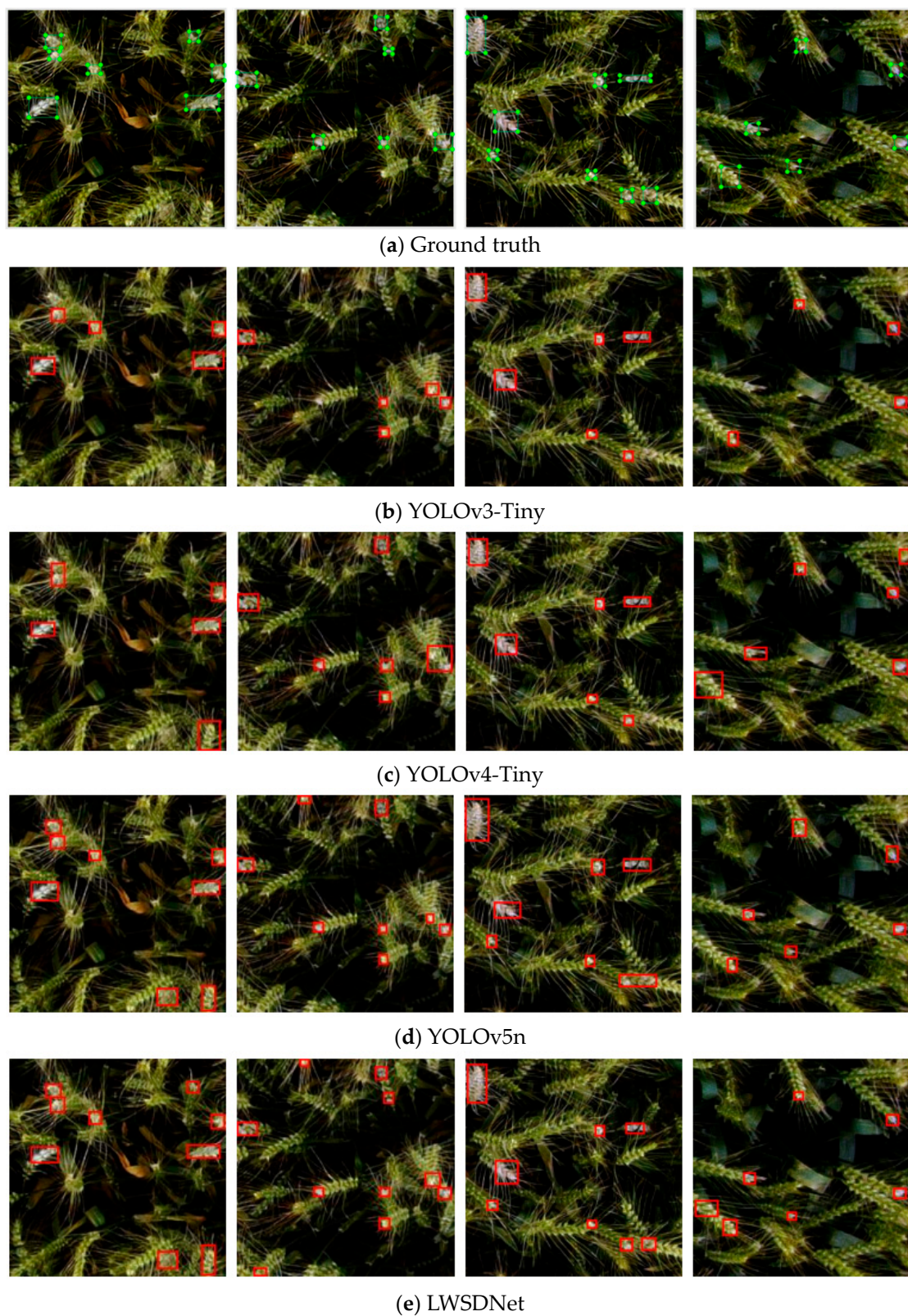


Figure 12. Visualization of lightweight network detection results. The green boxes represent the Ground Truth of wheat scab, and the red boxes represent the predicted area of wheat scab.

4. Discussion

Conventional object detection models often suffer from large parameter sizes and computational complexities, which pose limitations when applied to the task of wheat

scab detection. The detection process entails UAV capturing field wheat images, which are subsequently transmitted to a server through communication technology. The server then performs preprocessing and forward propagation to generate the detection results, which are finally transmitted back to the UAV for display. However, due to the two-way data communication involved, real-time detection becomes unattainable, and a high-performance server is needed.

In order to achieve real-time and accurate detection of wheat scab in field environments, the deployment of the proposed LWSDNet network on UAV devices is essential. It is worth noting that the designed LWSDNet exhibits lower parameter count and computational complexities comparable to MobileNetV2. Additionally, LWSDNet achieves notable improvements in average precision for detection. Consequently, the computational demands of LWSDNet can be easily met by the UAV terminal. As part of future research endeavors, the deployment of LWSDNet on UAV using Android Studio and neural network inference frameworks is aimed at enabling real-time detection of wheat scab in field environments. In addition, in order to improve the generalization of the model, based on the research in this paper, domain adaptation methods will continue to be used in the future to enable the model to be applied to wheat scab data of different varieties and periods.

5. Conclusions

This study introduced a novel lightweight method for detecting wheat scab using UAV. The method employed overlapping cropping of UAV wheat images to fulfill the input requirements of the network, and it applied image contrast enhancement techniques for illuminating the processing of the UAV images. To address the resource constraints of UAV, the study presented LWSDNet, a lightweight wheat scab detection network designed with MixConv. To enhance the network's capacity for extracting scab features, a scab feature enhancement module and a MixConv adaptive feature fusion module were introduced, enabling the detection of scab of various sizes. Additionally, a knowledge distillation strategy was proposed, involving the fusion of scab feature and response information. During network training, LWSDNet served as the student network, while a teacher network was devised based on it, allowing LWSDNet to acquire knowledge pertaining to the teacher network's features and output responses. Experimental comparisons with commonly used object detection networks demonstrated that LWSDNet achieved a higher AP in wheat scab detection, all while maintaining the minimum parameter size and computational complexity. The comparative analysis with lightweight object detection models confirmed that LWSDNet offered better performance in wheat scab detection, exhibiting smaller parameter counts and computational complexity.

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

Funding: This research was funded by Anhui Natural Science Foundation, grant number 2208085MC60, Natural Science Research Project of Anhui Provincial Education Department, grant number 2023AH050084, and National Natural Science Foundation of China, grant numbers 62273001, 32372632.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: The authors offer sincere thanks to the funding support provided by this research project and to all participants of the paper for their contributions.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Figueroa, M.; Hammond-Kosack, K.E.; Solomon, P.S. A review of wheat diseases—A field perspective. *Mol. Plant Pathol.* **2018**, *19*, 1523–1536. [CrossRef] [PubMed]
2. Dweba, C.C.; Figlan, S.; Shimelis, H.A.; Motaung, T.E.; Sydenham, S.; Mwadzingeni, L.; Tsilo, T.J. Fusarium head blight of wheat: Pathogenesis and control strategies. *Crop Prot.* **2017**, *91*, 114–122. [CrossRef]
3. Sun, Y.; Jiang, Z.; Zhang, L.; Dong, W.; Rao, Y. SLIC_SVM based leaf diseases saliency map extraction of tea plant. *Comput. Electron. Agric.* **2019**, *157*, 102–109. [CrossRef]
4. Singh, V.; Misra, A.K. Detection of plant leaf diseases using image segmentation and soft computing techniques. *Inf. Process. Agric.* **2017**, *4*, 41–49. [CrossRef]
5. Ahmed, I.; Yadav, P.K. A systematic analysis of machine learning and deep learning based approaches for identifying and diagnosing plant diseases. *Sustain. Oper. Comput.* **2023**, *4*, 96–104. [CrossRef]
6. Abdu, A.; Mokji, M.; Sheikh, U. Machine learning for plant disease detection: An investigative comparison between support vector machine and deep learning. *IAES Int. J. Artif. Intell.* **2020**, *9*, 670. [CrossRef]
7. Pallathadka, H.; Ravipati, P.; Sekhar Sajja, G.; Phasinam, K.; Kassanuk, T.; Sanchez, D.T.; Prabhu, P. Application of machine learning techniques in rice leaf disease detection. *Mater. Today Proc.* **2022**, *51*, 2277–2280. [CrossRef]
8. Kumar Nagothu, S.; Anitha, G.; Siranthini, B.; Anandi, V.; Siva Prasad, P. Weed detection in agriculture crop using unmanned aerial vehicle and machine learning. *Mater. Today Proc.* **2023**. [CrossRef]
9. Jaisakthi, S.M.; Mirunalini, P.; Thenmozhi, D.; Vatsala. Grape Leaf Disease Identification using Machine Learning Techniques. In Proceedings of the 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 21–23 February 2019; pp. 1–6.
10. Bao, W.; Zhao, J.; Hu, G.; Zhang, D.; Huang, L.; Liang, D. Identification of wheat leaf diseases and their severity based on elliptical-maximum margin criterion metric learning. *Sustain. Comput. Inform. Syst.* **2021**, *30*, 100526. [CrossRef]
11. Larijani, M.R.; Asli-Ardeh, E.A.; Kozegar, E.; Loni, R. Evaluation of image processing technique in identifying rice blast disease in field conditions based on KNN algorithm improvement by K-means. *Food Sci. Nutr.* **2019**, *7*, 3922–3930. [CrossRef]
12. Maski, P.; Thondiyath, A. Plant Disease Detection Using Advanced Deep Learning Algorithms: A Case Study of Papaya Ring Spot Disease. In Proceedings of the 2021 6th International Conference on Image, Vision and Computing (ICIVC), Qingdao, China, 23–25 July 2021; pp. 49–54.
13. Kerkech, M.; Hafiane, A.; Canals, R. Deep leaning approach with colorimetric spaces and vegetation indices for vine diseases detection in UAV images. *Comput. Electron. Agric.* **2018**, *155*, 237–243. [CrossRef]
14. Jakjoud, F.; Hatim, A.; Bouaaddi, A. Deep Learning application for plant diseases detection. In Proceedings of the 4th International Conference on Big Data and Internet of Things, Rabat, Morocco, 23–24 October 2020; pp. 1–6.
15. Soumo, A.; Ndeda, R.; Aoki, S.; Murungi, L. Comparison of Deep Learning Architectures for Late Blight and Early Blight Disease Detection on Potatoes. *Open J. Appl. Sci.* **2022**, *12*, 723–743. [CrossRef]
16. Oppenheim, D.; Shani, G.; Erlich, O.; Tsrur, L.L.J.P. Using Deep Learning for Image-Based Potato Tuber Disease Detection. *Phytopathology* **2019**, *109*, 1083–1087. [CrossRef]
17. Li, L.; Zhang, S.; Wang, B. Plant Disease Detection and Classification by Deep Learning—A Review. *IEEE Access* **2021**, *9*, 56683–56698. [CrossRef]
18. Bao, W.; Liu, W.; Yang, X.; Hu, G.; Zhang, D.; Zhou, X. Adaptively spatial feature fusion network: An improved UAV detection method for wheat scab. *Precis. Agric.* **2023**, *24*, 1154–1180. [CrossRef]
19. Li, J.; Qiao, Y.; Liu, S.; Zhang, J.; Yang, Z.; Wang, M. An improved YOLOv5-based vegetable disease detection method. *Comput. Electron. Agric.* **2022**, *202*, 107345. [CrossRef]
20. Zhang, Y.; Zhou, G.; Chen, A.; He, M.; Li, J.; Hu, Y. A precise apple leaf diseases detection using BCTNet under unconstrained environments. *Comput. Electron. Agric.* **2023**, *212*, 108132. [CrossRef]
21. Bao, W.; Zhu, Z.; Hu, G.; Zhou, X.; Zhang, D.; Yang, X. UAV remote sensing detection of tea leaf blight based on DDMA-YOLO. *Comput. Electron. Agric.* **2023**, *205*, 107637. [CrossRef]
22. Liu, D.; Gao, S.; Chen, P.; Cheng, L. A generality hard channel pruning with adaptive compression rate selection for HRNet. *Pattern Recognit. Lett.* **2023**, *168*, 107–114. [CrossRef]
23. Liu, J.; Zhuang, B.; Zhuang, Z.; Guo, Y.; Huang, J.; Zhu, J.; Tan, M. Discrimination-Aware Network Pruning for Deep Model Compression. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 4035–4051. [CrossRef]
24. Liu, Y.; Zhang, W.; Wang, J. Adaptive multi-teacher multi-level knowledge distillation. *Neurocomputing* **2020**, *415*, 106–113. [CrossRef]
25. Liu, X.; Zhu, Z. Knowledge Distillation for Object Detection Based on Mutual Information. In Proceedings of the 2021 4th International Conference on Intelligent Autonomous Systems (ICoIAS), Wuhan, China, 14–16 May 2021; pp. 18–23.
26. Dong, N.; Zhang, Y.; Ding, M.; Xu, S.; Bai, Y. One-stage object detection knowledge distillation via adversarial learning. *Appl. Intell.* **2022**, *52*, 4582–4598. [CrossRef]
27. Jafari, A.; Rezagholizadeh, M.; Sharma, P.; Ghodsi, A. Annealing Knowledge Distillation. Available online: <https://aclanthology.org/2021.eacl-main.212/> (accessed on 25 June 2024).
28. Bao, W.; Yang, X.; Liang, D.; Hu, G.; Yang, X. Lightweight convolutional neural network model for field wheat ear disease identification. *Comput. Electron. Agric.* **2021**, *189*, 106367. [CrossRef]

29. Yang, Z.; Li, Z.; Jiang, X.; Gong, Y.; Yuan, Z.; Zhao, D.; Yuan, C. Focal and Global Knowledge Distillation for Detectors. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 4633–4642.
30. Tan, M.; Le, Q.V. MixConv: Mixed Depthwise Convolutional Kernels. In Proceedings of the British Machine Vision Conference, Cardiff, UK, 9–12 September 2019.
31. Romero, D.W.; Brintjes, R.J.; Tomczak, J.M.; Bekkers, E.J.; Hoogendoorn, M.; Gemert, J.C. FlexConv: Continuous Kernel Convolutions with Differentiable Kernel Sizes. *arXiv* **2021**, arXiv:2110.08059.
32. Wang, W.; Wang, J. Double Ghost Convolution Attention Mechanism Network: A Framework for Hyperspectral Reconstruction of a Single RGB Image. *Sensors* **2021**, *21*, 666. [[CrossRef](#)] [[PubMed](#)]
33. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
34. Yang, H.; Deng, X.; Shen, H.; Lei, Q.; Zhang, S.; Liu, N. Disease Detection and Identification of Rice Leaf Based on Improved Detection Transformer. *Agriculture* **2023**, *13*, 1361. [[CrossRef](#)]
35. Yang, Z.; Yang, W.; Yi, J.; Liu, R. WH-DETR: An Efficient Network Architecture for Wheat Spike Detection in Complex Backgrounds. *Agriculture* **2024**, *14*, 961. [[CrossRef](#)]

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