Applied Soft Computing Journal 195 (2026) 115017

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Attention-based and context-aware knowledge distillation for enhancing crop disease detection

Xiangyuan Zhu [a], Taotao Mao [a], Jianguo Chen [b,*], Feifan Peng [a], Keqin Li [c,d]

[a] School of Computer Science and Software, Zhaoqing University, Zhaoqing 526061, China
[b] School of Software Engineering, Sun Yat-sen University, Zhuhai 519082, China
[c] College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China
[d] Department of Computer Science, State University of New York, New Paltz, NY 12561, USA

HIGHLIGHTS

- A pixel-wise attention mechanism based on the gradient of the loss function is designed to distinguish the importance of each pixel. This mechanism assigns varying weights to pixels according to their contribution to the training loss, contrasting with conventional methods that apply equal weights through channel-wise weighting.
- An ACKD approach is proposed to enhance semantic understanding through channel-wise attention, emphasize the foreground with spatial attention, identify salient pixels via pixel-wise attention, and capture global contextual relationships through contextual distillation.
- The effectiveness of the ACKD method is demonstrated through extensive experiments on four open-source datasets—Strawberry Diseases, Tomato-Village, Tomato Leaf Diseases, and Bean Plant Pathologies—and it achieves state-of-the-art performance when integrated into a popular one-stage you only look once (YOLO) detector, showcasing significant improvements in precision, recall, and mean average precision (mAP).

ARTICLE INFO

Keywords:
Crop disease detection
Deep learning
Knowledge distillation
Pixel-wise attention

ABSTRACT

Effective crop disease detection is essential for maintaining agricultural productivity, safeguarding food security, and ensuring the prosperity of farmers. However, the high computational demands of current deep neural network models often make them unsuitable for resource-constrained agricultural applications. Hence, there is a pressing demand for lightweight model designs for crop disease detection. To address this, we propose an attention-based and context-aware knowledge distillation (ACKD) approach that precisely evaluates pixel-wise contributions. Initially, a fine-grained pixel-wise attention mechanism identifies salient diseased pixels by emphasizing their gradients relative to the training loss. Secondly, spatial and channel-wise attention mechanisms capture broader spatial correlations and semantic details among pixels. Finally, contextual distillation extracts global contextual relationships, enabling the model to focus on informative pixels. By integrating these mechanisms, the ACKD approach transfers knowledge based on pixel-wise salience, thereby enhancing crop disease detection accuracy. Extensive experiments on the Strawberry Diseases, Tomato-Village, Tomato Leaf Diseases, and Bean Plant Pathologies datasets demonstrate that the ACKD achieves average precision (P) of 80.2%, recall (R) of 76.4%, F1 score of 78.1%, and mean average precision (mAP) of 80.4%. These results highlight the ACKD's strong generalization across diverse crop datasets and its robust potential for agricultural applications. To our knowledge, this is the first application of pixel-wise attention for knowledge distillation in agricultural scenarios.

1. Introduction

Tomatoes, beans, and strawberries are favored by consumers worldwide and pivotal to the agricultural industry. Cultivated in climates ranging from temperate to tropical, they generate considerable revenue and are key players in the global fresh produce market. Unfortunately, these delicate crops are challenging to grow due to their susceptibility to diseases like powdery mildew, leaf spots, bacterial infections, and

* Corresponding author.
Email addresses: zxycs@zqu.edu.cn (X. Zhu), mttfmt0@163.com (T. Mao), chenjg33@mail.sysu.edu.cn (J. Chen), pff355556695@163.com (F. Peng), lik@newpaltz.edu (K. Li).

https://doi.org/10.1016/j.asoc.2026.115017
Received 31 October 2024; Received in revised form 6 March 2026; Accepted 8 March 2026
Available online 10 March 2026

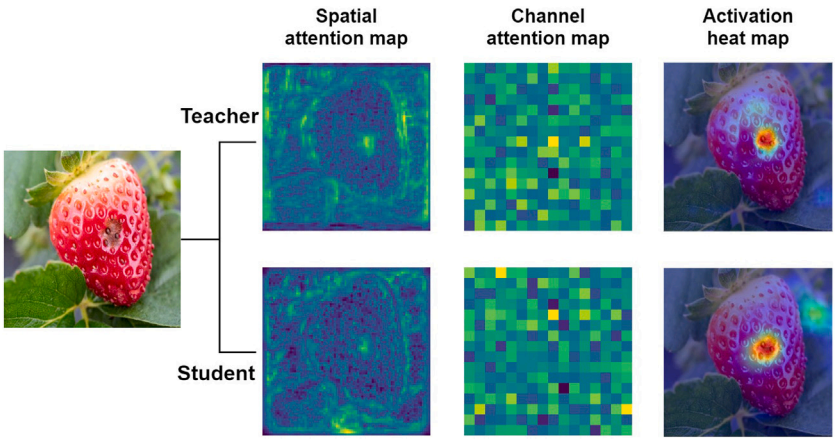


Fig. 1. Feature discrepancies in the teacher (YOLOv7) and student (YOLOv7-tiny) models: spatial, channel attention, and activation heat maps.

viral infections, which significantly reduce both yields and fruit quality [1]. While traditional manual inspection methods are time-consuming and subjective, computer vision has emerged as a transformative solution for automated plant disease monitoring in precision agriculture [2], with applications ranging from strawberry disease detection [3] and object counting [4] to geographical origin identification of Dendrobium officinale [5]. Nevertheless, these state-of-the-art deep learning models are computationally intensive, hindering their practical applicability in resource-constrained agricultural settings. Therefore, deploying lightweight models with fewer parameters is crucial in agricultural contexts for their ability to facilitate swift disease identification and management, ensuring both a productive harvest and the economic prosperity of farmers.

Knowledge distillation (KD) is a model compression technique in machine learning where a smaller network, known as the student, learns from a larger network, referred to as the teacher. The goal of KD is to transfer the teacher's knowledge to the student model, enabling it to perform tasks with reduced computational resources while retaining the teacher's accuracy. Although many knowledge distillation approaches have been tailored for object detection, resulting in significant advancements in fields such as infrared small target identification [6], autonomous driving [7], dense object detection [8], and facial expression recognition [9], research in crop disease detection remains limited.

Crop disease detection presents unique challenges that are not commonly found in general object detection tasks. Diseases can manifest in various forms, such as discoloration, lesions, and deformities, making their detection with conventional methods quite complex. Moreover, achieving precision and accurate localization is difficult, especially when the symptoms are subtle or blend in with the healthy plant tissues. The intricate growth environment of crops further compounds these challenges. Although lightweight models based on quantization-and-pruning strategies [10] and attention mechanisms [11] have been explored for agricultural image analysis, their deployment for crop disease detection in resource-constrained environments remains notably limited.

Given the complexities of crop disease detection, applying knowledge distillation emerges as a promising research area. The success of this approach hinges on effectively selecting features for distillation, which is essential for tackling the challenges mentioned. In Ref. [12], the authors introduced a Feature Interaction Module to enhance feature learning for the students through channel and spatial attention. Additionally, in Ref.[13], the authors leveraged an attention mechanism to identify similarities between the features of teacher and student networks. By transferring knowledge from a well-trained teacher model to a more efficient student model, we can develop systems that require fewer computational resources while maintaining high accuracy in detecting and localizing diseases. This exploration of knowledge distillation techniques could greatly enhance agricultural productivity and sustainability.

Diverse knowledge is encoded in feature maps extracted from intermediate layers of neural networks. To guide effective distillation, we visualize and analyze the feature discrepancies between the teacher and student detectors using spatial and channel attention maps, as well as activation heat maps derived from Grad-CAM [14]. This analysis facilitates the identification of differences in the learned features, as illustrated in Fig. 1. Upon examining the spatial attention (SA) maps, it is evident that the importance of spatial positions differs. The teacher model provides a clearer foreground, with more distinct boundaries and disease locations. Meanwhile, the minimal differences in the background indicate that not all spatial positions are equally significant to the models' performance. Furthermore, channel-wise attention (CA) varies significantly between the models, suggesting semantic and contextual disparities in information processing. In the last column of Fig. 1, the teacher model accurately concentrates on the diseased regions; however, the student model attends to irrelevant features near the diseased strawberry.

It becomes clear that discriminating salient regions within feature maps is crucial for identifying effective features for distillation. To this end, we propose an ACKD approach. The ACKD combines attention distillation, including spatial, channel-wise, and pixel-wise distillation (PA), with contextual distillation. Our approach evaluates the distinct contributions of pixels and channels, enabling the student model to locate salient features as indicated by pixel-wise and channel-wise attention calculations. Moreover, to capture semantic information more effectively, we integrate GcBlock [15] into our model to extract the contextual relationships among pixels throughout the entire image. To our knowledge, this is the first exploration of pixel-wise attention based on gradients for object-level knowledge distillation, particularly in agricultural scenarios. The main contributions of this paper are summarized as follows:

- We design a pixel-wise attention mechanism based on the gradient to distinguish the saliency of each pixel. This mechanism assigns varying weights to pixels according to their contribution to the training loss, contrasting with conventional methods that apply equal weights through channel-wise weighting.
- We propose an ACKD approach that enhances semantic understanding through channel-wise attention, emphasizes the foreground with spatial attention, identifies salient pixels via pixel-wise attention, and captures global contextual relationships by contextual distillation.
- We demonstrate the effectiveness of the ACKD method through extensive experiments on four open-source datasets—Strawberry

Diseases, Tomato-Village, Tomato Leaf Diseases, and Bean Plant Pathologies—and it achieves state-of-the-art performance when integrated into a popular one-stage you only look once (YOLO) detector.

The paper is structured as follows. Section 2 introduces the YOLO algorithm, crop disease detection, and knowledge distillation for object detection. Section 3 discusses the attention mechanism and contextual distillation in the ACKD algorithm. Section 4 evaluates our method against state-of-the-art approaches. Finally, Section 5 summarizes the paper.

2. Related work

2.1. Overview of YOLO algorithm

The YOLO algorithm is a state-of-the-art technique in object detection, revolutionizing the execution of real-time detection tasks. Unlike traditional object detection methods that require multiple computations to identify objects in an image, YOLO performs object detection as a single regression problem, directly predicting the bounding boxes and class probabilities from the input image. This is achieved by dividing the image into a grid of cells, with each cell assigned class probabilities and bounding box coordinates. YOLO's key innovation lies in its ability to simultaneously detect objects at multiple scales and locations within an image, significantly improving its speed and efficiency. The algorithm has undergone several iterations, starting with the initial YOLOv1 [16], followed by YOLOv2 [17], YOLOv3 [18], and the most recent versions, YOLOv7 [19] and YOLOv8 [20], each introducing enhancements in both accuracy and speed. YOLO's efficiency and robustness have positioned it as a favored choice for various applications, encompassing video surveillance, autonomous vehicles, and real-time interactive applications. In this paper, we choose a YOLOv7 detector to evaluate our proposed method.

2.2. Crop disease detection

Crop disease detection is a classic application in agricultural object detection, with significant advancements in recent years. Various deep-learning models have been explored to enhance the accuracy and efficiency of crop disease identification. As reported in Ref. [21], an improved YOLO model was constructed for cotton disease and pest recognition, employing an efficient channel attention mechanism, Focal Loss function, and Hard-Swish activation function. In the study presented in Ref. [22], a progressive learning and region proposal module was designed to reduce the negative effects of features resembling vegetable diseases. According to Ref. [23], a convolutional neural network (CNN) model was presented for detecting maize diseases. This model leverages an auxiliary classifier generative adversarial network (ACGAN) to expand the training dataset, thereby improving the model's ability to generalize. Additionally, the model employs transfer learning to adapt to the complexities of various cultivation environments. The Internet of Things (IoT) platform StrawberryTalk, as detailed in Ref. [24], exemplifies the integration of technology for image-based strawberry disease detection, improving performance by mitigating wind interference and optimizing camera configurations. Despite advancements in object detection technology, agricultural applications still face challenges such as variability in lighting conditions, complex backgrounds, and the diversity of crop types. The requirements for real-time processing and lightweight algorithms further complicate the deployment of efficient detection systems.

2.3. Knowledge distillation for object detection

Early works in knowledge distillation introduced the concept of using the teacher's soft targets as an additional source of information to guide the student's training. Since then, the field has evolved to encompass a variety of distillation strategies, such as decoupled knowledge distillation for classification and regression [25], dynamic distillation [26], student-centered distillation [27], localization distillation [28], and multi-teacher distillation [29]. These strategies aim to capture different aspects of the teacher's knowledge. While most of these approaches have been presented for image classification, significant advancements have been made in applying knowledge distillation to object detection. Object detection is inherently more complex than image classification because it involves not only the classification of objects but also their precise localization. In particular, crop disease detection via knowledge distillation remains an under-explored area of research.

Knowledge is typically encoded in feature maps, which makes feature-based distillation a preferred method for transferring knowledge from the intermediate layers of CNNs. The FGD method, introduced in Ref. [30], combines focal and global distillation approaches for object detection. It employs focal distillation through a scale mask to extract knowledge from the foreground and background separately. Then, it leverages the GcBlock [15], a block designed to learn global interactions between these extracted features. GKD-BMFI, presented in Ref. [31], was designed to weigh feature importance with gradients in object-level knowledge distillation. It utilizes the gradients of the detection loss relative to the feature maps. Unlike FGD, GKD-BMFI identifies valuable distillation regions by focusing on ground-truth bounding boxes and their neighboring pixels. However, after calculating the gradients, GKD-BMFI performs a global average pooling to derive a single weight for each channel, potentially obscuring the importance of individual pixels within the channel. The IRKD method, proposed in Ref.[6] for infrared small target detection, incorporates a unified channel-spatial attention module to distinguish feature distinctions between the teacher and student models. Concurrently, GcBlock is applied to capture contextual information.

Despite the advancements, these state-of-the-art approaches treat all pixels within identified knowledge regions equally, overlooking the distinct contributions each pixel can make to accurate classification and precise localization. This uniform treatment significantly hinders the effectiveness of knowledge transfer, highlighting the need for a more sophisticated approach that recognizes the individual significance of each pixel.

3. Methodology

This section presents the overall architecture of the proposed ACKD. As shown in Fig. 2, the ACKD comprises two key components: attention-based distillation and context-aware distillation, detailed in Sections 3.1 and 3.2, respectively.

3.1. Attention-based distillation

We introduce an attention-based distillation designed to address the limitations of uniform pixel treatment. By prioritizing pixels based on their task-specific attention, our method enhances the model's ability to learn from the most informative pixels and streamlines the knowledge transfer process.

3.1.1. Pixel-wise attention

In this section, we introduce a gradient-based salient pixel-wise attention (PA) mechanism to discern pixel-level knowledge. It is important to recognize that different pixels within the same feature map play distinct roles in knowledge distillation. To identify which pixels should receive greater attention during distillation, we draw inspiration from Grad-CAM [14], a technique known for its ability to highlight important regions in images. The contribution of each pixel in the feature maps can be determined by the loss gradient of that pixel. This process is analogous to assigning a weight to each pixel, where the gradient signifies the pixel's importance. Similar to how humans focus on salient regions when perceiving images, the gradient of each pixel acts as an indicator that influences the loss. In essence, the salient pixels are characterized by the

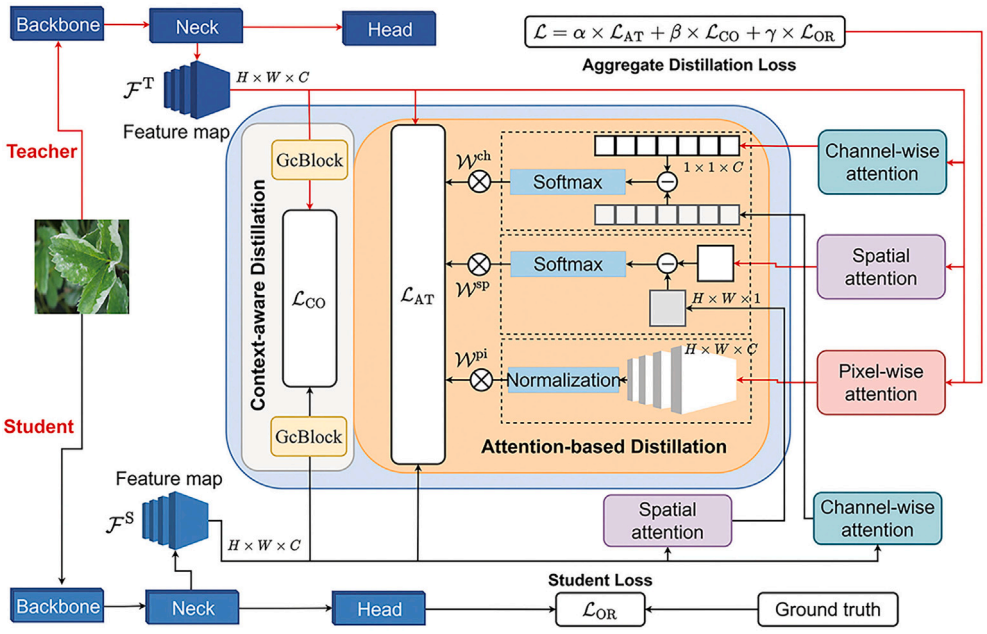


Fig. 2. Illustration of the proposed ACKD method.

highest gradients, signifying their critical role in the neural network's decision-making process.

The pixel-wise attention $\mathcal{W}^{pi}$ is defined as:

$$\mathcal{W}^{pi} = \text{Norm}\left(\frac{\partial \mathcal{L}}{\partial x}\right) \tag{1}$$

where $\mathcal{L}$ indicates the aggregate distillation loss of the model, x represents a pixel in a feature map $\mathcal{F}^T$, the superscript T denotes the teacher model, and $\text{Norm}(\cdot)$ denotes the min-max normalization operation. After performing Eq. (1), $\mathcal{W}^{pi}$ matches the dimension of $\mathcal{F}^T$ and has values in the range $[0, 1]$. Applying this weight to $\mathcal{F}^T$ through multiplication generates a pixel-wise saliency map that highlights the most informative pixels.

The teacher applies the pixel-wise attention based on the aggregate distillation loss $\mathcal{L}$ and its pixels x in the feature map $\mathcal{F}^T$. The resulting pixel-wise salience map, with dimensions $H \times W \times C$, indicates the importance of each pixel in the teacher's feature map.

Proposition 1. *The salience of the pixel x is directly proportional to the magnitude of the gradient $\frac{\partial \mathcal{L}}{\partial x}$.*

Proof. Let $\mathcal{L}$ be the aggregate distillation loss function defined on $\mathbb{R}^{H \times W \times C}$, and assume that it is differentiable at the pixel $x = (x_1, x_2, \ldots, x_C)$ of the feature map $\mathcal{F}^T$, where H, W, and C denote the height, width, and number of channels of $\mathcal{F}^T$, respectively. Using the first-order Taylor expansion at x, the change in $\mathcal{L}$ around x can be approximated as:

$$\mathcal{L}(x + \Delta x) \approx \mathcal{L}(x) + \frac{\partial \mathcal{L}}{\partial x} \cdot \Delta x$$

where $\cdot$ represents inner product, and the gradient vector $\frac{\partial \mathcal{L}}{\partial x}$ is given by:

$$\frac{\partial \mathcal{L}}{\partial x} = \left(\frac{\partial \mathcal{L}}{\partial x_1}(x), \frac{\partial \mathcal{L}}{\partial x_2}(x), \ldots, \frac{\partial \mathcal{L}}{\partial x_C}(x)\right)$$

The change in the loss function $\mathcal{L}$ around x can be expressed as:

$$\Delta \mathcal{L} = \mathcal{L}(x + \Delta x) - \mathcal{L}(x)$$

Using the Taylor expansion approximation, we have:

$$\Delta \mathcal{L} \approx \frac{\partial \mathcal{L}}{\partial x} \cdot \Delta x$$

The salience of the pixel x is determined by its impact on the loss function $\mathcal{L}$. From the above approximation, it is evident that the change in the loss function $\Delta \mathcal{L}$ is directly proportional to the gradient $\frac{\partial \mathcal{L}}{\partial x}$ at the pixel x. Specifically:

- If $\left|\frac{\partial \mathcal{L}}{\partial x}\right|$ is large, a small change Δx in the pixel value x will result in a larger change $\Delta \mathcal{L}$ in the loss function. This indicates that the pixel x is highly salient, as it has a significant impact on $\mathcal{L}$.
- Conversely, if $\left|\frac{\partial \mathcal{L}}{\partial x}\right|$ is small, the change $\Delta \mathcal{L}$ in the loss function will be smaller for the same Δx. This suggests that the pixel x is less salient, as it has a minimal impact on $\mathcal{L}$.

Therefore, the salience of a pixel x is directly proportional to the magnitude of the gradient $\frac{\partial \mathcal{L}}{\partial x}$. This relationship provides a formal justification for using gradient-based methods to measure pixel salience in the context of a differentiable loss function $\mathcal{L}$. □

Utilizing gradients offers three main advantages, particularly in knowledge distillation and model compression tasks. First, gradients help detect salient image pixels, which most strongly influence the model's predictions. This ability is crucial for pinpointing the discriminative areas of the image, allowing for effective knowledge transfer from a teacher model to a student model.

Second, the gradient-based approach is beneficial for weakly supervised disease localization. By emphasizing the pixels that most influence the model's output, we can effectively identify diseased areas within the image, even when labeled data are scarce. This targeted focus improves localization accuracy by concentrating on the pixels most indicative of the presence of disease.

Lastly, the calculation of gradients is computationally efficient and instructively beneficial. It enables the student model to learn the decision boundaries with fewer parameters by focusing on the gradients of the teacher model's outputs. This approach maintains high accuracy while reducing complexity, as the student model is trained to recognize and respond to the most critical features of the input data.

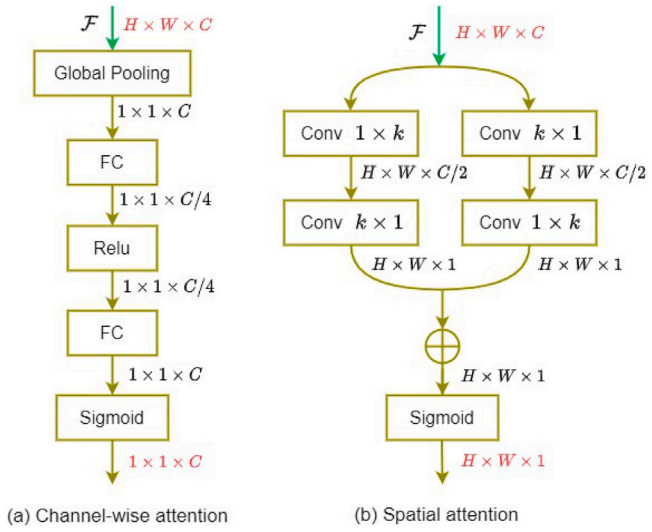


Fig. 3. Illustration of channel-wise attention and spatial attention.

3.1.2. Channel-wise attention

In CNNs, the number of output channels corresponds to the number of filters. Applying diverse filters is an effective technique for extracting semantic information at different levels, ranging from simple features like edges and textures to more complex patterns like shapes and objects. To preserve high-level features that are sensitive to multiple scales and receptive fields, channel-wise attention (CA) [32] is applied, as depicted in Fig. 3(a). The CA mechanism allocates larger weights to channels that exhibit high responses to salient objects.

Let the input feature map $\mathcal{F}$ have dimensions of $H \times W$ (height by width) and C channels. The global average pooling (GAP) operation P can be represented as:

$$P(c) = \frac{1}{H \cdot W} \sum_{i=1}^{H} \sum_{j=1}^{W} \mathcal{F}(c)_{ij} \tag{2}$$

where c denotes the channel index ($c \in [1, C]$), and $\mathcal{F}(c)_{ij}$ represents the pixel value at the i-th row and j-th column in the c-th channel of the input feature map. $P(c)$ represents the output value of the c-th channel after GAP.

Two fully connected (FC) operations are sequentially applied to extract channel-wise semantics. Channel-wise attention $Ch(\mathcal{F})$ is defined as:

$$Ch(\mathcal{F}) = \text{Sigmoid}(\text{FC}_2(\text{ReLU}(\text{FC}_1(P(c), W_1)), W_2)) \tag{3}$$

where $\text{FC}_1(\cdot)$ and $\text{FC}_2(\cdot)$ are fully connected layers parameterized by W_1 and W_2, rectified linear unit (ReLU) is the activation function, and the Sigmoid function normalizes the output to a range between 0 and 1. This process yields a vector of size $1 \times 1 \times C$, with values indicating the importance of each channel.

As shown in Fig. 2, both the teacher and the student apply channel-wise attention based on their respective feature maps $\mathcal{F}^T$ and $\mathcal{F}^S$. The resulting vectors, each with a size of $1 \times 1 \times C$, represent the channel-wise attention maps for the teacher and student.

3.1.3. Spatial attention

In crop disease detection, images are characterized by rich foreground information and intricate backgrounds. To reduce background interference, we aim to extract detailed foreground knowledge by focusing on salient objects and minimizing distracting textures. Therefore, we adopt spatial attention [32] to prioritize the foreground, rather than treating all spatial pixels equally, thereby facilitating the extraction of features crucial for identifying crop diseases.

Given the feature map $\mathcal{F}$, we design two parallel branches to extract spatial attention, each applying consecutive convolutional operations with kernel sizes of $1 \times k$ and $k \times 1$. Spatial attention $S(\mathcal{F})$ is defined as:

$$\begin{aligned} S_1 &= \text{Conv}_2(\text{Conv}_1(\mathcal{F}, W_3), W_4) \\ S_2 &= \text{Conv}_1(\text{Conv}_2(\mathcal{F}, W_4), W_3) \\ S(\mathcal{F}) &= \text{Sigmoid}(S_1 + S_2) \end{aligned} \tag{4}$$

where $\text{Conv}_1(\cdot)$ and $\text{Conv}_2(\cdot)$ denote convolutional operations parameterized by W_3 and W_4, respectively. S_1 and S_2 represent the features from the two branches, respectively. The element-wise addition of S_1 and S_2 is followed by applying the Sigmoid function to normalize the output to a range between 0 and 1, resulting in a spatial attention map of size $H \times W \times 1$.

As shown in Fig. 2, both the teacher and the student apply spatial attention based on their respective feature maps $\mathcal{F}^T$ and $\mathcal{F}^S$. The resulting spatial attention maps, each with a size of $H \times W \times 1$, highlight the importance of different spatial locations within the feature maps.

3.1.4. Attention-based distillation loss

Based on Eqs. (3) and (4), channel saliency $\mathcal{W}^{\mathbf{ch}}$ and spatial saliency $\mathcal{W}^{\mathbf{sp}}$ are defined as:

$$\begin{aligned} \mathcal{W}^{\mathbf{ch}} &= C \cdot \text{Softmax}\left(\left|Ch(\mathcal{F}^{\mathbf{T}}) - Ch(\mathcal{F}^{\mathbf{S}})\right| / \tau\right) \\ \mathcal{W}^{\mathbf{sp}} &= H \cdot W \cdot \text{Softmax}\left(\left|S(\mathcal{F}^{\mathbf{T}}) - S(\mathcal{F}^{\mathbf{S}})\right| / \tau\right) \end{aligned} \tag{5}$$

where $S(\cdot)$ and $Ch(\cdot)$ denote the spatial and channel saliency functions, respectively, τ is a scaling factor, and the superscripts T and S denote the teacher and student models, respectively.

There are distinct differences between the features of students and teachers. Instead of directly using the teacher's features for guidance, we utilize $\mathcal{W}^{\text{pi}}$, $\mathcal{W}^{\text{sp}}$, and $\mathcal{W}^{\text{ch}}$, which are derived from pixel-wise attention, spatial saliency, and channel saliency, to instruct the students. Consequently, the attention-based loss $\mathcal{L}_{\text{AT}}$ is defined as:

$$\mathcal{L}_{\text{AT}} = \frac{1}{H \cdot W \cdot C} \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{c=1}^{C} \mathcal{W}^{\text{pi}}_{i,j,c} \mathcal{W}^{\text{sp}}_{i,j} \mathcal{W}^{\text{ch}}_{c} \| \mathcal{F}^{\text{T}}_{i,j,c} - f(\mathcal{F}^{\text{S}}_{i,j,c}) \|_2^2 \tag{6}$$

where $\mathcal{F}^{\text{S}}$ and $\mathcal{F}^{\text{T}}$ denote the feature maps of the student and teacher models, respectively. The function $f(\cdot)$ is used to align $\mathcal{F}^{\text{S}}$ and $\mathcal{F}^{\text{T}}$. The weighting factors $\mathcal{W}^{\text{pi}}$, $\mathcal{W}^{\text{sp}}$, and $\mathcal{W}^{\text{ch}}$ emphasize the relative importance of different pixel locations and channels in the loss calculation.

As shown in Fig. 2, the attention-based distillation loss $\mathcal{L}_{\text{AT}}$ is computed using three saliency weights: the channel saliency $\mathcal{W}^{\text{ch}}$, the spatial saliency $\mathcal{W}^{\text{sp}}$, and the pixel-wise attention $\mathcal{W}^{\text{pi}}$. $\mathcal{W}^{\text{ch}}$ and $\mathcal{W}^{\text{sp}}$ are derived from the differences between the teacher and student feature maps. $\mathcal{W}^{\text{pi}}$ is obtained by normalizing the gradient of the aggregate distillation loss $\mathcal{L}$ with respect to the pixels in the teacher's feature map. These weights are then combined to guide the student model by emphasizing important features.

3.2. Context-aware distillation

Context is vital in object detection, particularly for identifying crop diseases in complex agricultural scenes cluttered with similar appearances. It enhances the model's understanding by analyzing how the arrangement of pixels, visual features, and semantic meaning influence spatial relationships. This contextual knowledge is crucial for accurately localizing diseases within an image. To capitalize on this advantage, we propose a context-aware distillation method that transfers contextual knowledge from the teacher model's feature maps to the student model, enhancing detection accuracy.

We apply a GcBlock [15] to extract global contextual information from an image, facilitating the contextual knowledge transfer from the

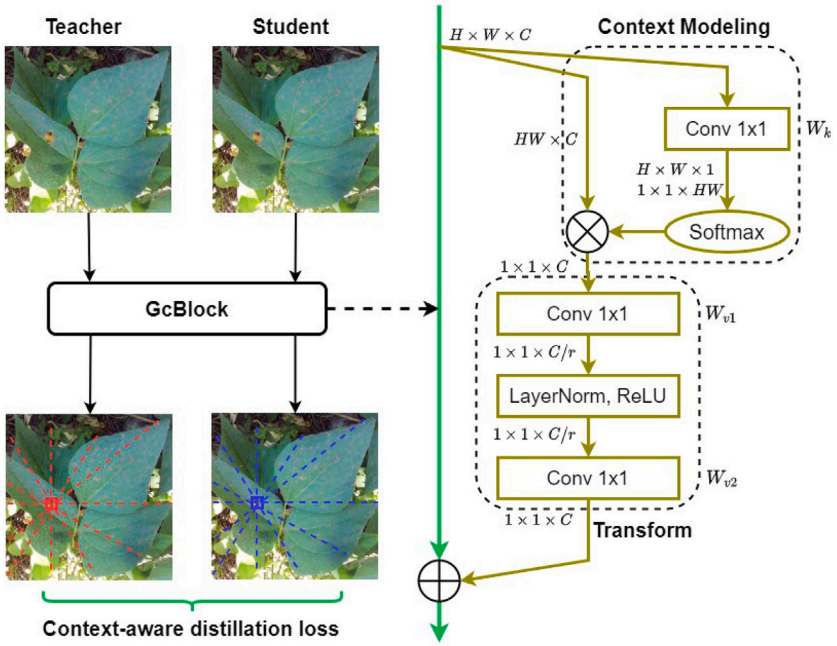


Fig. 4. Illustration of context-aware distillation with GcBlock.

teacher to the student, as shown in Fig. 4. The context-aware distillation loss is defined below.

$$\mathcal{G}(\mathcal{F}) = \mathcal{F} + W_{v2}\mathrm{ReLU}\left(\mathrm{LN}\left(W_{v1}\sum_{j=1}^{N_p}\frac{e^{w_k\mathcal{F}_j}}{\sum_{m=1}^{N_p}e^{w_k\mathcal{F}_m}}\mathcal{F}_j\right)\right)$$

$$\mathcal{L}_{\mathrm{CO}} = \sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{c=1}^{C}\|\mathcal{G}(\mathcal{F}^{\mathrm{T}}_{i,j,c}) - \mathcal{G}(\mathcal{F}^{\mathrm{S}}_{i,j,c})\|_2^2 \tag{7}$$

where W_k, W_{v1}, and W_{v2} represent convolutional layers, LN and ReLU denote layer normalization and activation functions, respectively. $\mathcal{G}$ denotes the GcBlock transformation, and N_p is the positional size of the feature map $\mathcal{F}$.

As shown in Fig. 2, the context-aware distillation process involves computing the GcBlock transformations for both the teacher and student models. The context-aware distillation loss $\mathcal{L}_{\mathrm{CO}}$ is then calculated based on the differences between these transformations, facilitating the transfer of contextual knowledge from the teacher to the student.

3.3. Detection loss of the student model

In general, any object detection network can serve as the student model. For clarity, we select the YOLOv7-tiny network as our student model for this explanation. The loss function for YOLOv7-tiny, denoted as $\mathcal{L}_{\mathrm{OR}}$, is a sum of the objectness loss $\mathcal{L}_{\mathrm{obj}}$, classification loss $\mathcal{L}_{\mathrm{cls}}$, and bounding box regression loss $\mathcal{L}_{\mathrm{box}}$, each weighted by their respective importance ratios. The student loss $\mathcal{L}_{\mathrm{OR}}$ can be expressed as:

$$\mathcal{L}_{\mathrm{OR}} = \lambda_1 \times \mathcal{L}_{\mathrm{obj}} + \lambda_2 \times \mathcal{L}_{\mathrm{cls}} + \lambda_3 \times \mathcal{L}_{\mathrm{box}} \tag{8}$$

where λ_1, λ_2, and λ_3 are hyperparameters that determine the relative contribution of each loss component to the total loss during training.

Objectness loss $\mathcal{L}_{\mathrm{obj}}$ and classification loss $\mathcal{L}_{\mathrm{cls}}$ are calculated using binary cross-entropy (BCE) between the predicted scores and the corresponding true labels:

$$\begin{aligned}\mathcal{L}_{\mathrm{obj}} &= \mathrm{BCE}\,(\hat{y}_{\mathrm{obj}}, y_{\mathrm{obj}})\\ \mathcal{L}_{\mathrm{cls}} &= \mathrm{BCE}\,(\hat{y}_{\mathrm{cls}}, y_{\mathrm{cls}})\end{aligned} \tag{9}$$

where $\hat{y}_{\mathrm{obj}}$ and $\hat{y}_{\mathrm{cls}}$ are predicted objectness and classification scores, and y_{obj} and y_{cls} are their true labels, respectively. BCE$(\cdot)$ denotes the binary cross-entropy function.

Bounding box regression loss $\mathcal{L}_{\mathrm{box}}$ is determined by the CIoU (complete intersection over union) [19] metric between the predicted bounding box coordinates $\hat{y}_{\mathrm{box}}$ and the true bounding box coordinates y_{box}:

$$\mathcal{L}_{\mathrm{box}} = 1 - \mathrm{CIoU}(\hat{y}_{\mathrm{box}}, y_{\mathrm{box}}) \tag{10}$$

For the details of CIoU, refer to [19].

As shown in Fig. 2, the student model's detection loss $\mathcal{L}_{\mathrm{box}}$ is computed based on the predicted results from the detection head and ground truth, highlighting the contributions of objectness, classification, and bounding box regression losses.

3.4. Aggregate distillation loss

An attention-based, context-aware distillation method has been designed to guide the student model's learning. The student model is trained using the following aggregate loss function for distillation.

$$\mathcal{L} = \alpha\mathcal{L}_{\mathrm{AT}} + \beta\mathcal{L}_{\mathrm{CO}} + \gamma\mathcal{L}_{\mathrm{OR}} \tag{11}$$

where α, β, and γ are hyperparameters that balance the loss function components. $\mathcal{L}_{\mathrm{AT}}$ denotes the attention-based distillation loss, $\mathcal{L}_{\mathrm{CO}}$ represents the context-aware distillation loss, and $\mathcal{L}_{\mathrm{OR}}$ refers to the detection loss of the student model. As shown in Fig. 2, $\mathcal{L}$ is used to train the student model and calculate the pixel-wise saliency $\mathcal{W}^{\mathrm{pi}}$. The pipeline of the ACKD model is described by Algorithm 1.

Algorithm 1 outlines the training procedure of the ACKD model. Within each training epoch, the algorithm processes each batch of the training dataset through three essential steps: (1) Forward Pass (lines 4-6): It computes the feature maps $\mathcal{F}^{\mathrm{T}}$ and $\mathcal{F}^{\mathrm{S}}$ from the teacher and student models, respectively. (2) Attention Weight Calculation (lines 7-9): It calculates the PA, CA, and SA attention weights, which guide the student model in capturing critical features from the teacher model. (3) Loss Computation (lines 10-14): It computes the attention-based distillation loss $\mathcal{L}_{\mathrm{AT}}$, context-aware distillation loss $\mathcal{L}_{\mathrm{CO}}$, and the original detection loss $\mathcal{L}_{\mathrm{OR}}$, aggregating them into a total loss $\mathcal{L}$. After processing each

Algorithm 1 Attention-based and context-aware knowledge distillation (ACKD) model.

Input:
The pre-trained teacher model $\mathcal{N}_T$, training set $\mathcal{D}$, balancing hyperparameters α, β, γ, and learning rate η.
Output:
Student model $\mathcal{N}_S$ with parameter θ_S.
1: Initialize θ_S randomly; Initialize losses: $\mathcal{L}_{AT} \leftarrow 0$, $\mathcal{L}_{CO} \leftarrow 0$, $\mathcal{L}_{OR} \leftarrow 0$, $\mathcal{L} \leftarrow 0$; Initialize hyperparameters: $\alpha \leftarrow 1$, $\beta \leftarrow 2.5 \times 10^{-5}$, $\gamma \leftarrow 1$, $\eta \leftarrow 0.01$;
2: **for** each epoch **do**
3: **for** each batch of training data $(x, y) \in \mathcal{D}$ **do**
4: Forward pass:
5: Compute feature maps $\mathcal{F}^T$ from the teacher model $\mathcal{N}_T$;
6: Compute feature maps $\mathcal{F}^S$ from the student model $\mathcal{N}_S$;
7: Compute attention weights:
8: Pixel-wise attention weight $\mathcal{W}^{pi}$ by Eq. (1);
9: Channel-wise attention weight $\mathcal{W}^{ch}$ and spatial attention weight $\mathcal{W}^{sp}$ by Eq. (5);
10: Compute losses:
11: Attention-based distillation loss $\mathcal{L}_{AT}$ by Eq. (6);
12: Context-aware distillation loss $\mathcal{L}_{CO}$ by Eq. (7);
13: Original detection loss $\mathcal{L}_{OR}$ by Eq. (8);
14: Compute total loss by Eq. (11): $\mathcal{L} \leftarrow \alpha\mathcal{L}_{AT} + \beta\mathcal{L}_{CO} + \gamma\mathcal{L}_{OR}$;
15: **end for**
16: Update parameters: $\theta_S \leftarrow \theta_S - \eta\nabla_{\theta_S}\mathcal{L}$;
17: **end for**
18: **return** The student model $\mathcal{N}_S$ with updated parameters θ_S.

batch, the student model's parameters θ_S are updated via backpropagation (line 16). This iterative process continues across multiple epochs until the model converges, ultimately yielding the optimized student model.

The computational complexity of these steps is crucial for the efficiency and scalability of the ACKD model. During the forward pass, the complexity of extracting feature maps $\mathcal{F}^T$ and $\mathcal{F}^S$ is determined by the respective teacher and student models. For attention weight calculation, PA and CA have a complexity of $O(H \times W \times C)$, while SA has $O(H \times W \times C^2 \times k)$. In terms of loss computation, $\mathcal{L}_{AT}$ and $\mathcal{L}_{CO}$ have $O(H \times W \times C)$, and $\mathcal{L}_{OR}$ depends on the specific student model. Here, H, W, and C denote the height, width, and channel size of the features, respectively, and k denotes the kernel size in spatial attention. Despite incorporating multiple attention mechanisms and a global context block, the overall complexity remains manageable and scalable, ensuring enhanced performance without prohibitive computational overhead.

While the above complexity analysis establishes theoretical bounds, we further quantify the empirical training cost. The gradient-based pixel attention requires backpropagating gradients $\nabla_{\mathcal{W}^{pi}}\mathcal{L}$ through Eq. (6). This gradient computation is efficiently realized by treating $\mathcal{W}^{pi}$ as a differentiable tensor within PyTorch's computational graph. During backpropagation, gradients for all spatial and channel dimensions are computed simultaneously via optimized tensor operations, eliminating the need for explicit per-element gradient coding. Empirical evaluation shows ACKD introduces moderate computational overhead, with 23.1% additional training time per epoch and 3.5% increased GPU memory consumption relative to the student baseline, confirming its practicality for resource-constrained applications. Detailed ablation studies are provided in Table 6 of Section 4.4.1.

4. Experiments

The experiments employ the advanced YOLOv7 as the teacher model and its compact variant, YOLOv7-tiny, as the student model for crop disease object detection in agricultural settings. YOLOv7 is renowned for its high performance, while YOLOv7-tiny is appreciated for its smaller size and lower computational demands, making it ideal for resource-constrained applications. We aim to assess the effectiveness of transferring sophisticated features from the teacher to the student model using the ACKD method.

4.1. Experimental design and metrics

Our evaluations are conducted on a server equipped with an Intel(R) Xeon(R) Gold 6330 processor and three NVIDIA A100 graphics processing units (GPUs), running Ubuntu 20.04. The experiments utilize PyTorch 1.12.0 as the deep learning framework, with compute unified device architecture (CUDA) version 11.3 for GPU acceleration. The training schedule comprises 600 epochs, a batch size of 16, an initial learning rate of 0.01, and a cosine decay factor of 0.2 to adjust the learning rate over time. We employ stochastic gradient descent (SGD) as the optimization algorithm, resizing input images to 640×640 pixels. The distillation temperature, a hyperparameter that influences the softening of logits in the distillation process, is set to 0.8 based on ablation experiments. For the aggregate distillation loss, the hyperparameters are empirically set to $\alpha = 1$, $\beta = 2.5 \times 10^{-5}$, and $\gamma = 1$ to balance loss magnitudes and ensure stable training. Specifically, $\alpha = \gamma = 1$ leverages the natural scaling of $\mathcal{L}_{AT}$ (normalized by $H \times W \times C$) and $\mathcal{L}_{OR}$ (YOLOv7 default), whereas the small β prevents the unnormalized $\mathcal{L}_{CO}$ from overwhelming the total loss and causing gradient instability. The teacher and student detection loss parameters are set to $\lambda_1 = 0.7$, $\lambda_2 = 0.3$, and $\lambda_3 = 0.05$, corresponding to the default settings in the YOLOv7 model. Spatial attention is modulated with $k = 9$, which is adopted from the reference [32].

The detection performance of the models is evaluated using precision (P), recall (R), mean average precision at IoU 0.5 (mAP@0.5), and mean average precision at IoU ranging from 0.5 to 0.95 (mAP@0.5:0.95). Computational efficiency is evaluated in terms of model parameters, training time per epoch, peak GPU memory, and inference speed (frames per second, FPS).

4.2. Datasets

Building upon this experimental foundation, we evaluate the performance of ACKD across four diverse open-source datasets: Strawberry Diseases [33], Tomato-Village [34], Tomato Leaf Diseases [35], and Bean Plant Pathologies [36]. These datasets, which offer a rich collection of images and annotations for various plant diseases, are detailed below and serve as benchmarks for assessing the model's training and testing efficacy. Fig. 5 depicts representative examples of the four key datasets.

The **Strawberry Diseases Dataset** [33] consists of 1943 images and categorizes seven types of strawberry diseases: gray mold fruit, powdery mildew fruit, anthracnose fruit, powdery mildew leaf, leaf spot, calcium deficiency leaf, and angular leaf spot. It includes images collected by the artificial intelligence laboratory (AI Lab) at Jeonbuk National University, as well as images sourced from the Internet by us. The dataset is divided into a training set of 1547 images and a validation set of 396 images.

Moving on to tomato diseases, the **Tomato-Village Dataset** [34], comprising 14,368 images and 161,175 bounding boxes, is divided into training and validation subsets with 11,493 and 2875 images, respectively. It captures real-world contexts for detecting six types of tomato diseases: late blight, leaf miner, magnesium deficiency, nitrogen deficiency, potassium deficiency, and spotted wilt virus.

Another dataset focused on tomato diseases is the **Tomato Leaf Diseases Dataset** [35], which consists of 706 images, divided into training and validation subsets with 645 and 61 images, respectively. Unlike the Tomato-Village dataset, which focuses on a broader range of tomato diseases, this dataset is specifically designed to detect seven types of conditions affecting tomato leaves, including bacterial spot, early blight, late blight, leaf mold, target spot, and black spot, as well as to identify healthy leaves.

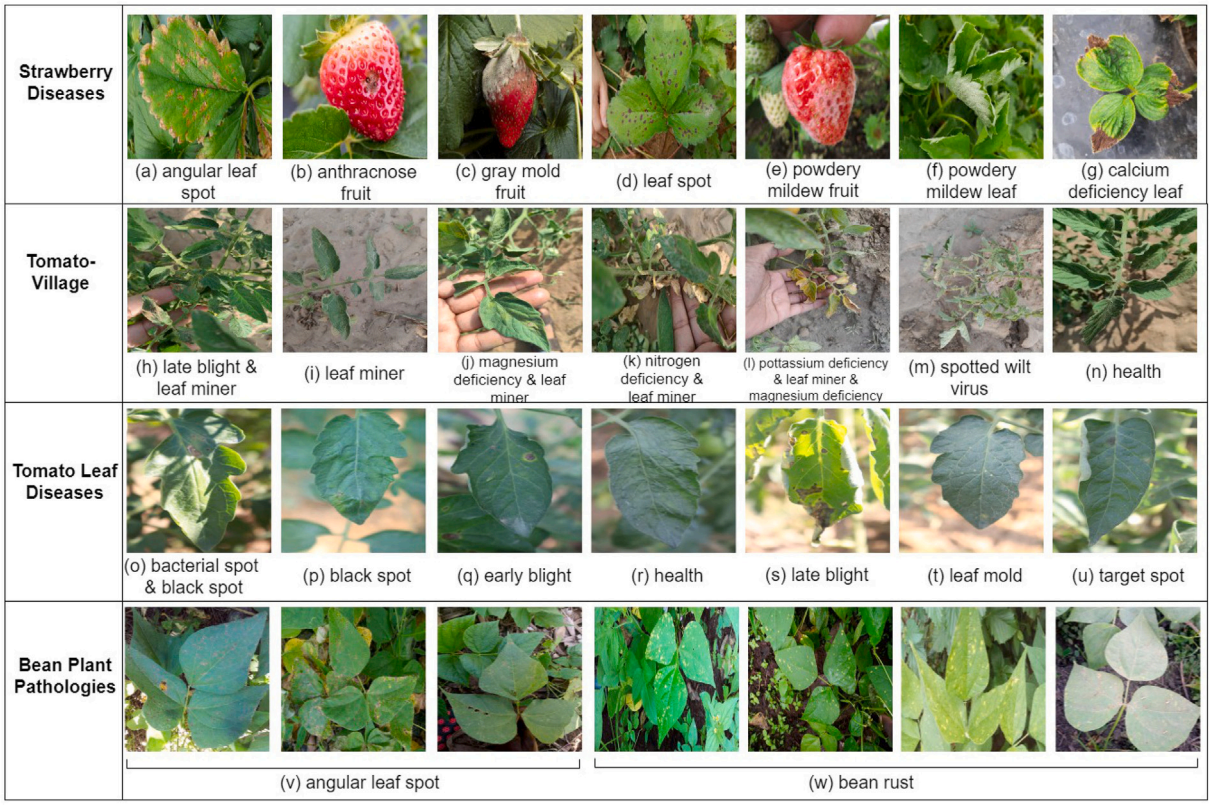


Fig. 5. Representative examples of the four datasets.

Table 1
Details of the four datasets.

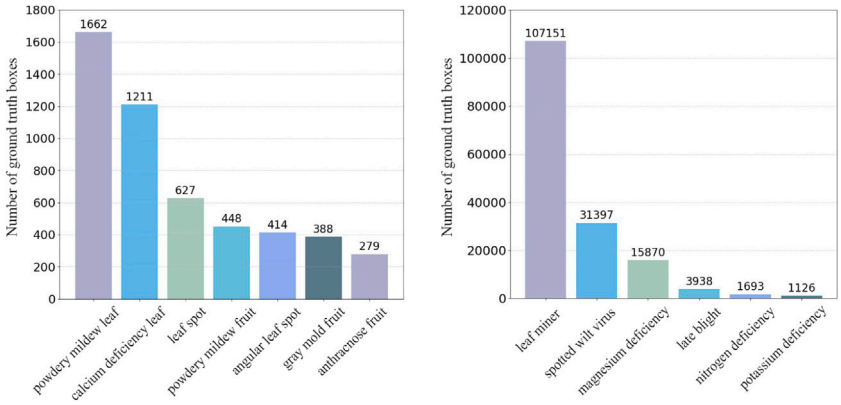
Dataset	Number of categories	Training set	Validation set	Ground truth boxes			Average boxes per image
				Small	Medium	Large	
Strawberry Diseases	7	1547	396	38	940	4051	2.59
Tomato-Village	6	11,493	2875	32,893	92,193	36,137	11.21
Tomato Leaf Diseases	7	645	61	1538	517	265	3.29
Bean Plant Pathologies	2	3512	990	418	2363	13,272	3.56

Lastly, the **Bean Plant Pathologies Dataset** [36], a subset of the Makerere University Beans Image Dataset, consists of 4988 images curated for diagnosing diseases in bean crops. It includes 3512 training images, 990 validation images, and 486 test images. The images are divided into two main disease classes: angular leaf spot and bean rust, both affecting bean leaves, along with healthy images to improve the model's detection capabilities.

The aforementioned four datasets provide a comprehensive testbed for evaluating the ACKD algorithm. The fields in Table 1 highlight the datasets' diversity and complexity in the following four aspects, which are critical for assessing the algorithm's robustness and adaptability.

- **Image Scales**: The datasets vary significantly in size, ranging from the smallest *Tomato Leaf Diseases* dataset (645 training images and 61 validation images) to the largest *Tomato-Village* dataset (11,493 training images and 2875 validation images). This diversity allows us to assess the algorithm's performance across different data magnitudes.
- **Object Size Distribution**: The datasets exhibit diverse object size distributions, with the *Tomato Leaf Diseases* dataset containing 1538 small boxes, the *Strawberry Diseases* dataset having 4051 large boxes, the *Bean Plant Pathologies* dataset comprising 13,272 large boxes, and the *Tomato-Village* dataset featuring 92,193 medium boxes. This variation challenges the algorithm to adapt to different object sizes.
- **Object Density**: The datasets differ in object density, with the *Tomato-Village* dataset having the highest average number of boxes per image (11.21) and the *Strawberry Diseases* dataset having the lowest (2.59). This diversity evaluates the algorithm's ability to handle varying disease complexities.
- **Application Challenges**: In practical agricultural scenarios, the number of objects often varies significantly across categories. For instance, as shown in Fig. 6, the *Strawberry Diseases* dataset includes 1662 objects in the powdery mildew leaf category and only 279 objects in the anthracnose fruit category. Similarly, the *Tomato-Village* dataset contains 107,151 objects in the leaf miner category but only 1126 objects in the potassium deficiency category. This disparity exacerbates the challenge for the ACKD algorithm to adapt effectively.

To address the challenges posed by the diverse characteristics of these datasets, we employ ten data augmentation techniques in preprocessing data. These techniques enhance the generalization ability of the ACKD model by exposing it to a variety of real-world agricultural scenarios. Specifically, the techniques include HSV augmentation (for Hue, Saturation, and Value), translation, scale adjustment,



(a) Strawberry Diseases Dataset (b) Tomato-Village Dataset

Fig. 6. Ground truth boxes in the Strawberry Diseases and Tomato-Village datasets.

shear transformation, perspective transformation, horizontal flip, mosaic augmentation, mixup augmentation, copy-paste augmentation, and paste-in augmentation.

4.3. Comparison study

4.3.1. Comparison with state-of-the-art methods

To evaluate the effectiveness of our ACKD method, we conduct a comparative analysis with state-of-the-art object detection distillation methods, including FGD [30], IRKD [6], and GKD-BMFI [31]. These methods represent the latest advancements in the field, each employing innovative technical approaches to enhance the student model's performance.

As shown in Table 2, ACKD consistently enhances the student model's detection capability across all four agricultural datasets, yielding improvements of 0.3%–2.2% in F1 score and 1.4%–2.9% in mAP@0.5. These gains effectively narrow the performance gap between the lightweight student and heavyweight teacher models, demonstrating robust cross-domain knowledge transfer despite varying disease patterns and visual complexities.

Beyond quantitative accuracy gains, ACKD demonstrates superior stability, with consistent improvements in comprehensive metrics (F1 and mAP) despite dataset-specific trade-offs in precision and recall. In contrast, the compared methods occasionally suffer performance degradation: FGD exhibits declines in mAP@0.5:0.95 on the Strawberry Diseases and Bean Plant Pathologies datasets, while IRKD shows slight precision drops on the Tomato-Village dataset. Notably, on the Tomato Leaf Diseases dataset—where the student baseline unexpectedly surpasses its teacher (a weak teacher scenario)—ACKD uniquely achieves further improvements in recall and mAP@0.5, whereas the competing approaches consistently suffer performance degradation. This robustness highlights ACKD's unique advantage in handling such challenging teachers, a critical capability for practical deployment where teacher performance is variable or unreliable.

The consistent F1 score improvements across diverse datasets reflect distinct precision–recall optimization patterns. These results indicate that ACKD achieves favorable precision–recall balances across different dataset characteristics: it improves precision (up to 5.1%) for fine-grained discrimination tasks such as strawberry disease identification, while enhancing recall (up to 7.3%) for comprehensive coverage

Table 2
Comparative analysis of ACKD with three alternative methods (%) and McNemar's test.

Dataset	Model	P	R	F1 score	mAP@0.5	mAP@0.5:0.95	Z value/significant?
Strawberry Diseases	YOLOv7 (T)	**88.2**	**84.9**	**86.5**	**91.9**	**71.0**	2.09/Yes
	YOLOv7-tiny (S)	84.5	81.5	83.0	87.9	65.6	2.04/Yes
	FGD [30]	86.0 (+1.5)	82.8 (+1.3)	84.4 (+1.4)	88.7 (+0.8)	64.7 (−0.9)	2.02/Yes
	IRKD [6]	84.3 (−0.2)	85.9 (+4.4)	85.1 (+2.1)	90.3 (+2.4)	66.0 (+0.4)	2.91/Yes
	GKD-BMFI [31]	85.2 (+0.7)	81.0 (−0.5)	83.0 (+0.0)	87.7 (−0.2)	65.2 (−0.4)	2.66/Yes
	Ours	**89.6 (+5.1)**	**81.0 (−0.5)**	**85.1 (+2.1)**	**90.7 (+2.8)**	**66.0 (+0.4)**	–
Tomato-Village	YOLOv7 (T)	**97.6**	**96.0**	**96.8**	**98.3**	**76.9**	22.75/Yes
	YOLOv7-tiny (S)	88.2	82.3	85.1	86.5	51.3	14.83/Yes
	FGD [30]	**89.9 (+1.7)**	**85.4 (+3.1)**	**87.6 (+2.5)**	89.0 (+2.5)	49.5 (−1.8)	3.31/Yes
	IRKD [6]	87.4 (−0.8)	84.7 (+2.4)	86.0 (+0.9)	87.5 (+1.0)	47.9 (−3.4)	8.76/Yes
	GKD-BMFI [31]	88.1 (−0.1)	85.1 (+2.8)	86.6 (+1.5)	88.3 (+1.8)	48.4 (−2.9)	5.84/Yes
	Ours	89.5 (+1.3)	85.3 (+3.0)	87.3 (+2.2)	**89.4 (+2.9)**	**49.7 (−1.6)**	–
Bean Plant Pathologies	YOLOv7 (T)	**66.2**	**62.2**	**64.1**	**66.7**	**37.1**	3.36/Yes
	YOLOv7-tiny (S)	73.9	52.4	61.3	62.7	30.7	4.12/Yes
	FGD [30]	**73.3 (−0.6)**	52.5 (+0.1)	61.2 (+0.1)	61.0 (−1.7)	29.3 (−1.4)	5.59/Yes
	IRKD [6]	67.4 (−6.5)	57.3 (+4.9)	61.9 (+0.6)	63.6 (+0.9)	30.4 (−0.3)	8.02/Yes
	GKD-BMFI [31]	66.5 (−7.4)	**60.8 (+8.4)**	**63.5 (+2.2)**	64.6 (+1.9)	32.0 (+1.3)	7.11/Yes
	Ours	67.6 (−6.3)	59.7 (+7.3)	63.4 (+2.1)	**65.0 (+2.3)**	**32.3 (+1.6)**	–
Tomato Leaf Diseases	YOLOv7 (T)	**82.6**	75.2	**78.7**	74.3	44.2	2.33/Yes
	YOLOv7-tiny (S)	75.2	**77.7**	76.4	**75.2**	**47.2**	2.12/Yes
	FGD [30]	**82.1 (+6.9)**	72.6 (−5.1)	**77.1 (+0.7)**	73.3 (−1.9)	45.7 (−1.5)	2.11/Yes
	IRKD [6]	76.6 (+1.4)	71.1 (−6.6)	73.7 (−2.7)	74.9 (−0.3)	**47.0 (−0.2)**	2.14/Yes
	GKD-BMFI [31]	77.1 (+1.9)	73.5 (−4.2)	75.3 (−1.1)	72.9 (−2.3)	43.6 (−3.6)	2.00/Yes
	Ours	74.0 (−1.2)	**79.6 (+1.9)**	76.7 (+0.3)	**76.6 (+1.4)**	45.2 (−2.0)	–

scenarios like bean plant pathology detection. The dataset-specific gains in precision and recall—as reflected in the unified F1 scores—demonstrate that ACKD achieves an effective detection balance, mitigating both false positives (unnecessary treatment costs) and false negatives (disease propagation risks), thereby enhancing agricultural diagnostic utility across diverse disease patterns.

4.3.2. Standardized McNemar's test

To rigorously assess the statistical significance of our model's performance, we conduct the standardized McNemar's test to compare ACKD with state-of-the-art models across the four datasets. Specifically, we obtain the predictions from both our model and the comparison models, and then construct a 2×2 contingency table. Based on this table, we calculate the McNemar's test statistic Z value. We set the significance level at 0.05, corresponding to a Z value of 1.96. We then compare the Z values obtained in our experiments to determine the significance of the results. The Z value is defined as:

$$Z_{ij} = \frac{|r_{ij} - r_{ji}|}{\sqrt{r_{ij} + r_{ji}}} \tag{12}$$

where r_{ij} represents the number of bounding boxes correctly detected by model i but incorrectly detected by model j.

As shown in the last column of Table 2, the Z values from McNemar's test are greater than 1.96. This indicates that the results obtained from ACKD and the other models are statistically different at the 95% confidence level.

4.3.3. Quantitative analysis for disease categories

Table 3 presents the mAP@0.5 scores for individual disease categories across the four datasets. ACKD demonstrates consistent superiority in fine-grained disease recognition, achieving optimal or near-optimal performance in the majority of categories. Notably, it surpasses the teacher model in challenging cases such as strawberry angular leaf spot (97.5% vs. 95.9%) and tomato late blight (92.8% vs. 86.1%), indicating effective capture of discriminative features for subtle morphological patterns.

Furthermore, the compared methods exhibit high variance across categories. For instance, FGD excels in nitrogen deficiency, whereas it degrades substantially in angular leaf spot. IRKD yields significant gains in leaf mold yet suffers severe drops in black spot. GKD-BMFI improves anthracnose fruit rot detection while exhibiting degradation in powdery mildew fruit. In contrast, ACKD elevates mAP@0.5 in all categories except black spot, demonstrating robust knowledge transfer across diverse disease patterns. These results confirm that ACKD's gains generalize consistently across disease categories, validating its applicability for comprehensive crop health monitoring systems.

Under weak teacher scenarios, as exemplified by the Tomato Leaf Diseases dataset, ACKD further demonstrates its robustness. In early blight and target spot—where the student baseline surpasses the teacher (82.1% vs. 81.4% and 76.4% vs. 65.9%, respectively)—ACKD maintains comparable performance (81.1% and 76.4%), whereas the compared methods suffer notable degradation. This stability underscores ACKD's capability to preserve valid knowledge even when teacher models are suboptimal.

4.3.4. Efficiency and generalization capability of ACKD

To assess the generalization capability of the ACKD algorithm, we conduct experiments on four disease categories: angular leaf spot, bean rust, late blight, and target spot. Angular leaf spot and late blight serve as cross-dataset categories, appearing across the Strawberry Diseases and Bean Plant Pathologies datasets, and the Tomato-Village and Tomato Leaf Diseases datasets, respectively. Bean rust and target spot are evaluated as single-dataset categories within the Bean Plant Pathologies and Tomato Leaf Diseases datasets exclusively. This design validates the algorithm's performance under varying conditions.

As shown in Table 4, ACKD consistently surpasses teacher models while matching or exceeding student baselines in mAP@0.5. For cross-dataset categories, ACKD achieves superior precision and comprehensive detection accuracy despite marginal recall trade-offs. For single-dataset categories, it maintains strong performance in primary metrics, with favorable precision–recall balance. These patterns confirm robust generalization across diverse agricultural settings.

Given the encouraging results across various disease categories, we further assess the versatility of the ACKD algorithm by integrating it into different object detection frameworks. Specifically, we conduct experiments using the YOLOv7 and RetinaNet frameworks on the Strawberry Disease dataset. As shown in Table 5, ACKD consistently achieves the highest mAP@0.5 across all frameworks, surpassing both student baselines and competing methods. Notably, ACKD achieves the highest performance for both YOLOv7-tiny and YOLOv7-ResNet50, despite their differing backbones but shared detection framework. Similarly, ACKD leads in mAP@0.5 for both YOLOv7-ResNet50 and RetinaNet-ResNet50, which share the same backbone but employ distinct detection

Table 3
Quantitative analysis of ACKD and three alternative methods on mAP@0.5 for disease categories in four datasets (%).

Dataset	Disease Category	Teacher (YOLOv7)	Student (YOLOv7-tiny)	ACKD	FGD	IRKD	GKD-BMFI
Strawberry Diseases	angular leaf spot	**95.9**	95.5	**97.5**	92.2	94.9	89.4
	anthracnose fruit rot	**88.2**	80.9	85.2	83.1	85.9	**87.1**
	calcium deficiency leaf	**92.8**	92.6	**92.6**	92.1	92.5	90.2
	gray mold	**91.9**	88.5	90.0	**93.6**	88.3	89.3
	leaf spot	**94.0**	89.7	**93.2**	90.0	92.7	85.8
	powdery mildew fruit	**90.3**	87.7	87.9	85.7	**88.5**	84.8
	powdery mildew leaf	**91.1**	89.1	89.3	88.6	**89.5**	87.1
Tomota- Village	late blight	**97.9**	89.1	**91.6**	**91.6**	90.9	90.5
	leaf miner	**97.2**	80.1	**82.4**	81.9	79.4	80.5
	magnesium deficiency	**98.8**	91.4	92.9	**93.1**	92.6	92.3
	nitrogen deficiency	**97.7**	87.4	92.5	**93.0**	90.9	92.6
	potassium deficiency	**99.6**	83.3	**86.8**	84.5	82.7	84.3
	spotted wilt virus	**98.8**	87.7	**90.1**	89.8	88.7	89.5
Tomato Leaf Diseases	bacterial spot	**74.5**	68.3	**74.5**	**74.5**	68.3	74.5
	black spot	**65.7**	62.7	59.1	**62.4**	55.9	59.0
	early blight	81.4	**82.1**	**81.1**	77.3	75.4	78.6
	late blight	86.1	**91.9**	**92.8**	92.3	92.5	88.8
	leaf mold	**46.8**	45.5	46.0	32.5	**55.3**	45.7
	target spot	65.9	**76.4**	76.4	74.5	**77.1**	63.9
Bean Plant Pathologies	angular leaf spot	**71.4**	64.7	**75.9**	62.7	67.9	68.1
	bean rust	**62.1**	59.5	**62.3**	59.3	59.3	61.1

Table 4
Performance and generalization of ACKD on four categories (%).

Disease Category	Model	P	R	mAP@0.5	mAP@0.5:0.95
angular leaf spot	ACKD	**74.6**	69.9	**75.9**	**43.9**
	YOLOv7 (T)	65.0	**70.0**	71.4	41.2
	YOLOv7-tiny (S)	72.9	56.0	64.7	32.1
bean rust	ACKD	70.1	**55.8**	**62.3**	31.2
	YOLOv7 (T)	67.4	54.4	62.1	**33.1**
	YOLOv7-tiny (S)	**74.9**	48.7	59.5	29.3
late blight	ACKD	**86.5**	88.7	**92.8**	**73.9**
	YOLOv7 (T)	77.3	**93.8**	86.1	66.1
	YOLOv7-tiny (S)	86.0	93.1	91.9	72.7
target spot	ACKD	**72.7**	**83.3**	**76.4**	32.2
	YOLOv7 (T)	54.5	**83.3**	65.9	**35.5**
	YOLOv7-tiny (S)	71.4	**83.3**	**76.4**	32.2

Table 5
Detection performance (%), model sizes (M), and inference speed (FPS) of tested frameworks on the Strawberry Diseases Dataset.

Method	P	R	F1 Score	mAP@ 0.5	mAP@ 0.5:0.95	Parameters	FPS
YOLOv7 (T)	**88.2**	**84.9**	**86.5**	**91.9**	**71.0**	36.51	134.1
YOLOv7-tiny (S)	84.5	81.5	83.0	87.9	65.6	**6.03**	**199.2**
FGD [30]	86.0	82.8	84.4	88.7	64.7	6.19	194.5
IRKD [6]	84.3	85.9	85.1	90.3	66.0	6.20	195.5
GKD-BMFI [31]	85.2	81.0	83.0	87.7	65.2	6.19	197.4
Ours	**89.6**	**81.0**	**85.1**	**90.7**	**66.0**	6.82	194.9
YOLOv7-ResNet101 (T)	**86.6**	77.7	81.9	**86.2**	**64.7**	67.40	69.2
YOLOv7-ResNet50 (S)	84.3	78.2	81.1	85.0	63.2	**48.41**	**99.9**
FGD [30]	86.0	**78.3**	82.0	85.2	63.2	48.57	97.7
IRKD [6]	87.0	76.4	81.4	85.4	63.3	48.58	98.5
GKD-BMFI [31]	85.7	78.2	81.8	85.2	63.3	48.57	98.2
Ours	**86.5**	77.9	**82.0**	**85.7**	**63.5**	49.20	96.9
Retina-ResNet101 (T)	**85.6**	**81.3**	**83.4**	**86.6**	**67.5**	57.40	76.3
Retina-ResNet50 (S)	85.8	71.7	78.1	83.4	60.8	**34.01**	**120.6**
FGD [30]	82.5	77.5	79.9	84.8	62.2	34.17	118.9
IRKD [6]	**87.1**	75.3	**80.8**	85.0	62.7	34.18	115.4
GKD-BMFI [31]	81.2	**77.9**	79.5	83.3	60.2	34.17	119.8
Ours	82.5	77.7	80.0	**85.5**	**63.5**	34.80	114.3

frameworks. These findings confirm that the performance of ACKD is invariant to the choice of detection framework or backbone, thereby highlighting its broad applicability and effectiveness.

ACKD achieves substantial parameter reduction with no degradation in F1 scores or mAP values, maintaining model sizes and inference speeds comparable to those of lightweight student baselines while reducing parameters by up to 81.3% and improving throughput by at least 40.0% relative to heavyweight teachers. This synthesis of model compactness and computational efficiency validates that ACKD enhances detection accuracy without sacrificing real-time performance, rendering it ideally suited for resource-constrained agricultural applications, such as mobile or embedded systems, where high accuracy and low computational overhead are essential.

4.3.5. Interpreting visual results

Figs. 7 to 10 present activation heat maps for the four crop disease benchmarks, validating the quantitative analysis and demonstrating ACKD's consistent knowledge transfer. The first column shows ground truth with diseased areas marked in red, while columns two through seven display Grad-CAM-generated heat maps for different models. These maps vividly indicate the diseased regions the models target, underscoring the challenges in detecting small, crowded, and overlapping diseases in agricultural settings. For example, the fifth row of Fig. 7 shows an image with three disease categories, where much of the infected area is substantially overlapped or intersected.

In complex agricultural settings, ACKD effectively identifies diseases by focusing on salient regions, as demonstrated by its detection of small disease objects like bacterial spot, black spot, early blight, and leaf mold in the Tomato Leaf Diseases Dataset (Fig. 8). ACKD's pixel-wise attention mechanism, which uses gradients to distinguish and localize the most salient pixels within small, overlapping disease areas, significantly contributes to its outstanding performance in detecting subtle diseases.

Furthermore, ACKD excels at detecting medium to large disease objects, such as angular leaf spot and bean rust in the Bean Plant Pathologies Dataset, as shown in Fig. 10. It surpasses the teacher model in specific instances, including row 1 of Fig. 9, rows 2 and 6 of Fig. 8, and rows 3, 5, and 7 of Fig. 10. This superior performance is attributed to ACKD's ability to minimize noisy pixel interference and focus intently on foreground diseased regions.

Overall, comparative analysis and visualization across the four datasets demonstrate that ACKD consistently improve the student model's performance. This improvement is attributed to its novel attention-based and context-aware approach, which identifies salient and discriminative regions within feature maps and employs pixel-wise attention based on gradients. The integration of spatial and channel-wise attention with the GcBlock module enhances the model's precision in identifying disease areas of varying sizes. These capabilities render ACKD a promising solution for broad application in agricultural practices, particularly in detecting disease objects of different sizes.

4.4. Ablation analysis

Ablation experiments are conducted on the Strawberry Diseases Dataset to evaluate the impact of pixel-wise attention (PA), spatial attention (SA), channel-wise attention (CA), GcBlock, and the distillation temperature τ.

4.4.1. Impact of PA, SA, CA, and GcBlock

To systematically assess attention mechanisms, individual and combined effects are evaluated. As shown in Table 6, GcBlock and SA provide the most substantial individual gains in mAP@0.5, followed by PA and CA. Combination strategies outperform isolated mechanisms, with full integration (PA + SA + CA + GcBlock) achieving optimal accuracy. Other combinations (PA + CA, SA + CA + GcBlock) also demonstrate improvements, validating multi-mechanism benefits. These accuracy gains incur varying computational costs: PA and SA emerge as primary contributors to training time (approximately 19.2% increase), while CA and GcBlock impose modest overheads (approximately 11.5% and 15.4%, respectively). Notably, memory consumption remains stable across all configurations, with the full integration requiring merely 0.9 GB additional GPU memory, representing a 3.5% increase over the baseline.

In summary, attention mechanisms substantially enhance distillation efficacy with manageable resource overhead. Multi-component integration proves more effective than isolated modules, aligning with findings in Refs. [6,30,31] and validating the advantages of diverse attention strategies for optimizing model performance.

4.4.2. Impact of the distillation temperature τ

Table 7 reveals that ACKD's performance exhibits sensitivity to the temperature hyperparameter τ. Optimal balance across evaluation metrics is achieved at $\tau = 0.8$, where the model attains peak mAP@0.5:0.95, despite marginally lower mAP@0.5 compared to $\tau = 0.5$. This configuration yields the most robust detection performance, effectively optimizing the precision–recall trade-off while maximizing comprehensive localization accuracy.

Notably, $\tau = 0.8$ demonstrates particular efficacy in small object detection, as evidenced by substantial improvements in both AP_s (15.0%) and AR_s (90.0%) compared to other settings where small-object metrics remain at zero. This marked enhancement in fine-grained feature discrimination underscores the critical role of temperature calibration for detecting subtle pathological patterns. Consequently, $\tau = 0.8$ is recommended as the default configuration for achieving balanced, high-performance detection in practical agricultural applications.

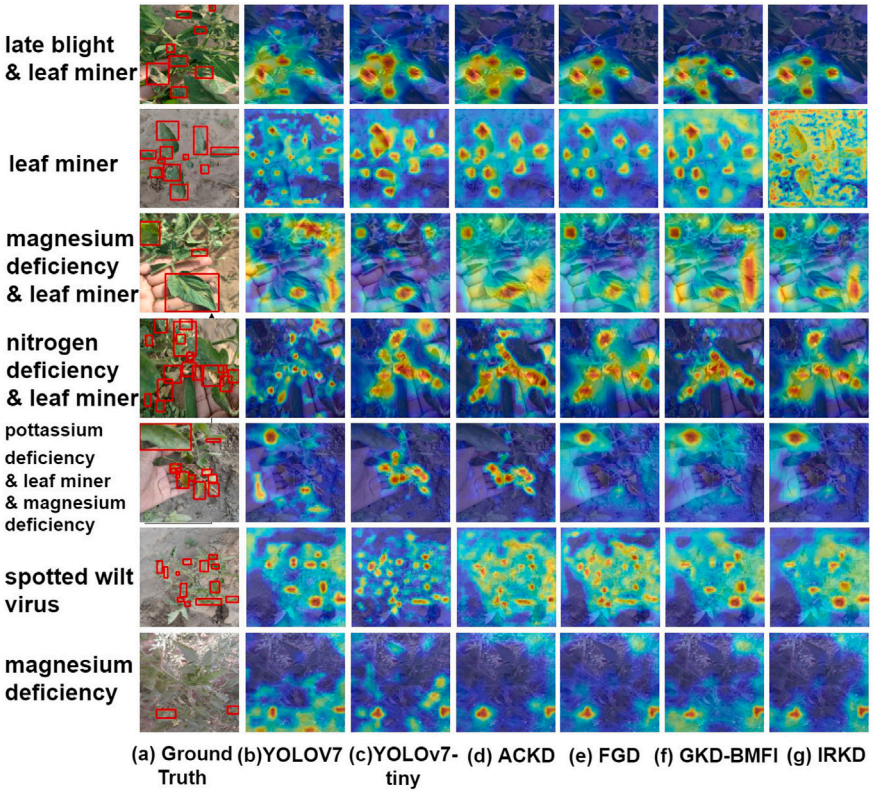


Fig. 7. Visualization of representative images from the tomato-village dataset.

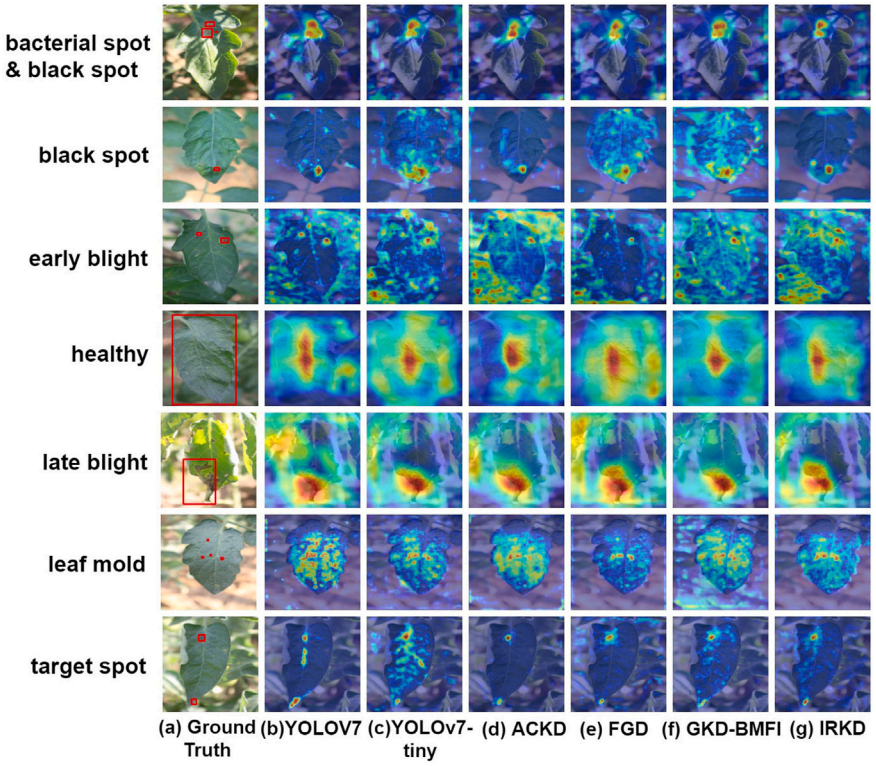


Fig. 8. Visualization of representative images from the tomato leaf diseases dataset.

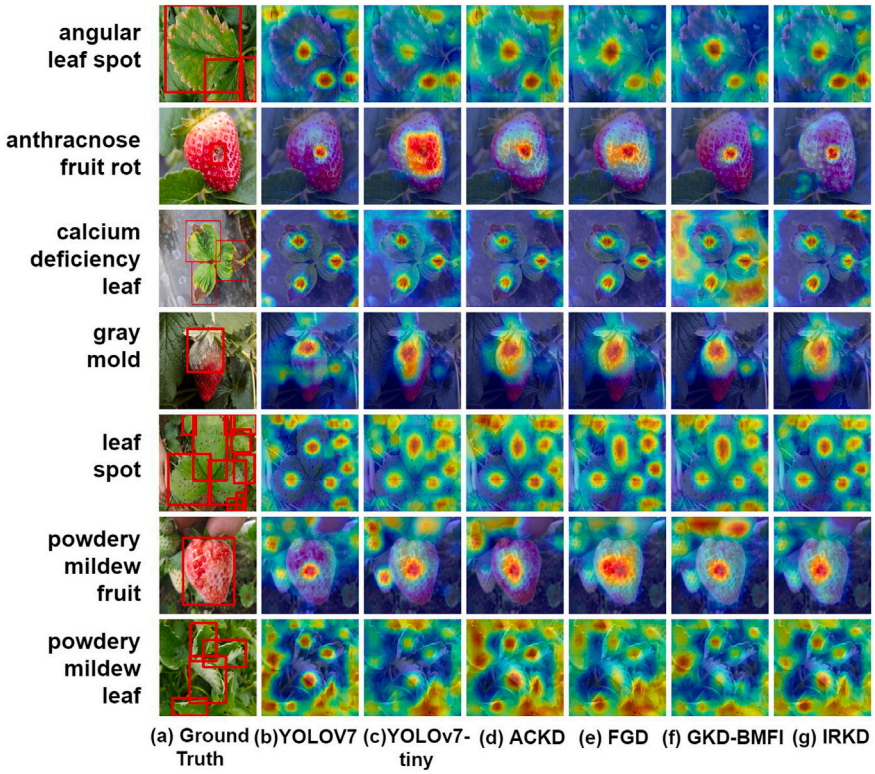


Fig. 9. Visualization of representative images from the Strawberry Diseases Dataset.

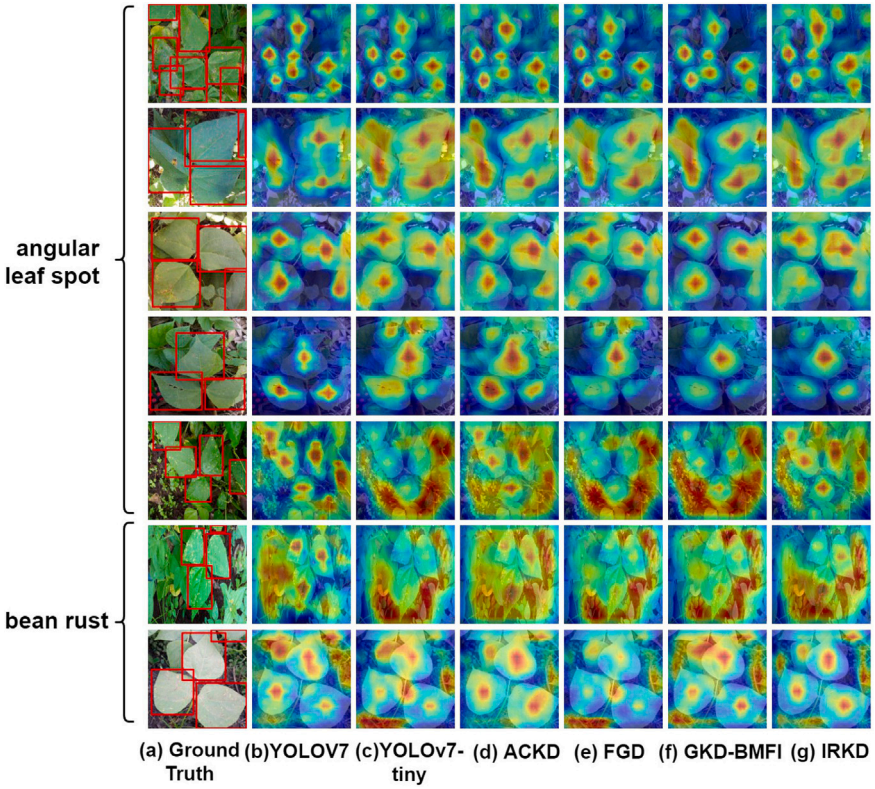


Fig. 10. Visualization of representative images from the bean plant pathologies dataset.

Table 6
Impact of PA, SA, CA, and GcBlock on detection metrics (%) and computational overhead (time/epoch and GPU memory) for Strawberry Diseases Dataset.

Student (YOLOv7-tiny)	Attention combination	P	R	F1 score	mAP@0.5	mAP@0.5:0.95	Time/Epoch (s)	GPU memory (GB)
✓	–	84.5	81.5	83.0	87.9	65.6	26.0	25.7
✓	PA	87.4	82.1	84.7	88.8	66.3	31.0	25.6
✓	SA	87.8	82.2	84.9	89.5	68.2	31.0	25.6
✓	CA	87.1	80.7	83.8	88.1	67.5	29.0	26.0
✓	GcBlock	87.8	80.7	84.1	89.6	67.6	30.0	26.0
✓	PA + SA	85.8	83.7	84.7	89.2	68.0	31.0	26.0
✓	PA + CA	84.6	85.6	85.1	90.0	68.5	30.0	25.6
✓	PA + GcBlock	86.4	82.3	84.3	89.5	67.2	31.0	25.8
✓	SA + CA	85.5	84.2	84.8	89.7	68.2	31.0	26.4
✓	CA + GcBlock	87.4	82.5	84.9	89.7	68.2	31.0	26.5
✓	PA + SA + CA	88.3	82.9	85.5	89.5	67.7	31.0	27.0
✓	PA + SA + GcBlock	89.1	81.3	85.0	89.8	67.3	31.0	26.5
✓	PA + CA + GcBlock	86.4	83.2	84.8	88.3	67.9	32.0	25.8
✓	SA + CA + GcBlock	87.1	81.9	84.4	90.0	68.2	31.0	26.5
✓	PA + SA + CA + GcBlock	89.6	81.0	85.1	90.7	66.0	32.0	26.6

Table 7
Ablation experiments for temperature hyperparameter (%).

τ	P	R	F1 score	mAP@0.5	mAP@0.5:0.95	AP_s	AP_m	AP_l	AR_s	AR_m	AR_l
0.2	86.8	84.6	85.7	89.5	67.1	0	51.7	71.4	0	70.7	79.5
0.3	85.6	83.6	84.6	89.8	67.7	0	44.8	71.8	0	70.1	79.1
0.4	84.8	84.8	84.8	89.2	68.0	0	42.3	71.9	0	67.0	80.4
0.5	89.6	81.0	85.1	90.7	66.0	0	48.8	69.0	0	72.5	70.9
0.6	87.1	81.4	84.2	89.7	67.6	0	55.8	71.6	0	71.2	79.6
0.7	87.4	82.2	84.7	89.1	67.9	0	49.6	72.0	0	69.0	79.6
0.8	87.3	83.6	85.4	90.1	68.5	15.0	46.6	72.7	90.0	68.0	80.4
0.9	83.8	85.4	84.6	89.9	68.0	0	45.9	72.5	0	70.9	80.4

Table 8
Testing results of our proposed method under varying standard deviations (SDs) of Gaussian noise on the strawberry diseases dataset (%).

Method	SDs	P	R	mAP@0.5	mAP@0.5:0.95	AP_s	AP_m	AP_l	AR_s	AR_m	AR_l
Teacher YOLOv7	0	88.2	84.9	91.9	71.0	90.0	59.6	74.2	90.0	72.2	82.6
	2	89.4	84.9	91.6	70.8	80.0	57.5	74.2	80.0	71.7	81.7
	4	89.6	83.3	91.1	70.2	80.0	59.4	73.2	80.0	70.9	80.7
	6	87.9	82.9	89.7	68.7	45.0	53.4	72.2	90.0	67.1	80.3
	8	88.3	76.1	85.7	66.0	10.0	49.9	68.9	80.0	64.9	78.8
Student YOLOv7-tiny	0	84.5	81.5	87.9	65.6	0.0	42.1	69.0	0.0	69.0	79.6
	2	87.6	78.8	89.6	64.7	3.2	42.2	68.4	70.0	67.0	79.4
	4	82.5	81.7	88.2	63.5	7.0	40.4	67.6	70.0	67.1	78.4
	6	85.5	75.7	86.2	61.8	0.0	38.1	65.8	0.0	66.9	77.9
	8	80.4	74.0	81.6	57.7	0.0	33.5	61.2	0.0	63.0	75.4
FGD [30]	0	86.0	82.8	88.7	64.7	0.0	45.3	68.1	0.0	70.5	78.2
	2	85.1	83.9	89.2	66.2	0.0	50.3	69.6	0.0	70.2	79.1
	4	85.9	82.4	88.6	64.8	0.0	47.9	68.8	0.0	69.1	78.8
	6	85.0	76.3	86.7	62.7	0.0	38.6	66.2	0.0	68.8	76.1
	8	79.3	76.2	83.1	59.5	0.0	35.3	63.0	0.0	66.6	74.6
IRKD [6]	0	84.3	85.9	90.3	66.0	5.5	52.7	69.7	60.0	72.2	79.1
	2	88.5	82.2	90.5	65.9	0.0	53.1	69.8	0.0	71.7	79.2
	4	82.7	84.7	89.3	65.1	3.7	50.3	68.8	60.0	67.7	78.5
	6	83.4	81.1	87.4	63.0	0.0	49.5	67.0	0.0	68.1	77.3
	8	80.9	75.6	83.0	59.0	0.0	37.0	63.0	0.0	62.9	75.4
GKD-BMFI [31]	0	85.2	81.0	87.7	65.2	0.0	38.8	69.6	0.0	59.5	79.7
	2	85.7	78.3	86.7	64.6	0.0	39.6	68.7	0.0	59.3	79.0
	4	86.9	76.7	85.7	63.3	0.0	44.7	67.7	0.0	58.8	78.5
	6	86.9	73.5	83.9	61.3	0.0	37.0	65.7	0.0	56.8	77.3
	8	83.0	71.2	80.9	58.2	0.0	33.4	62.4	0.0	61.9	76.0
Ours	0	89.6	81.0	90.7	66.0	0.0	48.8	69.0	0.0	72.5	70.9
	2	83.1	85.3	90.2	65.5	0.0	47.0	68.6	0.0	71.7	78.3
	4	90.1	78.7	89.5	64.6	0.0	44.8	67.7	0.0	72.0	77.8
	6	86.9	79.7	86.7	63.0	0.0	45.5	67.2	0.0	67.0	78.0
	8	82.9	77.9	83.2	60.5	0.0	40.4	64.3	0.0	67.5	77.2

4.5. Robustness analysis

To evaluate the robustness of our method, we conduct experiments with varying levels of Gaussian noise, a common type of real-world disturbance. The standard deviation (SD) is used to quantify the noise level.

As shown in Table 8, ACKD maintains robust detection performance under progressive Gaussian noise corruption. At the maximum tested intensity (SD = 8), the model retains mAP@0.5 at 83.2% and mAP@0.5:0.95 at 60.5%, representing degradations of merely 7.5% and 5.5% from clean input performance, respectively. This stability indicates

effective preservation of feature discriminability despite significant pixel-level perturbations.

In contrast, the student baseline and competing distillation methods exhibit substantially sharper accuracy decay, suffering mAP@0.5:0.95 reductions up to 7.9%, whereas ACKD limits degradation to 5.5%. While heavyweight teachers and FGD demonstrate comparable resilience, ACKD achieves this robustness with significantly fewer parameters, validating its efficiency for resource-constrained deployments. These results confirm the superior robustness of ACKD against input disturbances, ensuring reliable detection performance in challenging agricultural environments with noisy imaging conditions.

5. Conclusion

This paper introduces a novel attention-based and context-aware knowledge distillation (ACKD) approach for accurate crop disease detection. ACKD integrates attention and context distillation, enabling the student model to identify salient diseased regions even in complex agricultural settings accurately. By employing a pixel-wise attention mechanism based on gradients, ACKD assigns varying weights to pixels based on their contribution to the training loss, differing from traditional methods that uniformly distribute weights across channels. The results confirm ACKD as a significant advancement in crop disease detection, providing a more sophisticated and effective approach for identifying diseases in agricultural settings. Specifically, its lightweight architecture enables deployment on unmanned aerial vehicles (UAVs) and edge devices for real-time, large-scale field monitoring.

Future work in this domain could concentrate on the following pivotal objectives. First, integrating ACKD with IoT devices in precision agriculture is crucial for facilitating rapid deployment and establishing a comprehensive monitoring and management system. Second, developing an explainable knowledge distillation would enhance transparency, enabling farmers to understand the model's rationale for disease detection and its basis for treatment recommendations. Third, advancing UAV-deployable ACKD systems with multi-modal aerial surveillance and enhanced robustness to real-world perturbations, such as motion blur and dynamic illumination variations, is essential for scalable automated agricultural assessment.

CRediT authorship contribution statement

Xiangyuan Zhu: Writing – review & editing, Writing – original draft, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Taotao Mao:** Visualization, Validation, Software, Methodology, Data curation. **Jianguo Chen:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Formal analysis. **Feifan Peng:** Visualization, Software, Data curation. **Keqin Li:** Writing – review & editing, Methodology, Formal analysis.

Declaration of competing interest

The authors declare the following financial interests/personal relationships that may be considered potential competing interests:

Jianguo Chen reports that financial support was provided by the National Natural Science Foundation of China, the Guangxi Key Research and Development Program, the Pearl River Talent Plan, and the Natural Science Foundation of Guangdong Province of China. Xiangyuan Zhu reports that financial support was provided by the Key Field Project in Artificial Intelligence (Intelligent Robots) for Regular Higher Education Institutions of Guangdong Province and the Innovative Research Team of the Zhaoqing Big Data Engineering Technology Center. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was partially funded by the Key Field Project in Artificial Intelligence (Intelligent Robots) for Regular Higher Education Institutions of Guangdong Province [Grant No. 2025ZDZX3041], the National Natural Science Foundation of China [Grant No. 62372486], the Natural Science Foundation of Guangdong Province [Grant No. 2023A1515011179], the Guangxi Key Research and Development Program [Grant No. AB24010160], the Pearl River Talent Plan [Grant No. 2023QN10X579], and the Innovative Research Team of the Big Data Engineering Technology Center, Zhaoqing University.

Data availability

The source code is available at https://github.com/ZQU-BD/Demo/tree/main/YOLOv7-ACKD.

References

[1] H. Li, H. Zhang, J. Zhao, L. Huang, C. Ruan, Y. Dong, W. Huang, D. Liang, Automatic localization of image semantic patches for crop disease recognition, Appl. Soft Comput. 165 (2024) 112076, https://doi.org/10.1016/j.asoc.2024.112076

[2] W. Ding, M. Abdel-Basset, I. Alrashdi, H. Hawash, Next generation of computer vision for plant disease monitoring in precision agriculture: a contemporary survey, taxonomy, experiments, and future direction, Inf. Sci. 665 (2024) 120338, https://doi.org/10.1016/j.ins.2024.120338

[3] K. Aghamohammadesmaeilketabforoosh, S. Nikan, G. Antonini, J.M. Pearce, Optimizing strawberry disease and quality detection with vision transformers and attention-based convolutional neural networks, Foods 13 (12) (2024) 1869, https://doi.org/10.3390/foods13121869

[4] J. He, B. Liu, F. Cao, J. Xu, Y. Xiao, Few-shot object counting with dynamic similarity-aware in latent space, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–14, https://doi.org/10.1109/TGRS.2024.3350383

[5] C. Liu, F. Cao, Y. Diao, Y. He, S. Cai, Geographical origin identification of dendrobium officinale using variational inference-enhanced deep learning, Foods 14 (19) (2025) 3361, https://doi.org/10.3390/foods14193361

[6] J. Xue, J. Li, Y. Han, Z. Wang, C. Deng, T. Xu, Feature-based knowledge distillation for infrared small target detection, IEEE Geosci. Remote Sens. Lett. 21 (2024) 1–5, https://doi.org/10.1109/LGRS.2024.3369859

[7] Q. Lan, Q. Tian, Instance, scale, and teacher adaptive knowledge distillation for visual detection in autonomous driving, IEEE Trans. Intell. Veh. 8 (3) (2023) 2358–2370, https://doi.org/10.1109/TIV.2022.3217261

[8] Y. Song, P. Zhang, W. Huang, Y. Zha, T. You, Y. Zhang, Closed-loop unified knowledge distillation for dense object detection, Pattern Recognit. 149 (2024) 110235, https://doi.org/10.1016/j.patcog.2023.110235

[9] X. Zhang, J. Zhu, D. Wang, Y. Wang, T. Liang, H. Wang, Y. Yin, A gradual self distillation network with adaptive channel attention for facial expression recognition, Appl. Soft Comput. 161 (2024) 111762, https://doi.org/10.1016/j.asoc.2024.111762

[10] T. Shinde, An efficient and scalable framework for lightweight crop disease recognition in low-resource settings, in: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2025, pp. 5534–5541, https://doi.org/10.1109/CVPRW67362.2025.00550

[11] Q. Zhou, Y. Liu, L. Wu, A. Li, Y. Wu, A lightweight network with efficient channel attention for intelligent daylily maturity detection method, IEEE Trans. AgriFood Electron. 3 (2) (2025) 463–473, https://doi.org/10.1109/TAFE.2025.3566254

[12] T. Su, Q. Liang, J. Zhang, Z. Yu, G. Wang, X. Liu, Attention-based feature interaction for efficient online knowledge distillation, in: 2021 IEEE International Conference on Data Mining (ICDM), 2021, pp. 579–588, https://doi.org/10.1109/ICDM51629.2021.00069

[13] M. Ji, B. Heo, S. Park, Show, attend and distill: knowledge distillation via attention-based feature matching, in: The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21), vol. 35, 2021, pp. 7945–7952, https://doi.org/10.1609/aaai.v35i9.16969

[14] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618–626, https://doi.org/10.1109/ICCV.2017.74

[15] Y. Cao, J. Xu, S. Lin, F. Wei, H. Hu, GCNet: Non-local networks meet squeeze-excitation networks and beyond, in: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019, pp. 1971–1980, https://doi.org/10.1109/ICCVW.2019.00246

[16] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788, https://doi.org/10.1109/CVPR.2016.91

[17] J. Redmon, A. Farhadi, Yolo9000: better, faster, stronger, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6517–6525, https://doi.org/10.1109/CVPR.2017.690

[18] J. Redmon, A. Farhadi, Yolov3: an incremental improvement, arXiv:1804.02767, 2018.
[19] C.-Y. Wang, A. Bochkovskiy, H.-Y.M. Liao, Yolov7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, in: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 7464–7475, https://doi.org/10.1109/CVPR52729.2023.00721
[20] R. Varghese, M. Sambath, Yolov8: a novel object detection algorithm with enhanced performance and robustness, in: 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), 2024, pp. 1–6, https://doi.org/10.1109/ADICS58448.2024.10533619
[21] Y. Zhang, B. Ma, Y. Hu, C. Li, Y. Li, Accurate cotton diseases and pests detection in complex background based on an improved yolox model, Comput. Electron. Agric. 203 (2022) 107484, https://doi.org/10.1016/j.compag.2022.107484
[22] J. Zhou, J. Li, C. Wang, H. Wu, C. Zhao, Q. Wang, A vegetable disease recognition model for complex background based on region proposal and progressive learning, Comput. Electron. Agric. 184 (2021) 106101, https://doi.org/10.1016/j.compag.2021.106101
[23] E. Li, L. Wang, Q. Xie, R. Gao, Z. Su, Y. Li, A novel deep learning method for maize disease identification based on small sample-size and complex background datasets, Ecol. Inform. 75 (2023) 102011, https://doi.org/10.1016/j.ecoinf.2023.102011
[24] Y.-B. Lin, C.-Y. Liu, W.-L. Chen, C.-H. Chang, F.-L. Ng, K. Yang, J. Hsung, IoT-based strawberry disease detection with wall-mounted monitoring cameras, IEEE Internet Things J. 11 (1) (2024) 1439–1451, https://doi.org/10.1109/JIOT.2023.3288603
[25] R. Tang, Z. Liu, Y. Li, Y. Song, H. Liu, Q. Wang, J. Shao, G. Duan, J. Tan, Task-balanced distillation for object detection, Pattern Recognit. 137 (2023) 109320, https://doi.org/10.1016/j.patcog.2023.109320
[26] Z. Tu, X. Liu, X. Xiao, A general dynamic knowledge distillation method for visual analytics, IEEE Trans. Image Process. 31 (2022) 6517–6531, https://doi.org/10.1109/TIP.2022.3212905
[27] S. Yang, J. Yang, M. Zhou, Z. Huang, W.-S. Zheng, X. Yang, J. Ren, Learning from human educational wisdom: a student-centered knowledge distillation method, IEEE Trans. Pattern Anal. Mach. Intell. 46 (6) (2024) 4188–4205, https://doi.org/10.1109/TPAMI.2024.3354928
[28] Z. Zheng, R. Ye, Q. Hou, D. Ren, P. Wang, W. Zuo, M.-M. Cheng, Localization distillation for object detection, IEEE Trans. Pattern Anal. Mach. Intell. 45 (8) (2023) 10070–10083, https://doi.org/10.1109/TPAMI.2023.3248583
[29] D. Ma, K. Zhang, Q. Cao, J. Li, X. Gao, Coordinate attention guided dual-teacher adaptive knowledge distillation for image classification, Expert Syst. Appl. 250 (2024) 123892, https://doi.org/10.1016/j.eswa.2024.123892
[30] Z. Yang, Z. Li, X. Jiang, Y. Gong, Z. Yuan, D. Zhao, C. Yuan, Focal and global knowledge distillation for detectors, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 4633–4642, https://doi.org/10.1109/CVPR52688.2022.00460
[31] Q. Lan, Q. Tian, Gradient-guided knowledge distillation for object detectors, in: 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2024, pp. 423–432, https://doi.org/10.1109/WACV57701.2024.00049
[32] T. Zhao, X. Wu, Pyramid feature attention network for saliency detection, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 3080–3089, https://doi.org/10.1109/CVPR.2019.00320
[33] U. Afzaal, B. Bhattarai, Y.R. Pandeya, J. Lee, An instance segmentation model for strawberry diseases based on mask R-CNN, Sensors 21 (19) (2021), https://doi.org/10.3390/s21196565
[34] M. Gehlot, R.K. Saxena, G.C. Gandhi, Tomato-Village: a dataset for end-to-end tomato disease detection in a real-world environment, Multimedia Syst. 29 (6) (2023) 3305–3328, https://doi.org/10.1007/s00530-023-01158-y
[35] Roboflow, Tomato leaf diseases detection, 2024. https://www.kaggle.com/datasets/farukalam/tomato-leaf-diseases-detection-computer-vision?resource=download.
[36] The Marconi Research and Innovation Laboratory, Bean plant pathologies dataset for deep learning tasks, 2024. https://www.kaggle.com/datasets/msjahid/bean-crop-disease-diagnosis-and-spatial-analysis.