



Leaf rolling detection in maize under complex environments using an improved deep learning method

Yuanhao Wang^{1,2} · Xuebin Jing^{1,2} · Yonggang Gao³ · Xiaohong Han¹ · Cheng Zhao³ · Weihua Pan²

Received: 16 May 2024 / Accepted: 5 August 2024
© The Author(s) 2024

Abstract

Leaf rolling is a common adaptive response that plants have evolved to counteract the detrimental effects of various environmental stresses. Gaining insight into the mechanisms underlying leaf rolling alterations presents researchers with a unique opportunity to enhance stress tolerance in crops exhibiting leaf rolling, such as maize. In order to achieve a more profound understanding of leaf rolling, it is imperative to ascertain the occurrence and extent of this phenotype. While traditional manual leaf rolling detection is slow and laborious, research into high-throughput methods for detecting leaf rolling within our investigation scope remains limited. In this study, we present an approach for detecting leaf rolling in maize using the YOLOv8 model. Our method, LRD-YOLO, integrates two significant improvements: a Convolutional Block Attention Module to augment feature extraction capabilities, and a Deformable ConvNets v2 to enhance adaptability to changes in target shape and scale. Through experiments on a dataset encompassing severe occlusion, variations in leaf scale and shape, and complex background scenarios, our approach achieves an impressive mean average precision of 81.6%, surpassing current state-of-the-art methods. Furthermore, the LRD-YOLO model demands only 8.0 G floating point operations and the parameters of 3.48 M. We have proposed an innovative method for leaf rolling detection in maize, and experimental outcomes showcase the efficacy of LRD-YOLO in precisely detecting leaf rolling in complex scenarios while maintaining real-time inference speed.

Key message

In this study, we propose an improved object detection algorithm for detecting leaf rolling, a common adaptive response to environmental stresses. It achieves 81.6% mean average precision, surpassing existing methods.

Keywords Leaf rolling · Object detection · Maize · Deep learning

✉ Xiaohong Han
hanxiaohong@tyut.edu.cn

✉ Cheng Zhao
zhaocheng01@caas.cn

✉ Weihua Pan
panweihua@caas.cn

¹ College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan 030024, China

² Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China

³ Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China

Introduction

Maize stands as a fundamental staple crop, playing a pivotal role in ensuring food security. Additionally, it serves as a vital source of feed, energy, and forage (Tanumihardjo et al. 2020). However, drought emerges as a primary contributor to significant declines in maize crop yield (Farhangfar et al. 2015). To mitigate the adverse impacts of environmental stresses, plants have developed diverse mechanisms, among which leaf rolling is noteworthy. The rolling of leaves is a prevalent adaptive response seen in plants experiencing drought stress (Kadioglu et al. 2012). This physiological adaptation diminishes light interception, transpiration, and leaf dehydration. As a result, it emerges as a potentially valuable mechanism for drought avoidance, especially in arid regions (Kadioglu et al. 2007). Besides drought, leaf rolling can be triggered by various abiotic stresses like water

deficit and high temperature, there are also biotic stresses to consider, including insect infestation and fungal infections. Understanding the mechanisms behind leaf rolling alterations provides researchers with a distinct opportunity to enhance stress tolerance in crops exhibiting this trait, like maize (Kadioglu et al. 2012).

To gain a more profound understanding of leaf rolling as a mechanism, it is imperative to ascertain the occurrence and extent of this phenotype. Traditional leaf rolling detection has primarily been a manual process, known for its labor-intensive and time-consuming nature. Clarke visually assessed the degree of leaf rolling (Clarke 1986). Premachandra et al. assessed the extent of leaf rolling by quantifying the decrease in leaf width as a percentage caused by rolling (Premachandra et al. 1993). An analogous scoring method, which evaluates the percentage decrease in the width of the central part of the leaf due to rolling, was employed to establish the correlation between drought resistance and rolling (Saruhan et al. 2011). Zhang et al. computed the index of rolling by evaluating the widths of leaves in both their natural and unfolded states (Zhang et al. 2009). Sirault et al. developed a repeatable protocol to quantify leaf curvature. Micro-photographs of leaf cross-sections were taken, and two approaches were employed for quantifying leaf rolling: one based on the convex hull of the cross-section and the other using cubic smoothing splines for mathematical approximation. Both approaches yielded objective measurements (Sirault et al. 2015). Baret et al. investigated the viability of an efficient method for assessing leaf rolling in maize through aerial observation using UAVs, but no further applications were pursued (Baret et al. 2018). Visual scoring methods for leaf rolling are often subjective, while various assessment experiments can be both costly and inefficient. These low-throughput techniques present challenges when applied to large-scale phenotyping experiments. However, the research into high-throughput methods for determining leaf rolling within our investigation scope remains limited. Therefore, there exists an urgent demand for high-throughput methodologies, especially within the realm of field experiments.

Recently, the ongoing advancement of high-throughput plant phenotyping measurement and analysis technology has been accompanied by progress in artificial intelligence, notably in deep learning, contributing to plant phenotyping research (Jiang et al. 2020). Leaves, being integral components of plants, demand accurate detection and analysis, crucial for various applications such as species recognition (Mehdipour Ghazi et al. 2017; Waldchen et al. 2018a, b), disease diagnosis (Darwish et al. 2020; Martinelli et al. 2014), and vegetation analysis (Ding et al. 2020). Cutting-edge object detection algorithms in deep learning have found extensive applications in leaf detection, counting, and disease detection (Liu et al. 2020; Oo et al. 2018; Pal et al.

2023; Thai et al. 2023; Ubbens et al. 2018). These advancements lay the groundwork for our proposal of a method for detecting leaf rolling. The intricacies of dense leaves, characterized by occlusion, have consistently posed challenges in leaf-related tasks, thereby presenting difficulties in leaf rolling detection. Scale variations among leaves in different growth stages, alterations in leaf shape due to rolling, and background interference in complex environments are additional factors influencing our detection results. Our aim is to address these challenges and present a precise, high-throughput method for detecting leaf rolling in maize using an object detection algorithm.

This study introduces a method by integrating DCNv2 (Deformable ConvNets v2) (Zhu et al. 2019) alongside the CBAM (Convolutional Block Attention Module) (Woo et al. 2018) into YOLOv8. Our suggested method introduces DCNv2 to address deformation and scale disparities in leaf rolling detection in maize, and CBAM, a lightweight and effective attention mechanism, to strengthen feature extraction capability and feature validity. We term this method LRD-YOLO. The proposed LRD-YOLO model undergoes validation and testing on our dataset. Experimental findings showcase that our proposed method surpasses others in terms of accuracy, showcasing its effectiveness for detecting leaf rolling in maize. The contributions highlighted in this study are as follows:

- We created a dataset comprising maize leaves in different growth stages and with varying degrees of rolling in complex natural environments for leaf rolling detection in maize, meticulously labeling all data.
- We proposed a novel approach for leaf rolling detection in maize based on improved YOLOv8 with Deformable ConvNets v2 and Convolutional Block Attention Module.
- Through a comprehensive set of experiments on our dataset, we showcase that LRD-YOLO demonstrates exceptional performance in both accuracy and efficiency, surpassing other methods.

Materials and methods

Image acquisition

The images of maize were obtained from a greenhouse situated at the Shenzhen Experimental Base of the Chinese Academy of Agricultural Sciences, using the rear cameras of iPhone 13 and iPhone 14. Scientific water replenishment measures were implemented throughout the maize's growth cycle to manage water stress levels, resulting in varying degrees of leaf rolling, ranging from mild to severe.

As illustrated in Fig. 1 Samples of the data Fig. 1, these images were obtained under diverse conditions, including



Fig. 1 Samples of the data

overlap, occlusion, and multi-scale occurrences between leaves. The backgrounds featured a mix of weeds and wilted maize leaves, and light effects were also considered. The data collection took place in July 2023, yielding a total of 724 original maize images with multiple perspectives including 7878 individual target leaves, which were used to construct the dataset for this study.

Image annotation

To accurately assess the occurrence of maize leaf rolling, we employed the leaf rolling assessment criteria established by CIMMYT (Bänziger et al. 2000). The assessment involved measuring rolling on individual leaves, and the criteria are depicted in Fig. 2. In Stage 1, the leaf is unrolled and turgid, while from Stage 2 onwards, the leaf rim starts to roll. By Stage 3, the leaf blade displays pronounced rolling, appearing V-shaped; by Stage 4, the rolled leaf rim extends over a

section of the leaf blade. By Stage 5, the leaf is rolled tightly, resembling an onion.

In this study, the dataset is categorized into two classes based on the various stages of maize leaf rolling during labeling: leaf and rolled. During the classification process, leaves at Stage 1 are labeled as leaf, while leaves at Stage 2 to Stage 5 are labeled as rolled.

The images used in this study underwent annotation by the Labeling (Tzutalin 2015) software with the labeling file format being.txt. After the labeling process was finished, the labeled images were divided into training, validation, and test sets in an 8:1:1 ratio.

YOLOv8 model

YOLOv8 (Jocher et al. 2023), created by Ultralytics, stands as a cutting-edge YOLO model, demonstrating versatile applications in object detection and image classification

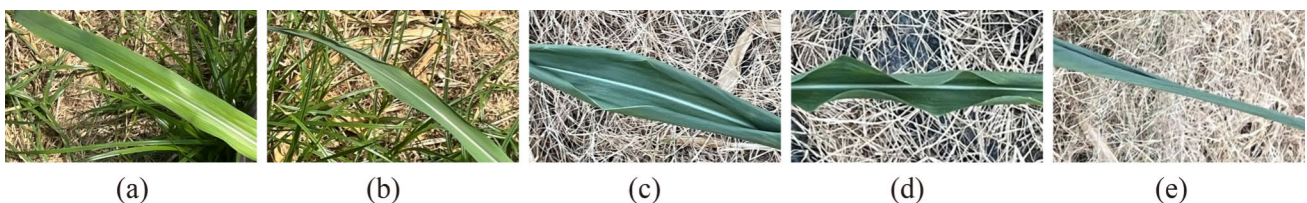


Fig. 2 Leaf rolling stage from 1 to 5. Stage 1, the leaf is unrolled and turgid; Stage 2, the leaf rim starts to roll; Stage 3, the leaf blade displays pronounced rolling, appearing V-shaped; Stage 4, rolled leaf

rim extends over a section of the leaf blade; Stage 5, the leaf is rolled tightly, resembling an onion

tasks. Ultralytics, known for their impactful YOLOv5 model (Jocher 2020), has once again set industry benchmarks with YOLOv8.

While YOLOv8 maintains the overarching network architecture of YOLOv5, encompassing the structural design of both backbone and neck while also considering various scale models, it introduces numerous modifications and improvements. YOLOv8 integrates the C2f module into its backbone, resulting in a reduction in the overall network size. The C2f module serves as the fundamental building block in the Backbone, featuring a smaller parameter count and superior feature extraction capabilities compared to the C3 module of YOLOv5. Refer to Fig. 3 for a graphical depiction illustrating the structures of the C3 and C2f modules. And introduce the Decoupled-Head concept (Ge et al. 2021). It retains the Path Aggregation Network (Liu et al. 2018) concept but removes the convolutional structure in the UpSampling stage. Furthermore, it discards the Anchor-Base, adopting the Anchor-Free approach. These improvements lead to increased performance in object detection, positioning YOLOv8 as the selected baseline model for our study.

Improvement of the YOLOv8 model

To improve the performance of detecting leaf rolling, we propose the LRD-YOLO, as depicted in Fig. 4 LRD-YOLO

addresses challenges associated with scale variation and occlusion in leaves at different growth stages.

To capture the scale variation induced by leaves at various growth stages, we incorporate the Deformable ConvNets v2 (DCNv2) into the model. Specifically, we substitute the convolution in the C2f module with the DCNv2. This adjustment aims to enhance the capability of the model in detecting leaves with deformations or significant scale variations. Additionally, to enhance leaf rolling detection in scenarios where leaves may occlude or overlap, we incorporate the CBAM before the small and medium detection heads. This strategic placement of the CBAM module aids in better detecting leaves that are subject to occlusion or overlap.

The proposed enhancements to the LRD-YOLO model significantly contribute to the overall accuracy and robustness of leaf rolling detection. Furthermore, these improvements enable the model to effectively adapt to the challenges posed by multiscale and occluded leaf detection within complex natural environments.

Deformable convnets v2

In traditional convolutional neural networks, convolution operations are performed at fixed positions within the input feature maps, as depicted in Fig. 5a. However, real-world scenarios often entail objects within images undergoing various transformations, such as deformations, rotations,

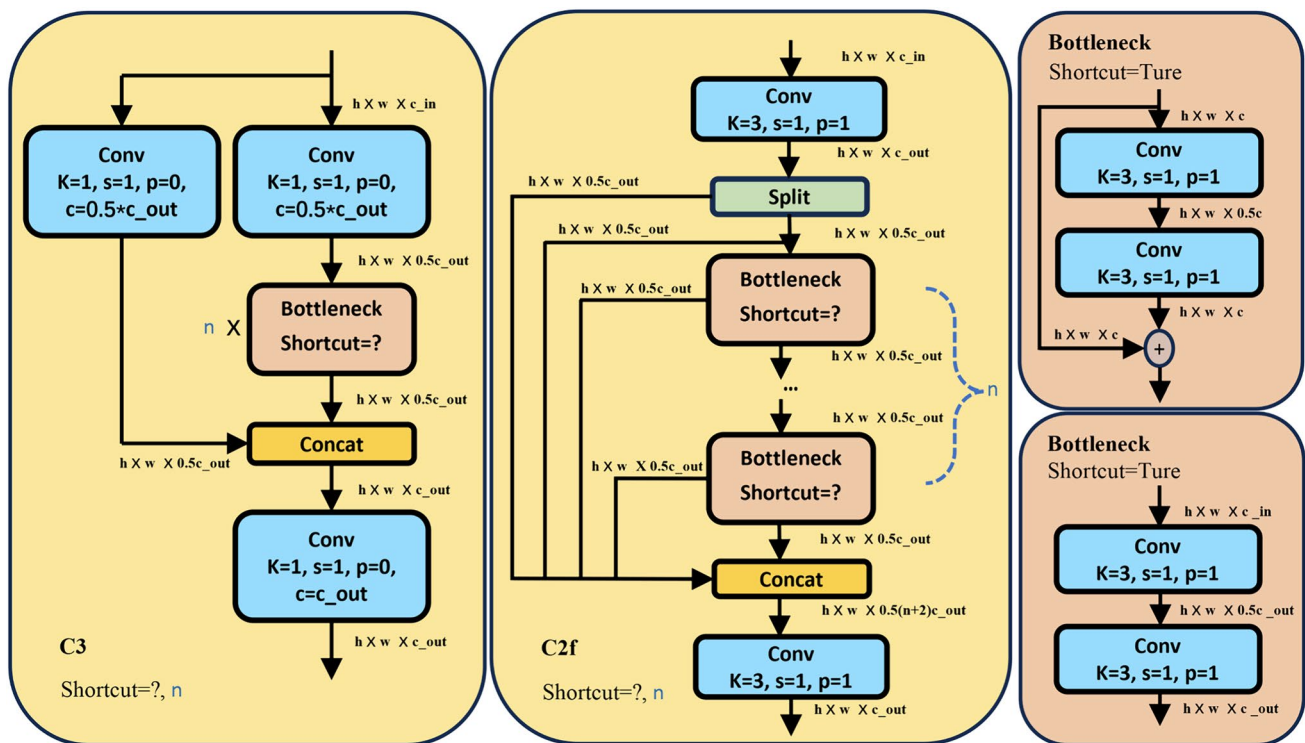


Fig. 3 The structures of the C3 and C2f modules

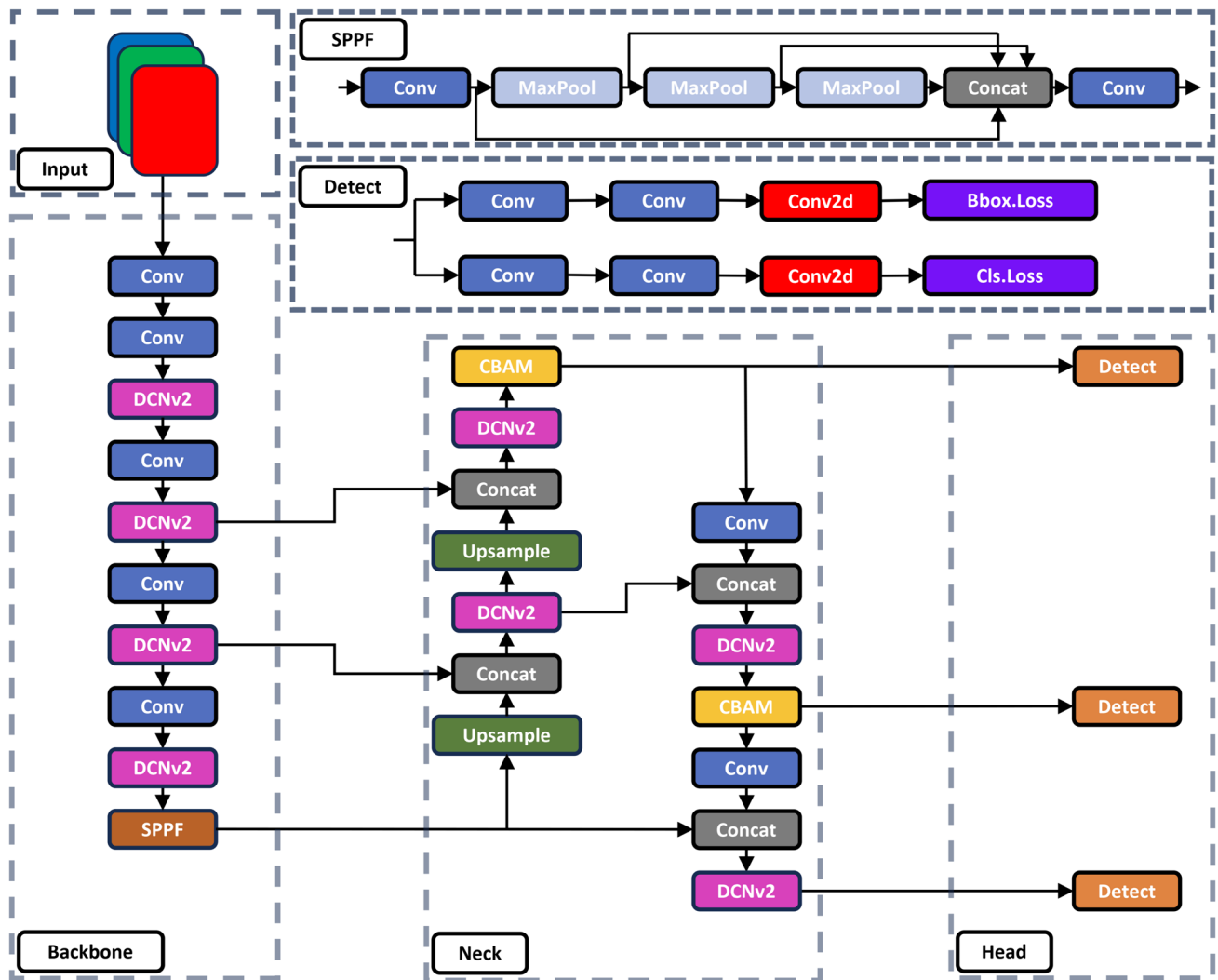


Fig. 4 Overall architecture of the proposed LRD-YOLO

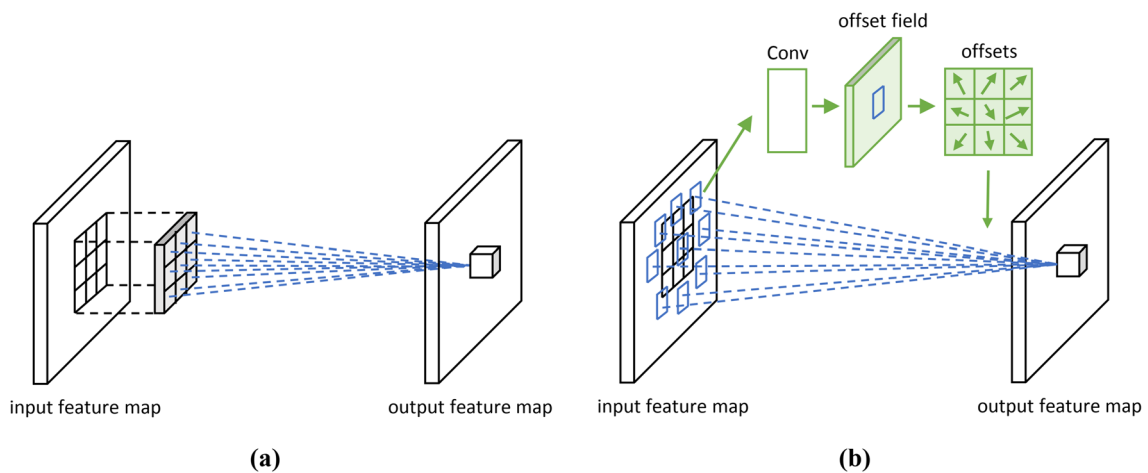


Fig. 5 Comparison of traditional convolution and deformable convolution. a Traditional convolution kernel. b Deformable convolution kernel

or changes in scale. These transformations pose challenges for traditional CNNs, impeding their ability to effectively capture relevant features. The DCN (Deformable Convolutional Networks) (Dai et al. 2017) is intricately designed to overcome the inherent constraints of conventional methodologies.

DCN addresses this limitation by introducing offsets ΔP_n to adapt convolutional kernels. By incorporating offsets into deformable convolutions, the convolutional kernels gain increased flexibility, enabling them to dynamically adjust their sampling positions. This flexibility enables the network to prioritize areas of interest within the input, effectively handling geometric variations and deformations. The representation of the deformable convolution operation is depicted below:

$$y(P_0) = \sum_{P_n \in R} w(P_n) \times (P_0 + P_n + \Delta P_n) \tag{1}$$

For a single feature map input, depicted in Fig. 5b, an extra 3×3 convolutional layer learns the offset. The output dimension matches the original feature map size. Deformable convolution starts with an interpolation operation using the generated offset, followed by standard convolution.

However, it is plausible that deformable convolution introduces extraneous regions that interfere with feature extraction, resulting in a degradation of algorithm performance. To address this issue, Deformable ConvNets v2 not only includes the offset for each sampling point but also incorporates a weight coefficient Δm_k to distinguish whether the introduced region aligns with our area of interest. The DCNv2 operation is formulated as:

$$y(P_0) = \sum_{P_n \in R} w(P_n) \times (P_0 + P_n + \Delta P_n) \Delta m_k \tag{2}$$

The weight coefficient is designed to distinguish between regions that align with the area of interest and those that do not. By incorporating these weight coefficients, DCNv2 can effectively filter out extraneous regions that may interfere

with feature extraction, thereby leading to an enhancement in the overall algorithm performance.

In summary, the offsets in DCN aim to pinpoint the location of regions containing valid information, while the incorporation of weight coefficients in DCNv2 serves to assign significance to these identified locations. Both mechanisms collectively ensure the precise extraction of valid information. Maize leaves undergo substantial geometric deformation during the rolling process, and there is also a challenge associated with considerable scale differences between leaves at various growth stages. Consequently, the application of Deformable ConvNets v2 proves instrumental in addressing both the deformation and scale disparities inherent in the detection of rolled maize leaves.

Convolutional block attention module

As an attention mechanism, CBAM is intended to amplify the representation capability of convolutional neural networks by concurrently emphasizing both channel-wise and spatial-wise features. In Fig. 6, the CBAM attention module's comprehensive structure is depicted, with the channel attention module focusing on essential features and the spatial attention module attending to their respective positions.

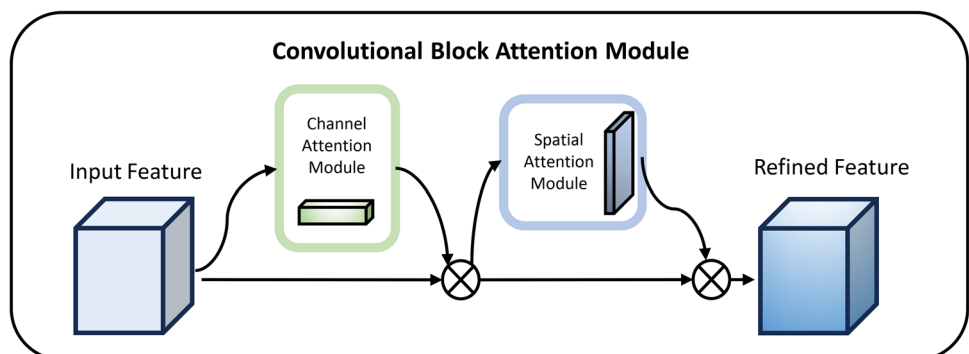
As depicted in Fig. 7a, the initial steps involve performing the pooling operation on the input feature map F to produce new feature maps. These are then concurrently input into a weight-sharing Multilayer Perceptron network, undergoing operations for dimensionality reduction and enhancement to manage parameter count. The resulting feature maps are activated using sigmoid activation, resulting in output feature maps M_c . These maps are subsequently multiplied by F to derive output $M_c(F)$.

The computation for the channel attention module is outlined as follows:

$$M_c = \sigma(\text{MLP}(\text{AvgPool}(F)) \oplus \text{MLP}(\text{MaxPool}(F))) \tag{3}$$

The spatial attention module uses the $M_c(F)$ as input. Initially, it conducts the pooling operation, resulting in the

Fig. 6 Overall architecture of CBAM. The module has two sub-modules: channel attention module and spatial attention module



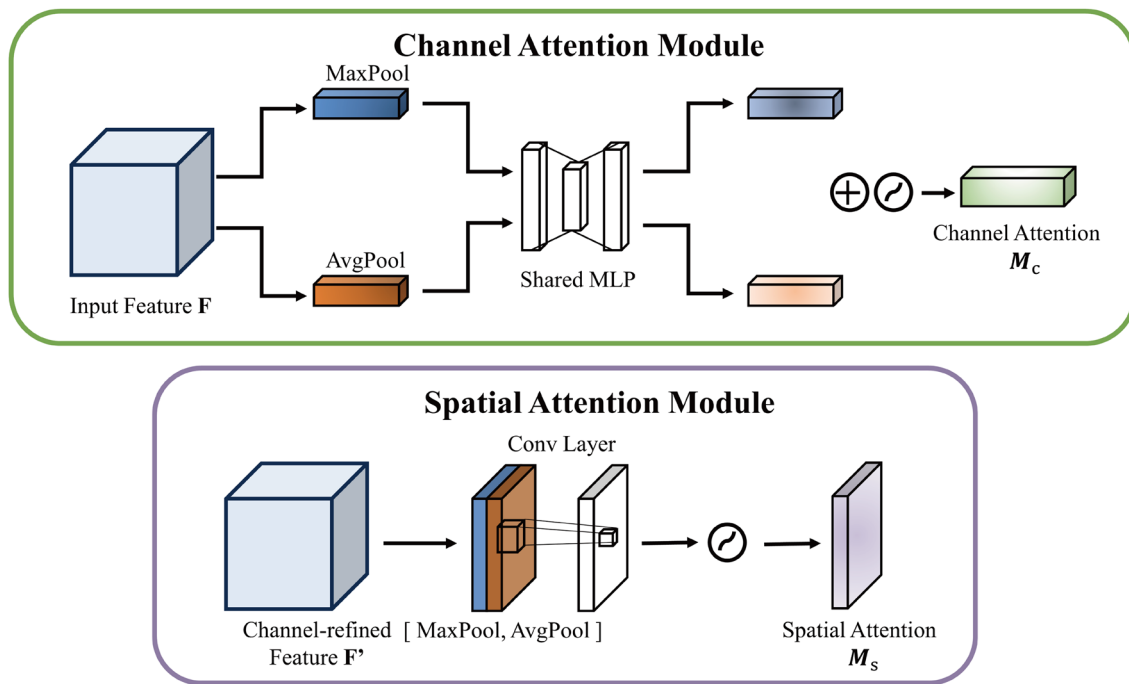


Fig. 7 Architecture of each attention sub-module. **a** Channel attention module. **b** Spatial attention module

generation of two distinct feature maps, which are subsequently concatenated across channels. Following this, a 7×7 convolutional kernel is employed to create a new feature map, with sigmoid activation applied to generate the feature map M_s . Finally, M_s is multiplied by $M_c(F)$ to yield the resulting output $M_s(F)$. The computation for the spatial attention module is expressed is outlined below:

$$M_s = \sigma(f^{(7 \times 7)}([\text{AvgPool}(F); \text{MaxPool}(F)])) \tag{4}$$

In summary, CBAM dynamically adjusts feature map weights, enhancing the model's ability to capture vital image features. As a strategic enhancement, we incorporated the CBAM module to extract features effectively and ensure their validity for leaf rolling detection in maize.

Experimental results

Environment of experiment

The experimental setting for this study operates on a Linux server equipped with 100GB of RAM and a Tesla V100S-PCIE graphics card, featuring Intel® Xeon® Gold 6230R CPUs@2.10GHz. PyTorch serves as the framework for experiments, with the software environment comprising CUDA11.1, Python 3.8.16, and Torch 1.10.1. During the training phase, we run the network for 150 epochs. We

define the size of input image as 640×640 and designate a batch size of 16. Utilizing the AdamW optimizer, we set the learning rate at 0.001667, momentum at 0.9, and weight decay at 0.0005.

Evaluation metrics

To thoroughly evaluate the proposed model for detecting leaf rolling in maize, we employed several evaluation metrics including FLOPs (floating point operations), precision, FPS (frames per second), recall, mAP (mean Average Precision), and the number of parameters. The following equations are utilized to compute the precision and recall:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{5}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{6}$$

The following equation is employed to compute mAP:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \tag{7}$$

In this equation, N is the categories, and AP_i is the average precision for the i th. A higher mAP score indicates more accurate detection.

FPS measures the inference speed, which is critical for assessing real-time model performance. FLOPs provide an estimate of the number of floating-point arithmetic operations necessary for a model during inference, while parameters encompass the trainable biases and weights in the neural network.

Ablation experiments

To assess the influence of each suggested enhancement of LRD-YOLO for leaf rolling detection in maize, we conducted ablation experiments. The hardware environment and parameter settings remained consistent throughout the ablation experiments.

Ablation experiments of the baseline and the LRD-YOLO

We first evaluate the effectiveness of our LRD-YOLO model against the baseline YOLOv8n model. The latter was trained using the same dataset as the former but lacked the incorporation of DCNv2 and the CBAM.

Table 1 Ablation experiment of the YOLOv8n model and the LRD-YOLO model displays the ablation experiment results. The comparison showcases that our two enhanced methods outperform the YOLOv8n model significantly. By incorporating the CBAM attention, the mAP increases by 2.4% to 78.9%, with only a slight increase of 0.03 M

parameters. Upon introducing the DCNv2 module into YOLOv8n, the mAP (80.5%) sees an improvement of 4.0%, and the FLOPs decrease from 8.9 to 8.0. By combining these two improved methods, LRD-YOLO significantly improves mAP(81.6%) by 5.1% and decreases the FLOPs from 8.9 to 8.0 with only a marginal increase of 0.32 M in the number of parameters.

As depicted in Fig. 8, we performed a detailed analysis of the changes in loss values. It's apparent that LRD-YOLO showcases a quicker reduction in loss compared to YOLOv8n on the validation set. This indicates the effectiveness of our enhancements.

The results indicate initial support for the effectiveness of improvements to the baseline YOLOv8n in detecting maize leaf rolling under complex environmental conditions.

Ablation experiments of the Deformable ConvNets v2

Next, we execute a more specific ablation analysis to assess the influence of DCNv2 on the performance of the LRD-YOLO. While the C2f component within YOLOv8 facilitates the acquisition of multi-scale features and broadens the scope of receptive fields, it concurrently raises computational demands and parameter counts. Furthermore, it demonstrates a lack of sensitivity to variations in the shape of the leaves. By replacing convolutional layers within the C2f component with DCNv2, we effectively alleviate

Table 1 Ablation experiment of the YOLOv8n model and the LRD-YOLO model

Methods	DCNv2	CBAM	Params (M)	FLOPs (G)	mAP@0.5 (%)
YOLOv8n			3.16	8.9	76.5
YOLOv8n+DCNv2	✓		3.46	8.0	80.5
YOLOv8n+CBAM		✓	3.19	8.9	78.9
LRD-YOLO	✓	✓	3.48	8.0	81.6

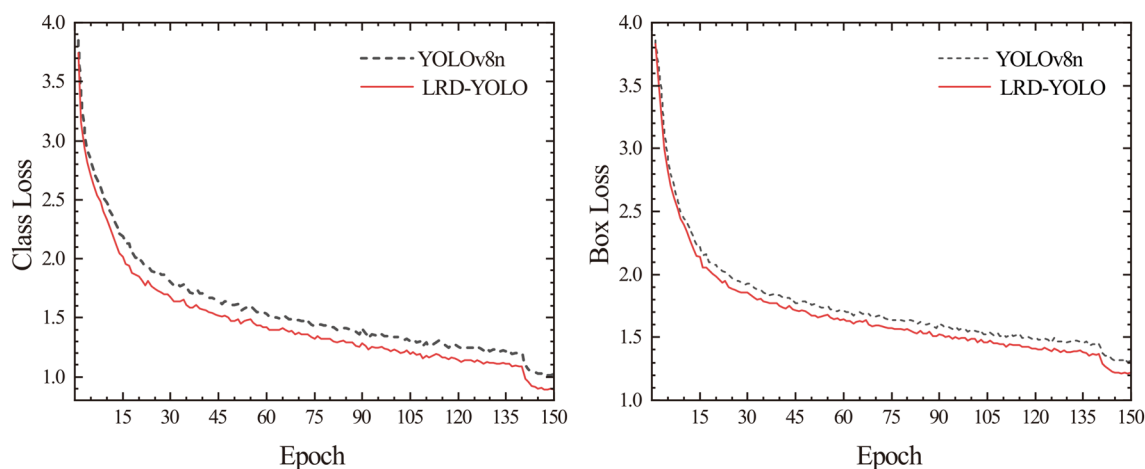


Fig. 8 Analysis of the training loss

Table 2 Comparison of adding DCNv2 to different positions

Backbone	Neck	Params (M)	FLOPs (G)	mAP@0.5 (%)
√		3.32	8.5	80.1
	√	3.30	8.4	79.4
√	√	3.46	8.0	80.5

Table 3 Comparison of adding CBAM module to different detection heads

Small	Medium	Large	Params (M)	FLOPs (G)	mAP0.5 (%)
√			3.16	8.9	77.1
	√		3.17	8.9	77.1
√	√	√	3.24	8.9	75.7
√	√		3.18	8.9	78.9

computational loads and bolster the performance of the baseline model. This enhancement proves especially significant for leaves manifesting notable scale fluctuations across growth phases and for those experiencing alterations in shape due to rolling.

The data in Table 2 highlights the performance contrast across various placements of DCNv2. Clearly, replacing convolutional layers within the C2f component of the baseline model, neither its neck nor its backbone, with DCNv2 yields enhancements in both mAP and FLOPs reduction. These outcomes emphasize the efficacy of incorporating DCNv2 into the C2f component, consequently amplifying the capability of LRD-YOLO to efficiently tackle the challenges posed by deformation and scale variations in identifying rolled maize leaves.

Ablation experiments of the convolutional block attention module

Finally, we examine the impact of CBAM on the efficacy of the LRD-YOLO. We incorporate the CBAM module before the various sizes of the detection head to evaluate its effect on our models.

Table 3 displays a comparison of performance across different placements of the CBAM module. Notably, integrating the CBAM module before the small and medium detection heads showcases the most significant enhancement in mAP. This improvement can be attributed to the dataset’s inclusion of small and medium-sized leaves, which are prone to occlusion and overlap. These outcomes validate the effectiveness of applying CBAM attention before the small and medium detection heads in mitigating missed detections of occluded and small targets.

In summary, the outcomes from all ablation experiments affirm that the integration of both DCNv2 and the CBAM module into the LRD-YOLO significantly enhances the accuracy of leaf rolling detection in maize, especially under challenging environmental conditions.

Comparison with state-of-the-art detection methods

Comparison of performance

We conducted a comprehensive performance evaluation on the test set, comparing LRD-YOLO model with six advanced methods: Faster R-CNN (Ren et al. 2017), SSD (Liu et al. 2016), YOLOv5n (G 2020), YOLOv6n (Li et al. 2022), YOLOv7-Tiny (Wang et al. 2022), and Real-Time Detection Transformer (RT-DETR) (Zhao et al. 2023). All experiments were executed on an NVIDIA TESLA V100s GPU, maintaining a consistent software environment. The performance analysis of these methods is presented in Table 4.

SSD and Faster R-CNN face challenges in achieving a harmonious balance between detection accuracy and inference speed. Burdened by an excess of parameters and arithmetic operations, Faster R-CNN exhibits a low inference speed of only 17.1 FPS. Conversely, while the SSD model showcases a reasonable speed of 48.6 FPS, its diminished precision makes it unsuitable for real-time tasks.

The YOLO methods, particularly adept at leaf rolling detection in maize, reveal distinctive performance characteristics. YOLOv5n stands out with the lowest FLOPs and Params, recorded at 4.2 G and 1.8 M, respectively, while YOLOv7-Tiny boasts the highest FPS at 76.3. Nevertheless,

Table 4 Performance comparison of LRD-YOLO with other detection methods

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	FPS	FLOPs (G)	Params (M)
Faster R-CNN	62.0	76.8	75.1	17.1	370.2	137.1
SSD	78.7	44.1	60.3	48.6	62.7	26.3
YOLOv5n	72.3	65.9	73.0	73.5	4.2	1.8
YOLOv6n	63.6	70.5	71.1	75.2	11.3	4.6
YOLOv7-Tiny	70.8	62.6	70.9	76.3	13.2	6.0
RT-DETR	83.3	72.4	79.5	31.1	108.3	32.8
LRD-YOLO	82.3	73.6	81.6	56.0	8.0	3.5

the detection precision, recall, and mAP metrics of YOLOv5n, YOLOv6n, and YOLOv7-Tiny do not align proportionately with their impressive inference speeds.

The Real-Time Detection Transformer (RT-DETR), an advanced end-to-end object detector devised by Baidu, stands out for its exceptional accuracy while maintaining real-time performance capabilities. RT-DETR exhibits outstanding performance on our dataset, achieving an impressive mAP of 79.5% and precision of 83.3%, surpassing other models within the YOLO series, all while sustaining a speed of 31.1 FPS.

Our proposed model, LRD-YOLO, emerges as the frontrunner with the highest mAP of 81.6%. Notably, its detection accuracy surpasses that of RT-DETR, achieving an improved fps of 56.0, requiring only 8.0 G FLOPs and

3.5 M parameters. These results underscore that our LRD-YOLO model is the optimal choice for leaf rolling detection in maize, successfully balancing both speed and accuracy in the domain.

Comparison of detection results

To further assess the efficacy of these methods, we carried out experiments to compare the actual effectiveness of seven object detection methods for leaf rolling detection. The results are illustrated in Fig. 9.

As depicted in the figures, leaves marked by yellow box or arrow exhibit varying degrees of occlusion and overlap, leading to false or missed detections for all models except the LRD-YOLO model. Faster R-CNN exhibits missed

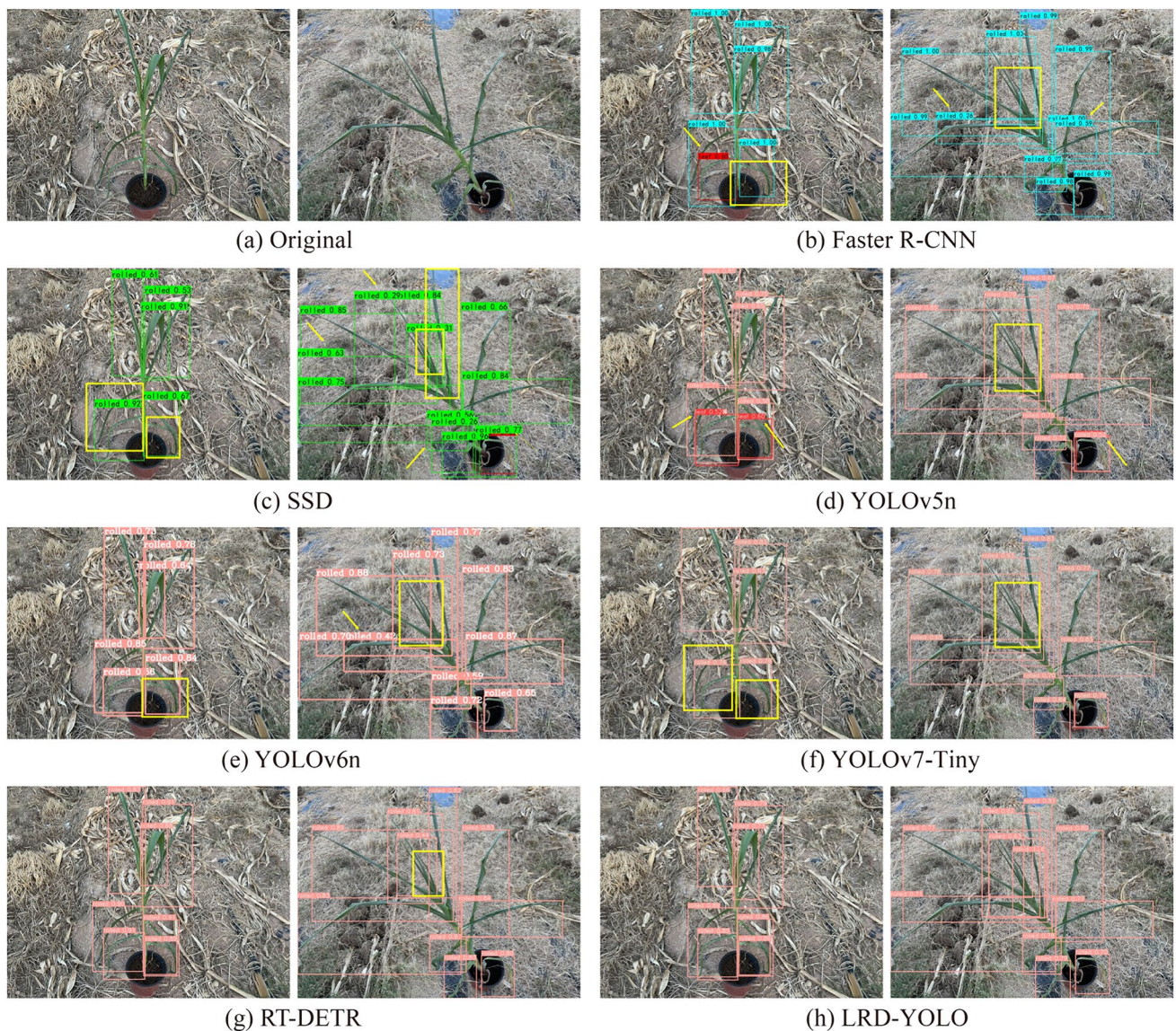


Fig. 9 Comparison of the detection results. The arrow points to the incorrect results, and the yellow box represents the missing target

detections when leaves overlap and occlude each other. Conversely, SSD is more prone to generating redundant detection boxes in dense scenarios. YOLOv5n incorrectly classified the rolled leaves in Fig. 9d as normal leaves, while both YOLOv6n and YOLOv7-Tiny displayed identical missed detections where leaves were either obscured or overlapped. RT-DETR showcased high accuracy in both images, with only one missed detection.

Only the LRD-YOLO model accurately predicted the position and quantity of the rolled leaves. These findings suggest that LRD-YOLO successfully addresses the challenge of detecting leaf rolling in maize under complex environmental conditions.

In summary, the comparison of performance and detection results further underscores the effectiveness of LRD-YOLO for leaf rolling detection in maize under intricate environmental conditions.

Robustness in adverse weather conditions

Although object detection methods have shown encouraging outcomes when applied to high-quality datasets, the ongoing challenge lies in precisely localizing objects within low-quality images taken in adverse weather conditions (Liu et al. 2022). To assess the robustness of LRD-YOLO, we conducted experiments comparing its effectiveness to the baseline model in leaf rolling detection under adverse weather conditions.

As depicted in Fig. 10, our data augmentation techniques to include more severe conditions such as bright light, rain, and fog in our test sets. Moreover, we have simulated scenarios where water droplets can obscure the lens during rainy conditions, as well as instances of mud splattering caused by windy weather.

The detection results of LRD-YOLO and YOLOv8n are illustrated in Fig. 11. While LRD-YOLO demonstrates robust performance under rainy conditions, it occasionally experiences false positives and misses in foggy and bright light environments when lens-obscuring water droplets are present. In comparison, the YOLOv8n model shows significant issues with false positives and misses across all adverse environments tested. These findings highlight LRD-YOLO's effectiveness in enhancing the baseline method's resilience to adverse weather conditions, significantly improving object detection accuracy in challenging environments.

In addition to applying the aforementioned data augmentation methods to our test set, we have extended these techniques to our training and validation sets, resulting in a training set comprising 4088 images and a validation set of 490 images. Based on this augmentation, we trained the LRD-weather model, which has been specifically designed to excel in severe weather conditions while maintaining high detection accuracy.

The performance of YOLOv8n, LRD-YOLO, and LRD-WEATHER on the test set is detailed in Table 5, the bolded section in Table 5 highlights the model with the highest score under the corresponding weather conditions. As shown, LRD-YOLO consistently improves mAP in mild weather conditions by 2.9%, 2.3%, 3.2%, and 5.0% over YOLOv8n, respectively, while maintaining high accuracy. Our model's performance under severe weather conditions is also demonstrated. However, in more extreme scenarios, both YOLOv8n and LRD-YOLO exhibit significant performance degradation, with the mAP metric dropping below 50% at Spatter_Severe conditions. In contrast, due to robust data augmentation during training and validation, the LRD-WEATHER model maintains over 75% accuracy and mAP metrics under severe extreme weather conditions,

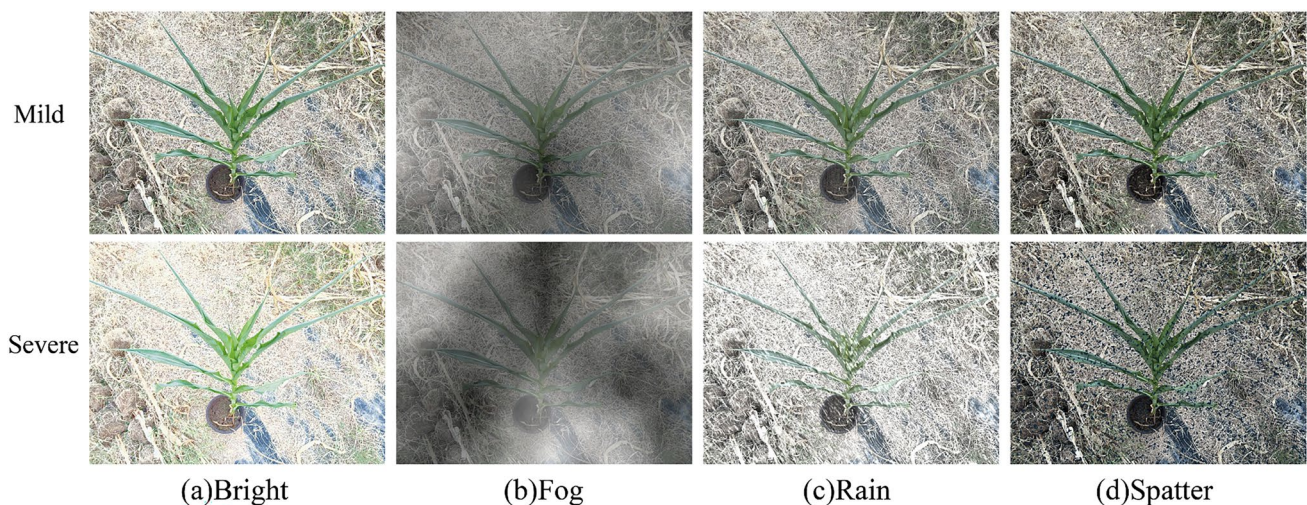


Fig. 10 Data augmentation for severe weather conditions

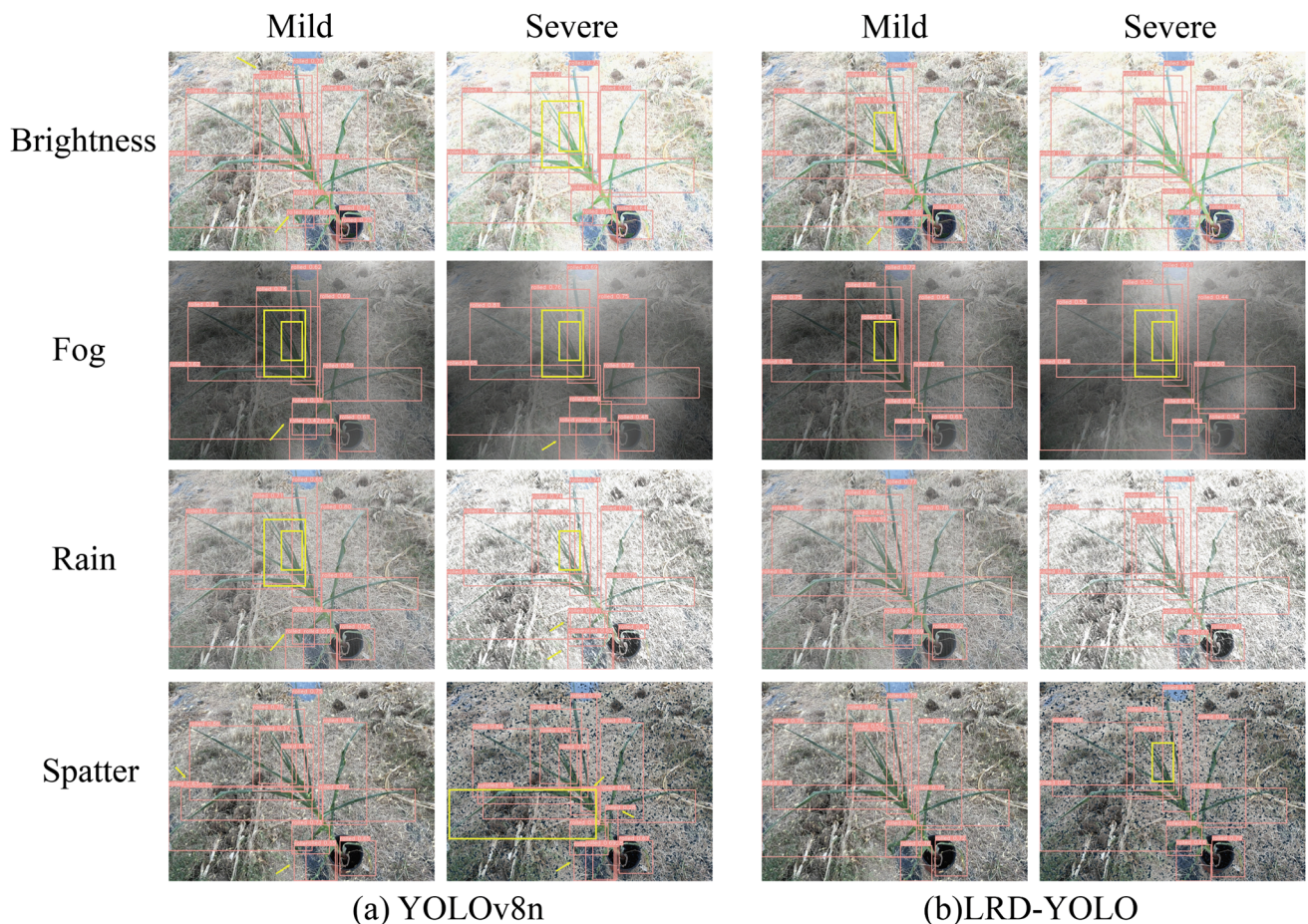


Fig. 11 Comparison of the detection results in adverse weather conditions. The arrow points to the incorrect results, and the yellow box represents the missing targets

showcasing its superior detection performance in challenging environments.

These results underscore the effectiveness of LRD-YOLO and LRD-WEATHER in enhancing the robustness of the baseline method against adverse weather conditions. They demonstrate the significant advancements our model brings to achieving precise object detection in challenging environmental contexts.

Discussion

Visualization of the detection results

To further underscore the efficacy of our improvements to the baseline model, we performed a detailed analysis of the results obtained by LRD-YOLO and the YOLOv8n model for maize leaf rolling detection. For this analysis, we utilized Grad-CAM (Selvaraju et al. 2020) visualization as a tool. Grad-CAM is designed to visualize the distinct contributions of various regions within a deep neural network to

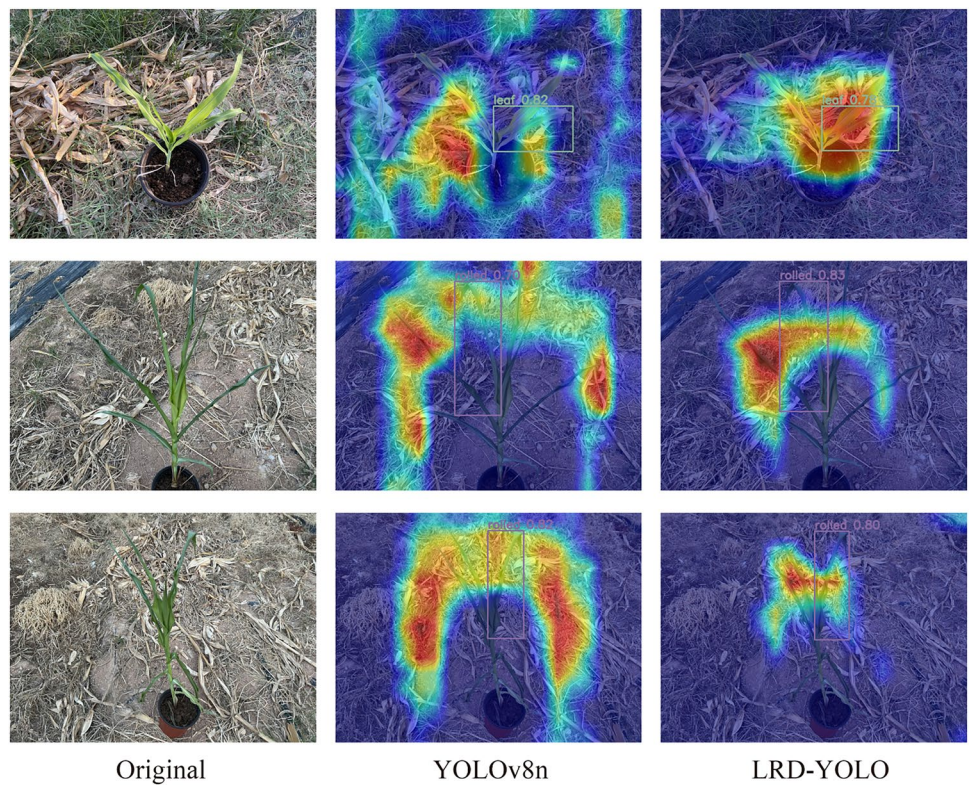
the prediction results. This method aids in pinpointing significant areas within images. Figure 12 presents a random selection of examples illustrating Grad-CAM visualizations generated by both LRD-YOLO and YOLOv8n on the test set. The Grad-CAM visualization provides valuable insights into the model's attention focus during leaf rolling detection in maize.

Upon careful examination of the Grad-CAM visualizations, our model exhibits a notable ability to concentrate on the specific area of the maize leaf where rolling occurs. For uncurled leaves, the model also maintains focus. The introduction of DCNv2 significantly enhances the model's proficiency in detecting leaves with diverse scale sizes and shape variations. In contrast, Grad-CAM visualizations from the YOLOv8n model display less precision, often extending to regions outside of the leaves. Remarkably, Grad-CAM visualizations from LRD-YOLO are characterized by increased focus and accuracy, capturing the key features of the leaves with precision. This underscores the excellent contribution of the CBAM module to our model. These findings highlight the effectiveness of our LRD-YOLO in improving the

Table 5 Performance comparison of YOLOv8n, LRD-YOLO and LRD-WEATHER

Weather conditions	Metrics	YOLOv8n	LRD-YOLO	LRD-WEATHER
Rain_Mild	mAP@0.5 (%)	74.4	77.3	77.6
	Precision (%)	73.9	77.0	78.6
	Recall (%)	67.7	68.2	73.8
Fog_Mild	mAP@0.5 (%)	66.9	69.2	77.1
	Precision (%)	69.5	71.2	80.9
	Recall (%)	61.3	61.8	71.1
Brightness_Mild	mAP@0.5 (%)	74.4	77.6	77.0
	Precision (%)	74.5	79.2	81.1
	Recall (%)	68.3	66.1	70.1
Spatter_Mild	mAP@0.5 (%)	73.8	78.8	77.0
	Precision (%)	78.7	80.7	82.9
	Recall (%)	63.1	70.9	69.6
Rain_Severe	mAP@0.5 (%)	63.5	64.6	75.2
	Precision (%)	61.7	62.3	80.0
	Recall (%)	58.6	60.3	68.2
Fog_Severe	mAP@0.5 (%)	59.1	66.6	76.2
	Precision (%)	60.9	71.0	77.4
	Recall (%)	55.0	62.5	72.2
Brightness_Severe	mAP@0.5 (%)	71.1	70.4	76.2
	Precision (%)	76.2	75.2	81.1
	Recall (%)	62.3	62.2	70.3
Spatter_Severe	mAP@0.5 (%)	42.5	47.5	76.5
	Precision (%)	43.2	57.5	76.4
	Recall (%)	42.1	44.6	75.4

Fig. 12 Grad-CAM visualization of LRD-YOLO and YOLOv8n



performance of the baseline YOLOv8n for detecting leaf rolling in maize. The LRD-YOLO model showcases an improved ability to navigate the complexities of the surrounding environment, ensuring robust performance even in the presence of interfering factors. The application of the Grad-CAM visualization technique further highlights the LRD-YOLO model’s enhanced focus on the key characteristics of maize leaves.

Lightweight improvement of the LRD-YOLO model

The integration of DCNv2 and CBAM significantly enhances the model's feature extraction and adaptability to shape and scale variations, but it also increases the complexity of the YOLOv8 model. These factors can pose challenges, particularly in resource limited settings such as small farms or remote areas without advanced computing infrastructure. Model complexity is as important a metric as accuracy, and while LRD-YOLO excels in accuracy, there are still opportunities for reduction in complexity.

We have taken steps to address the model's complexity and computational requirements. Specifically, we have employed the channel pruning algorithm (Layer-adaptive sparsity for the Magnitude-based Pruning) (Lee et al. 2020) to optimize the LRD-YOLO model. This approach aims to reduce network complexity by eliminating less critical channels, thereby improving computational efficiency. Detailed experimental results demonstrating the effectiveness of this optimization are provided in the following table.

As illustrated in Table 6, the pruned model demonstrates significant improvements over the original LRD-YOLO in terms of parameter reduction by 77.8%, 50% fewer FLOPs, and a 9% increase in inference speed. Importantly, despite these reductions, the pruned model maintains a marginal decrease of only 2.1% in mAP compared to the original, still surpassing the baseline YOLOv8n by 3%. This underscores the efficacy of our pruning strategy in balancing model complexity with performance.

Figure 13 visually represents the impact of our pruning approach on the convolutional layers, showcasing a substantial reduction in channel counts. This reduction signifies the successful optimization of model complexity, enhancing its suitability for resource constrained environments such as small farms and remote areas.

While our pruning efforts have significantly reduced the complexity of the model, we recognize that further improvements in inference speed are necessary. To address this challenge, we have explored alternative lightweight backbone networks as replacements for the original backbone in the LRD-YOLO model. Specifically, we evaluated MobileNetV3 (Howard et al. 2019), ShuffleNetV2 (Ma et al. 2018), and VanillaNet (Chen et al. 2023) with different layer configurations.

The experimental results presented in Table 7 highlight VanillaNet-9 as particularly promising, achieving a remarkable 52.5% improvement in inference speed compared to the original LRD-YOLO model. Although the accuracy of the model is reduced compared to LRD-YOLO, it is still slightly higher than the baseline model.

Table 6 Results of the pruning experiment

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	FPS	FLOPs (G)	Params (M)
YOLOv8n	76.4	68.5	76.5	77.3	8.9	3.16
LRD-YOLO	82.3	73.6	81.6	56.0	8.0	3.48
LRD-PRUNE	82.5	70.8	79.5	61.0	4.0	0.77

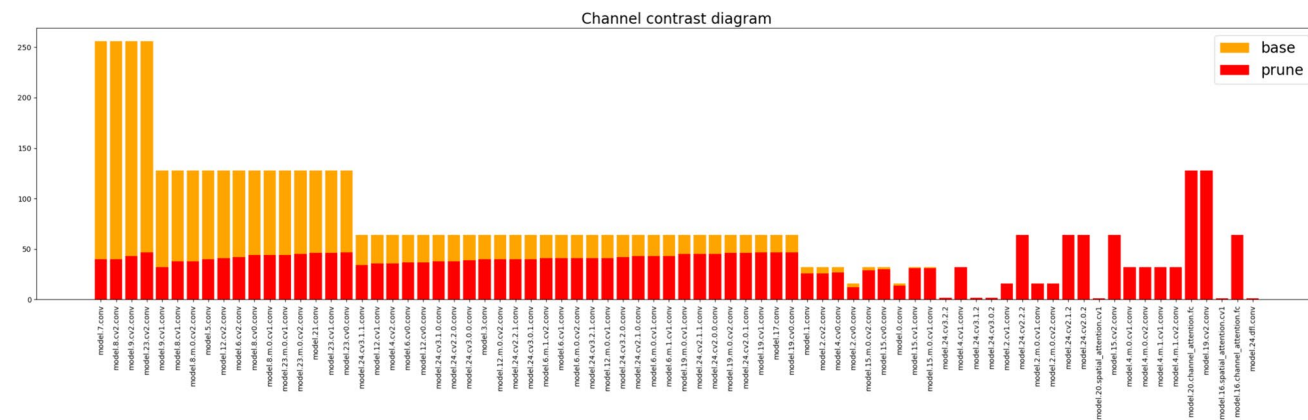


Fig. 13 Channels contrast of base and prune model

Table 7 Results of the backbone network experiment

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	FPS	FLOPs (G)	Params (M)
MobileNetV3	75.3	64.4	73.1	53.1	2.8	1.26
ShuffleNetV2	62.6	63.3	65.1	76.2	5.3	2.11
VanillaNet-5	77.3	65.4	72.3	87.9	5.3	2.12
VanillaNet-7	80.8	65.2	75.9	90.8	6.2	2.40
VanillaNet-9	80.1	67.9	76.7	85.4	6.4	2.47
VanillaNet-12	76.3	67.2	75.5	79.6	6.8	2.58

Inference speed is also improved over baseline. This enhancement is achieved while maintaining a low model complexity, demonstrating superior performance among the tested backbone networks.

Compared to other models in the YOLOv8 family (s, m, l), the YOLOv8n model stands out as the most lightweight variant. While the LRD-YOLO model introduces a slight increase in complexity compared to YOLOv8n, it remains a relatively lightweight solution suitable for a wide range of application scenarios.

Particularly for resource-constrained environments such as small farms or remote areas, the pruned LRD-YOLO model offers a practical and efficient solution. For scenarios demanding higher inference speeds, we have explored enhancing the LRD-YOLO model by integrating lightweight backbone networks like VanillaNet-9.

These optimizations directly address the concerns raised regarding computational demands and suitability for real-world agricultural applications. By significantly reducing model complexity while maintaining competitive performance metrics, our approach ensures that the pruned

LRD-YOLO model is well-equipped for practical deployment across varied agricultural settings.

Limitations

Our study's dataset, although diverse, may not be sufficiently large to capture all variations in leaf rolling across different maize varieties and environmental conditions. Advanced data augmentation methods could help enhance the dataset's diversity and richness, so we employed a comprehensive suite of seven methods, as illustrated in Fig. 14. These methods encompassed random cropping, cutout, brightness adjustment, flipping, noise addition, rotation, and shift.

The performance of LRD-YOLO after data augmentation is shown in the Table 8.

The rolling of maize leaves is a process that spans from mild to severe, manifesting phenotypic variations at different degrees of rolling. Although our suggested model can successfully accomplish the binary classification task of detecting rolled maize leaves, its efficacy is limited by the size of the dataset, impeding a comprehensive detection of

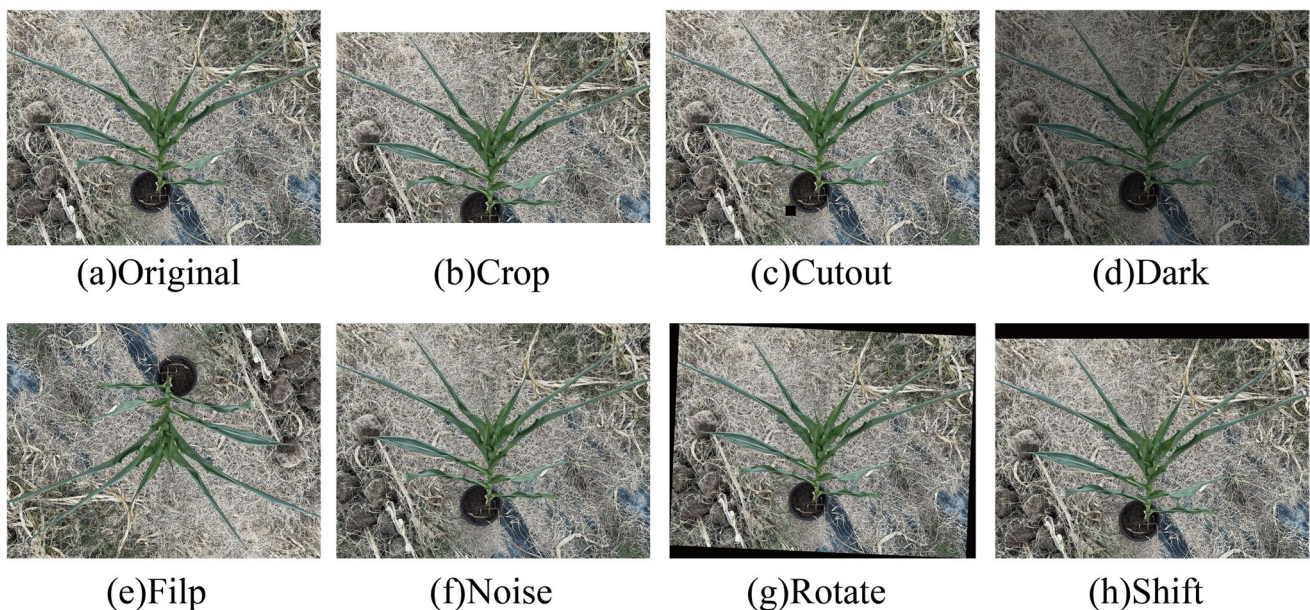
**Fig. 14** Example of data augmentation

Table 8 Results of data augmentation

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	Augmentation
LRD-YOLO	82.3	73.6	81.6	
LRD-YOLO	81.3	76.9	82.4	✓

the entire rolling process. Excessive classification leads to a decrease in the number of instances within each class, which poses challenges in properly training the model. Data augmentation alone cannot fundamentally address the issue of insufficient instances within each class and often leads to the problem of overfitting.

Moreover, the model requires a substantial amount of images to discern subtle differences in rolling degrees between different classes, a requirement not currently met by our dataset. In future work, we intend to establish a larger-scale dataset to delve deeper into the phenotypic characteristics of rolled maize leaves. And the imbalance across various stages of leaf rolling in our dataset is a critical issue that requires careful consideration as we expand our dataset. Future work will endeavor to cover leaf rolling caused by changes in soil type, climatic conditions and biotic stresses (e.g. pests and diseases) wherever possible. Our objective is to enhance the depth of the study and ultimately apply our research to field conditions.

Conclusion

We propose the LRD-YOLO model, an innovative approach for leaf rolling detection in maize with a focus on achieving high accuracy without compromising real-time inference speed. To initiate the study, a new leaf rolling dataset is meticulously collected, encompassing various challenges inherent in this task, such as severe occlusion, changes in leaf scale and shape, and complex background scenarios. The principal contributions of our approach involve integrating the CBAM mechanism into the YOLOv8 architecture. This integration enhances feature extraction capability and feature validity, thereby improving detection accuracy in occluded scenes and complex environments. Additionally, we introduce DCNv2 to better adapt to changes in target shape and scale. Following conducting experiments, our findings underscore the role of the LRD-YOLO in significantly improving detection accuracy for leaf rolling in maize, surpassing existing methods while maintaining real-time inference capabilities.

Author contributions YW contributed to conceptualization, data curation, investigation, methodology, software, validation,

writing—original draft, and writing—review and editing, as well as visualization. XJ was involved in data curation, investigation, methodology, validation, and writing—review and editing. YG participated in data curation, investigation, methodology, and writing—review and editing. XH and CZ contributed to methodology, project administration, supervision, and writing—review and editing. WP played a role in conceptualization, funding acquisition, methodology, project administration, supervision, and writing—review and editing.

Funding This work has been supported by the National Natural Science Foundation of China (Grant Nos. 32100501 and no.32300239), Shenzhen Science and Technology Program (Grant No. RCBS20210609103819020), the Innovation Program of Chinese Academy of Agricultural Sciences, National Key R&D Program of China (Grant No. 2023ZD04076).

Data availability Some of the data, source codes and more details about our project are in the GitHub (<https://github.com/WangYH1740/LRD-YOLO>). In addition, the original datasets are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors declare no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Bänziger M, Edmeades GO, Beck D, Bellon M (2000) Breeding for drought and nitrogen stress tolerance in maize: from theory to practice. CIMMYT, Mexico
- Baret F, Madec S, Irfan K, Lopez J, Comar A, Hemmerle M, Dutartre D, Praud S, Tixier MH (2018) Leaf-rolling in maize crops: from leaf scoring to canopy-level measurements for phenotyping. *J Exp Bot* 69:2705–2716
- Chen H, Wang Y, Guo J, Tao D (2023). VanillaNet: the power of minimalism in deep learning. <https://arxiv.org/abs/2305.12972>
- Clarke JM (1986) Effect of leaf rolling on leaf water loss in *Triticum* spp. *Can J Plant Sci* 66(4):885–891
- Dai J, Qi H, Xiong Y, Li Y, Zhang G, Hu H, Wei Y (2017) Deformable convolutional networks. <https://arxiv.org/abs/1703.06211>
- Darwish A, Ezzat D, Hassanien AE (2020) An optimized model based on convolutional neural networks and orthogonal learning particle swarm optimization algorithm for plant diseases diagnosis. *Swarm Evol Comput* 52:100616
- Ding Y, Li Z, Peng S (2020) Global analysis of time-lag and -accumulation effects of climate on vegetation growth. *Int J Appl Earth Observ Geoinformation* 92:102179

- Farhangfar S, Bannayan M, Khazaei HR, Baygi MM (2015) Vulnerability assessment of wheat and maize production affected by drought and climate change. *Int J Disaster Risk Reduct* 13:37–51
- Jocher G (2020) YOLOv5 by ultralytics. <https://github.com/ultralytics/yolov5>
- Jocher G, et al (2023) Ultralytics YOLO. <https://github.com/ultralytics/ultralytics>
- Ge Z, Liu S, Wang F, Li Z, Sun J (2021) YOLOX: exceeding YOLO series in 2021. <https://arxiv.org/abs/2107.08430>
- Howard AG, Sandler M, Chu G, Chen L-C, Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V, Le QV, Adam H (2019) Searching for MobileNetV3. *IEEE/CVF Int Conf Comput vis (ICCV)* 2019:1314–1324
- Jiang Y, Li C (2020) Convolutional neural networks for image-based high-throughput plant phenotyping: a review. *Plant Phenomics* 2020:4152816
- Kadioglu A, Terzi R (2007) A dehydration avoidance mechanism: leaf rolling. *Bot Rev* 73:290–302
- Kadioglu A, Terzi R, Saruhan N, Saglam A (2012) Current advances in the investigation of leaf rolling caused by biotic and abiotic stress factors. *Plant Sci* 182:42–48
- Lee J, Park S, Mo S, Ahn S, Shin J (2020) Layer-adaptive sparsity for the magnitude-based pruning. *International conference on learning representations*
- Li C, Li L, Jiang H, Weng K, Geng Y, Li L, Ke Z, Li Q, Cheng M, Nie W, Li Y, Zhang B, Liang Y, Zhou L, Xu X, Chu X, Wei X, Wei X (2022) YOLOv6: a single-stage object detection framework for industrial applications. <https://arxiv.org/abs/2209.02976>
- Liu J, Wang X (2020) Early recognition of tomato gray leaf spot disease based on MobileNetV2-YOLOv3 model. *Plant Methods* 16:83
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC (2016) SSD: single shot multibox detector. *Computer Vision – ECCV 2016*, pp 21–37
- Liu S, Qi L, Qin H, Shi J, Jia J (2018) Path aggregation network for instance segmentation. <https://arxiv.org/abs/1803.01534>
- Liu W, Ren G, Yu R, Guo S, Zhu J, Zhang L (2022) Image-adaptive YOLO for object detection in adverse weather conditions. *Proc AAAI Conf Artif Intell* 36:1792–1800
- Ma N, Zhang X, Zheng HT, Sun J (2018) ShuffleNet V2: practical guidelines for efficient CNN architecture design. Springer, Cham pp 122–1388
- Martinelli F, Scalenghe R, Davino S, Panno S, Scuderi G, Ruisi P, Villa P, Stroppiana D, Boschetti M, Goulart LR, Davis CE, Dandekar AM (2014) Advanced methods of plant disease detection. A review. *Agron Sustain Dev* 35:1–25
- Mehdipour Ghazi M, Yanikoglu B, Aptoula E (2017) Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing* 235:228–235
- Oo YM, Htun NC (2018) Plant leaf disease detection and classification using image processing. *Int J Res Eng* 5:516–523
- Pal A, Kumar V (2023) AgriDet: plant leaf disease severity classification using agriculture detection framework. *Eng Appl Artif Intell* 119:105754
- Premachandra GS, Saneoka H, Fujita K, Ogata S (1993) Water stress and potassium fertilization in field grown maize (*Zea mays* L.): effects on leaf water relations and leaf rolling. *J Agron Crop Sci* 170:195–201
- Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 39:1137–1149
- Saruhan N, Saglam A, Kadioglu A (2011) Salicylic acid pretreatment induces drought tolerance and delays leaf rolling by inducing antioxidant systems in maize genotypes. *Acta Physiol Plant* 34:97–106
- Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2020) Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vision* 128:336–359
- Sirault XR, Condon AG, Wood JT, Farquhar GD, Rebetzke GJ (2015) “Rolled-upness”: phenotyping leaf rolling in cereals using computer vision and functional data analysis approaches. *Plant Methods* 11:52
- Tanumihardjo SA, Mcculley L, Roh R, Lopez-Ridaura S, Palacios-Rojas N, Gunaratna NS (2020) Maize agro-food systems to ensure food and nutrition security in reference to the sustainable development goals. *Glob Food Secur* 25:100327
- Thai H-T, Le K-H, Nguyen NL-T (2023) FormerLeaf: an efficient vision transformer for cassava leaf disease detection. *Comput Electron Agric* 204:107518
- Tzutalin (2015). LabelImg. <https://github.com/tzutalin/labelImg>
- Ubbens J, Cieslak M, Prusinkiewicz P, Stavness I (2018) The use of plant models in deep learning: an application to leaf counting in rosette plants. *Plant Methods* 14:6
- Waldchen J, Mader P (2018) Plant species identification using computer vision techniques: a systematic literature review. *Arch Comput Methods Eng* 25:507–543
- Waldchen J, Rzanny M, Seeland M, Mader P (2018) Automated plant species identification-trends and future directions. *PLoS Comput Biol* 14:e1005993
- Wang C-Y, Bochkovskiy A, Liao H-YM (2022) YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. <https://arxiv.org/abs/2207.02696>
- Woo S, Park J, Lee J-Y, Kweon IS (2018) CBAM: convolutional block attention module. *Proceedings of the European conference on computer vision (ECCV)*, pp 3–19
- Zhang GH, Xu Q, Zhu XD, Qian Q, Xue HW (2009) SHALLOT-LIKE1 is a KANADI transcription factor that modulates rice leaf rolling by regulating leaf abaxial cell development. *Plant Cell* 21:719–735
- Zhao Y, Lv W, Xu S, Wei J, Wang G, Dang Q, Liu Y, Chen J (2023) DETRs beat YOLOs on real-time object detection. <https://arxiv.org/abs/2304.08069>
- Zhu X, Hu H, Lin S, Dai J (2019) Deformable ConvNets v2: More deformable, better results. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp 9308–9316

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.