Litchi flower targets integrating multi-scale attention and decoupled detection

by Deng, X., Wang, Y., Sun, H., Shen, Z., Tang, W., Cheein, F.A. and Chen, H.

Copyright, publisher and additional information: Publishers' version distributed under the terms of the Creative Commons Attribution License

DOI link to the version of record on the publisher's site

Deng, X., Wang, Y., Sun, H., Shen, Z., Tang, W., Cheein, F.A. and Chen, H. (2026) 'Litchi Flower Targets Integrating Multi-Scale Attention and Decoupled Detection', *Smart Agricultural Technology*, 14, article 102179.

Smart Agricultural Technology 14 (2026) 102179

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Litchi flower targets integrating multi-scale attention and decoupled detection

Xiaoling Deng [a,b,c], Yiwei Wang [a], Heguang Sun [a], Zifan Shen [a], Wenqian Tang [a], Fernando Auat Cheein [d,*], Haibo Chen [a,*]

[a] College of Artificial Intelligence and Low-Altitude Technology, South China Agricultural University, Guangzhou 510642, China
[b] National Center for International Collaboration Research on Precision Agricultural Aviation Pesticide Spraying Technology, Guangzhou 510642, China
[c] Guangdong Engineering Technology Research Center of Smart Agriculture, Guangzhou 510642, China
[d] Department of Engineering, Harper Adams University, Edgmond, Newport TF10 8NB England, United Kingdom

ARTICLE INFO

Keywords:
Litchi flower detection
UAV-based remote sensing
Deep learning
Small-object

ABSTRACT

In commercial litchi production, the scientific regulation of floral load is critical for ensuring consistent yields and high-quality fruits. However, litchi exhibits profuse and concentrated flowering. Accurately detecting these dense floral clusters in complex orchard environments is highly challenging. To address this challenge, this study proposes the Fusion Decoupled and Multi-scale Attention Network (FDMA-Net), an improved lightweight object detection model. The model employs High Performance GPU Network V2 (HGNetV2) as the backbone, incorporating depthwise separable convolutions (DWConv) to create a strong yet lightweight feature-extraction framework. A parameter-free simple attention module (SimAM) is integrated at the backbone output to enhance fine-grained feature discrimination. The C3k2_Star module, an enhanced cross-stage feature fusion block, reinforces context awareness and multi-scale feature fusion. Finally, a decoupled detection head integrating the spatial enhancement attention module (Detect_SEAM) leverages task decoupling and feature enhancement to improve localization accuracy and confidence in dense scenes. Experimental results indicate that the proposed model achieves a mean Average Precision (mAP@0.5) of 94.1%, outperforming the YOLOv11n baseline by 4.2%. Simultaneously, it reduces both the model size and floating-point operations (FLOPs) by 18.9% and 17.5%, respectively. Furthermore, robustness evaluations confirm that the model maintains high stability under non-ideal conditions. Consequently, the proposed method delivers superior performance in litchi flower detection, making it highly suitable for real-time, high-precision deployment on resource-constrained edge devices.

1. Introduction

Litchi, an important subtropical evergreen fruit tree native to China, occupies a prominent position in domestic and international fruit markets owing to its distinctive flavor and high nutritional value [1]. However, in litchi orchard management, the scientific regulation of floral load remains a key limiting factor for improving yield and quality [2]. Litchi exhibits profuse flowering and a concentrated flowering period, yet not all flowers successfully set fruit. Furthermore, the complexity of the flowering process, together with variable field conditions, directly influences fruit set and final yield [3,4]. Effective floral load management is thus critical to optimize fruit set and improve overall yield. If the load is excessive, flowers will deplete nutrients, which reduces fruit set and hinders fruit development. Conversely, an insufficient effective floral load leads to inadequate pollination and consequent yield loss [5]. Therefore, regulating the overall floral load and improving floral quality are core tasks for ensuring consistent yields and high-quality fruits in litchi orchards. In commercial production, these objectives are typically achieved through flower thinning [6].

Currently, three methods are used for litchi flower thinning, namely, manual, chemical, and mechanical thinning [7]. Manual thinning is labor-intensive and inefficient [8]. Chemical thinning removes flowers by spraying chemical agents onto the tree canopy, although relatively low in cost, it requires precise control of dosage and application timing and entails risks [9]. By contrast, mechanical thinning has substantial potential to reduce labor requirements and represents a key direction for

* Correspondence author.
E-mail addresses: FAuat@harper-adams.ac.uk (F. Auat Cheein), huanongchb@scau.edu.cn (H. Chen).

https://doi.org/10.1016/j.atech.2026.102179
Received 11 March 2026; Received in revised form 7 April 2026; Accepted 2 May 2026
Available online 5 May 2026
2772-3755/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

the future development of flower-thinning technologies. However, research on mechanical flower thinning remains at an early stage. Most current approaches operate in a non-selective manner. They lack accurate sensing of floral distribution and the capability for selective actuation, which constitutes a key bottleneck for further development [10]. To overcome this bottleneck, image processing and machine vision can be integrated to enable the accurate, real-time detection of litchi flowers. This integration provides critical visual input for next-generation intelligent machinery. Ultimately, it serves as a core technology for advancing mechanical flower thinning from "extensive" to "precision" management [11].

In recent years, acquiring high-resolution crop imagery with Unmanned Aerial Vehicles (UAVs) and applying machine learning—particularly deep learning—for crop growth monitoring and diagnosis has become a mainstream research focus [12].Barreto et al [13] developed a UAV imagery analysis workflow based on a fully convolutional network (FCN) that performs stem localization and pixel-level classification of crops and weeds in beet, maize, and strawberry fields, limiting counting error to 4.6% and substantially reducing the manual counting workload. Subsequently, Zhao et al [14] proposed Oriented and Small Wheat Spike Detection (OSWSDet), a YOLOv5-based detector that integrates Circular Smooth Label (CSL), a micro-scale detection layer, and Rotated Non-Maximum Suppression (RNMS), incorporates wheat-ear orientation features, and refines localization loss, achieving 90.5% mAP for UAV-based wheat-ear detection under complex field conditions. Furthermore, Reedha et al [15] applied the Vision Transformer (ViT) to crop–weed classification in high-resolution UAV-based remote-sensing imagery, leveraging data augmentation and transfer learning to outperform state-of-the-art models such as ResNet and EfficientNet under limited-data settings; specifically, the Vision Transformer-Base 16 (ViT-B16) model achieved a 0.994 F1-Score, surpassing conventional Convolutional Neural Network (CNN) models. Currently, there is a distinct lack of deep learning models specifically applied to litchi flower detection in the existing literature. Given this research gap, the aforementioned studies are reviewed to demonstrate the broad feasibility and significant potential of utilizing UAV-based remote sensing for detecting small, densely distributed agricultural targets. However, despite the significant progress achieved by these models in UAV-based agricultural detection, limitations persist in the fine-grained recognition of extremely small and densely distributed targets (e.g., flowers), particularly when addressing technical challenges associated with severe occlusion and complex background interference.

In an effort to achieve effective recognition under such complex conditions, previous research on flower detection has primarily explored two major paths based on application objectives: one focuses on detection and localization for precise agricultural operations, while the other focuses on automatic counting and estimation for yield prediction. For example, Shang et al [16] proposed a real - time apple flower detection method based on lightweight YOLOv5s. By replacing the backbone network of the original YOLOv5s with ShuffleNetv2 and replacing the Conv module in the Neck part with the Ghost module, the model size was significantly reduced and the detection speed was increased. This improved model showed good generalization in complex natural environments, with mAP@0.5 reaching 91.8% and a detection speed of 86.21 fps, demonstrating that the lightweight structure can significantly improve real - time performance while maintaining accuracy. Mu et al [17] established an instance segmentation model for apple flowers and king flowers based on Mask Region-based Convolutional Neural Network (Mask R – CNN), achieving the automatic recognition and localization of single king flowers in a natural orchard environment for the first time. This study effectively improved the small-object detection ability by introducing the Feature Pyramid Network (FPN) and ResNet-101 backbone. The mean Average Precision (mAP) for flower detection reached 86%, and the recognition accuracy of king flowers ranged from 65.6% to 98.7% at different flowering stages, significantly outperforming traditional color thresholding methods and semantic segmentation methods. In scenarios with a high density of flowers, Estrada et al [18] proposed a counting method using ground-based images of peach groves. To overcome detection difficulties caused by severe occlusion, they replaced traditional bounding-box detection with a multi-column convolutional neural network (MCNN), which regresses multi-scale flower features into a density map. While their results showed this regression method achieved lower counting errors than standard YOLO algorithms, density maps primarily estimate overall counts and lack the ability to precisely localize individual flowers. Motivated by the limitations of traditional detectors in such environments, the focus of this research is to construct a lightweight and robust detection framework capable of precisely recognizing extremely small and densely distributed targets, such as litchi flowers, amidst complex background interference.

This study proposes FDMA-Net, an improved, lightweight, high-precision detection model for the general problem of recognizing small, densely clustered, and heavily occluded targets, using litchi flowers as a representative case study. The specific objectives are as follows: (1) To design a lightweight and efficient backbone network with strong small-object sensitivity, by adopting HGNetV2 and combining it with DWConv. (2) To enhance feature fusion and contextual modeling for objects at different scales, by employing the C3k2_Star module to address the receptive field limitations in the conventional YOLO neck. (3) To improve the discrimination of fine-grained floral features and mitigate the inherent task conflicts between object classification and bounding box localization in dense scenes, by integrating the parameter-free SimAM attention mechanism and incorporating a decoupled detection head, Detect_SEAM. Comparative experiments with multiple mainstream models validate the proposed model's overall performance advantages across detection accuracy, model size, and inference efficiency.

2. Materials and methods

2.1. Image acquisition and dataset preparation

The UAV-based remote-sensing imagery used in this study was captured at the Litchi Culture Expo Park in Conghua District, Guangzhou, Guangdong Province, China (23.56° N, 113.61° E). Data collection was conducted using the DJI M30T enterprise drone, equipped with a three-axis mechanically stabilized gimbal. Fig. 1 provides an overview of the study area and the UAV platform. Its wide-angle camera is equipped with a 1/2-inch CMOS sensor (12 million pixels) and a lens with an equivalent focal length of 24 mm. Automatic flight missions were planned through FlightHub 2, and image collection covered 34 litchi trees, spanning from early March to mid-to-late April in 2024 and 2025, covering a two-year period. During these monitoring periods, flight missions were executed once a week, with data collection consistently taking place between 2:00 PM and 4:00 PM. The flight protocol was set as follows: the drone automatically flew and hovered at three sampling points (a total of 12 points) in each direction (east, south, west, and north) of each tree for photography. The flight height was kept constant at 6 m, and the camera gimbal maintained a tilt angle of 60° to align with the side of the tree canopy, ensuring clear imaging of litchi flowers while avoiding interference from the propeller airflow on the litchi branches. In total, over 600 remote-sensing images were collected.

Based on the curated high-resolution remote-sensing imagery, we used the LabelImg software (https://github.com/HumanSignal/labelImg) to manually annotate flowers with minimum bounding rectangles (MBRs). Given the large image footprint of UAV-based imagery and the small size of the objects, to improve the model's sensitivity to tiny objects, we tiled each original image into 1024 × 1024- pixel patches with 20% overlap to preserve the semantic integrity of objects at tile boundaries. Importantly, to prevent potential data leakage, the dataset split into training, validation, and test sets was strictly performed on the original full images prior to this cropping process. To improve the

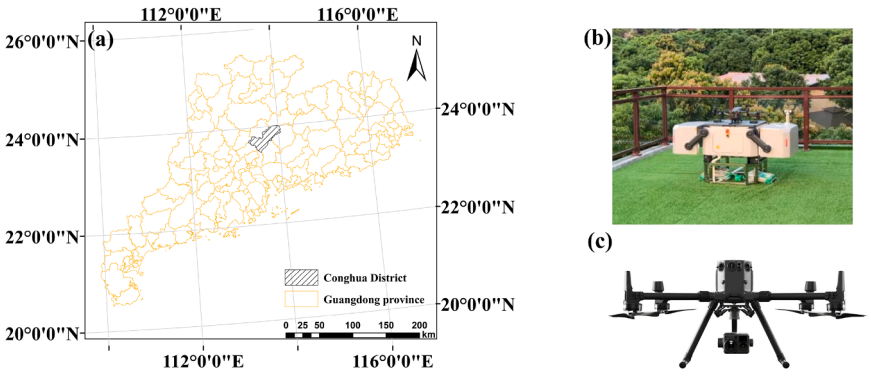

Fig. 1. Overview of the study area (a) Specific study location (b) DJI-deployed airport (c) The UAV used.

model's baseline robustness, we applied specific data augmentation techniques during the training process. Alongside structural augmentations such as MixUp, Mosaic, and random flips, we digitally simulated basic natural-light variations by randomly adjusting brightness, contrast, and gamma. To further challenge the model, the dataset was deliberately curated to include specific natural hard examples, namely highly dense floral clusters, severe inter-cluster occlusio, and targets with blurred color boundaries against the background foliage. Fig. 2 shows representative samples from the dataset, illustrating both natural captures and these digitally modified lighting conditions. Following these procedures, a benchmark comprising 3400 images was compiled and split into training (2380 images), validation (680 images), and test (340 images) sets at a ratio of 7:2:1, providing a solid basis for model training and evaluation. While these digital augmentations incrementally improve the model's stability against basic optical fluctuations, we acknowledge that true extreme field conditions introduce complex illumination effects that cannot be fully replicated digitally. Thus, this dataset primarily serves to validate the robust detection of litchi flowers, rather than guaranteeing absolute generalization across all meteorological environments.

2.2. The proposed FDMA-Net detection model

To address the detection challenges in UAV-based remote-sensing imagery—namely the small size of litchi flowers, heavy occlusions in dense scenes, and strong background clutter—we propose an improved object detector, FDMA-Net. Built on YOLOv11n, the model introduces systematic architectural optimizations and innovations; its overall framework is shown in Fig. 3. The core improvements focus on four aspects: adopting HGNetV2 to build a lightweight backbone that improves feature-extraction efficiency, introducing the C3k2_Star module to strengthen multi-scale feature fusion, integrating SimAM attention to enhance the discriminability of fine-grained features, and designing a decoupled detection head, Detect_SEAM, to improve localization and classification in dense scenes.

2.2.1. HGNetV2 backbone network

In UAV-based remote-sensing imagery, litchi flower objects are typically small, exhibit diverse poses, occur at high densities, and are readily occluded by foliage; these characteristics pose significant challenges for the feature extraction of deep-learning models [19]. In the

Fig. 2. Example images from the dataset, including natural captures (a, d, e) and digitally simulated lighting variations (b, c) applied for data augmentation.

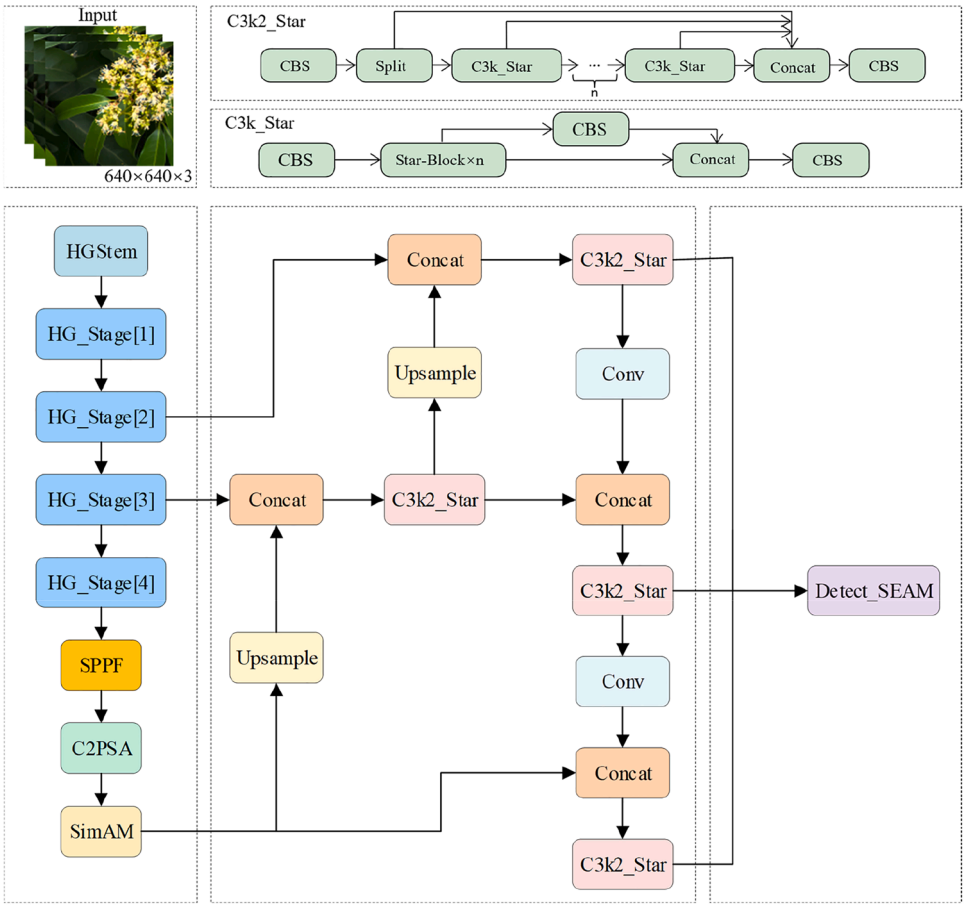

Fig. 3. Framework of the FDMA-Net Network. HGStem denotes the initial feature processing stage; HG_Stage [1] to HG_Stage [4] represent the four feature extraction stages (which will be explained later); C3k2_Star is a feature fusion module composed of CBS and Star_Block; the CBS module consists of a Convolutional layer, Batch Normalization, and a SiLU activation function in sequence; n represents the number of C3k_Star modules.

shallow layers of YOLO-series backbones, limited receptive fields and downsampling-induced spatial-information loss constrain the extraction of fine-grained cues (e.g., edges and textures) from small objects, often yielding under-represented features and, in turn, degraded detection accuracy [20].To address this issue, we introduce HGNetV2 [21] as the feature-extraction backbone (Fig. 4) and optimize it with DWConv, aiming to construct a feature-extraction framework that combines strong representational capacity with an outstanding lightweight design, thereby improving recognition accuracy on fine-grained objects and enhancing detection robustness.

HGNetV2 is an advanced backbone network specifically designed for high - resolution image and small - object detection. The overall network structure consists of modules such as HGStem, DWConv, and HGBlock, and feature extraction is mainly divided into four stages. HGStem first performs a 3 × 3 convolution on the input image, and then divides the feature map into two paths: one path goes through max - pooling, and the other path goes through two 2 × 2 convolutions to adjust the number of channels and size. After the results of the two paths are concatenated, two more convolutions are performed, and finally, a feature map with a size of 1/4 of the original image is output, thus reducing the computational load. The four HG_Stages of feature extraction are composed of DWConv and HGBlock modules. As shown in Fig. 4, the HGBlock module consists of a large number of standard convolutions. In HG_Stage [3] and HG_Stage[4], the HGBlock adopts multiple LightConv lightweight convolution methods, concatenating the convolution results of each layer with the input to increase the richness and diversity of target features, and then performing compression and excitation operations to output feature information. At the front - end of the HGBlock, DWConv is used. DWConv decomposes standard convolution into a depthwise spatial convolution followed by a 1 × 1 pointwise convolution [22]. While significantly reducing the number of parameters and FLOPs, its performance is close to that of the standard convolution. Therefore, we closely integrate DWConv into the multi - branch structure of HGNetV2 and use it as the core computing unit of the HGBlock. This design significantly reduces the computational overhead while retaining feature diversity, thus reducing the high - performance network to a compact model size and laying a solid foundation for the lightweighting of the model.

2.2.2. *SimAM spatial attention mechanism*

Litchi flower objects are small and densely clustered; their key discriminative cues (e.g., petals and stamens) are visually subtle, making this a typical fine-grained recognition task. Under heavy background clutter from leaves, branches, and shadows, these subtle yet critical cues are easily overwhelmed, making it difficult for the model to attend to relevant flower regions during feature extraction and thereby causing misses and false alarms. To strengthen the model's ability to resolve intra-object details, we introduce SimAM attention at the backbone output (Fig. 5) to reweight and enhance deep semantic feature maps, yielding more refined feature representations.

SimAM is a parameter-free attention mechanism with three-dimensional weights (3D weights), inspired by human visual attention; unlike conventional channel attention (e.g., SENet) or spatial attention, it aims to estimate the importance of every neuron in a feature map [23] . It does so by defining a simple yet effective energy function that quantifies the linear separability between a given neuron and the

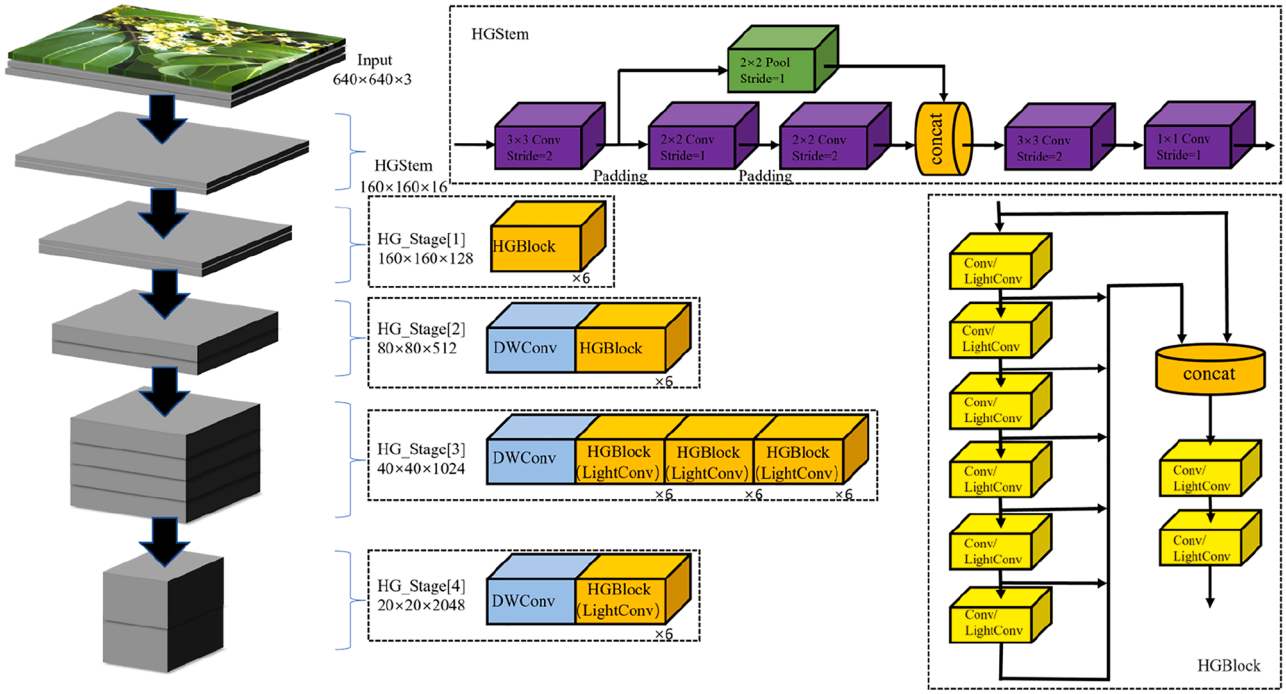

Fig. 4. Backbone Network Structure of HGNetV2. Pool denotes the pooling operation; DWConv denotes the depthwise convolution operation; LightConv denotes the lightweight convolution operation; Padding denotes the padding operation; HGBlock denotes the feature extraction module; 160 × 160 × 16, 160 × 160 × 128, 80 × 80 × 512, 40 × 40 × 1024, and 20 × 20 × 2048 are the dimensions of the corresponding output feature maps, respectively.

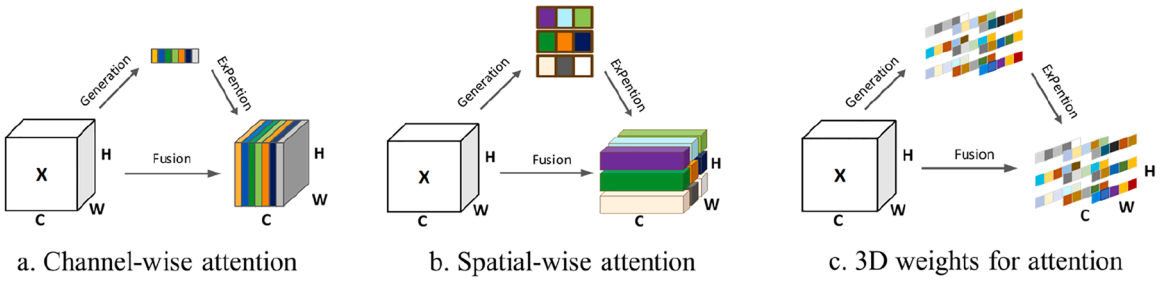

Fig. 5. Structure of the SimAM spatial attention mechanism. *X* is the input feature map, where *H*, *W*, and *C* denote the height, width, and number of channels of the feature map, respectively.

other neurons within the same channel. A lower energy indicates that a neuron is more dissimilar from its neighboring neurons and thus carries more distinctive and important information. Leveraging this energy function, SimAM efficiently computes per-neuron weights within a three-dimensional attention map without introducing additional learnable parameters, thereby maintaining a lightweight design while enabling fine-grained feature refinement [24].

The placement of the SimAM module is a deliberate architectural choice targeting the critical transition point between the backbone and the neck. The backbone's output at this stage represents the network's most semantically rich feature maps. However, these maps inevitably contain mixed signals: both critical fine-grained floral features and significant background clutter. Placing SimAM at this specific juncture—before the features enter the multi-scale fusion neck—allows it to act as an adaptive spatial and channel filter. It operates on these high-level semantic maps to refine the features in advance, adaptively suppressing weak background responses while enhancing features that are critical for flower recognition (e.g., neurons that encode petal contours or the morphology of flower clusters). By "pre-filtering" the feature maps in this way, SimAM ensures that the subsequent C3k2_Star neck module receives a more focused and discriminative input. This markedly increases the model's ability to sense and represent fine-grained features, thereby improving the detection and recognition accuracy of litchi flowers in complex backgrounds.

2.2.3. C3k2_Star module

In UAV-based remote-sensing imagery, variations in flight altitude and viewing angle cause litchi flower objects to exhibit pronounced scale variation, along with edge blurring and non-uniform spatial distribution. This requires detection models to not only resolve fine-grained features but also effectively aggregate multi-level contextual semantics; therefore, we integrate an improved fusion module—the C3k2_Star module [25] (Fig. 3).The core idea of this module is to replace the stack of standard bottleneck blocks in C3k2 with the Star-Block from StarNet, while preserving C3k2's lightweight design and introducing an explicit multi-scale feature-extraction mechanism to enhance information flow and semantic coupling.

C3k2_Star module follows the Cross Stage Partial (CSP) design paradigm at the macro level to improve computational efficiency and gradient propagation. Specifically, the input feature map F_{in} is first partitioned along the channel dimension into two parallel branches,F_{split1} and F_{split2}. The F_{split1} branch acts as a shortcut connection that undergoes only minimal convolutional processing, thereby preser-

ving a significant amount of the original feature information. The F_{split2} branch, in contrast, is fed into the core stack of Star-Blocks for in-depth, multi-scale feature extraction, following the methodology established by Jin et al [26] and Xu et al [27].

Star-Block is crucial for multi-scale perception. Internally, it employs a parallel, multi-branch architecture comprising one primary path and multiple auxiliary branches. By using convolutions configured with different receptive fields, the auxiliary branches capture spatial context at varying ranges. The outputs from all branches are adaptively aggregated using learnable attention weights α_i, and the computation can be defined as:

$$F_{out} = F_{main} + \sum_{i=1}^{n} \alpha_i \cdot R_i(F_{in}) \tag{1}$$

$$F_{C3k2_{Star}} = Concat\left(F_{split1}, \mathscr{B}_{Star}\left(F_{split2}\right)\right) \tag{2}$$

Here, F_{main} denotes the output of the primary path, and R_i denotes the nonlinear mapping of the i-th auxiliary branch. This design enables Star-Block to dynamically fuse local details with global context within a unified architecture. Ultimately, the output from the shortcut path, F_{split1} is concatenated with the output of F_{split2} (after being processed by the Star-Block) to form the final output of the C3k2_Star module. By integrating the efficient CSP paradigm with the Star-Block multi-scale unit, the C3k2_Star module substantially strengthens the network's ability to model pronounced scale variation, thereby maintaining robust discriminative performance when processing both long-range small objects and close-range large objects.

2.2.4. Detect_SEAM decoupled detection head

In UAV-based remote sensing for litchi-flower detection, the performance of the detection head largely governs the accuracy of localization and classification. Because litchi flowers are typically densely clustered and heavily occluded, a traditional coupled detection head that optimizes classification and regression along a single feature pathway is prone to objective conflict, leading to missed detections and localization errors for occluded or boundary-blurred flowers. To address this core challenge, we adopt Detect_SEAM, a decoupled detection head (Fig. 6), which centers on two principal strategies: task decoupling and feature enhancement [28].

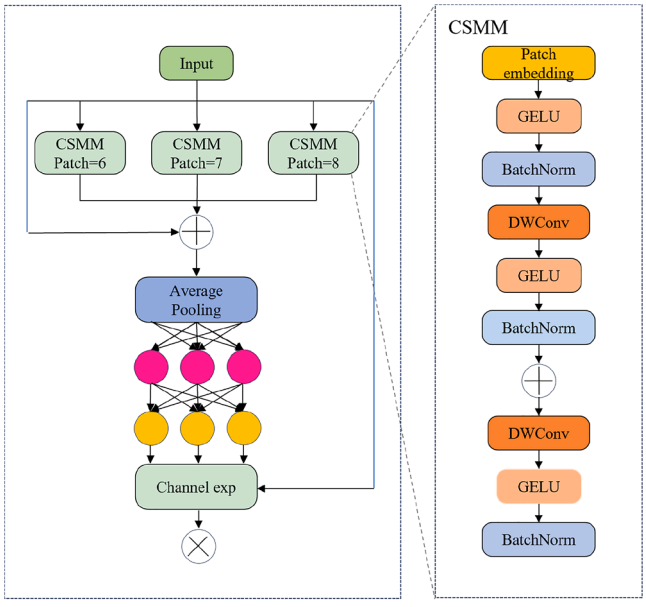

Fig. 6. The structural diagram of the core SEAM module. The SEAM architecture is illustrated on the left, with the internal structure of its CSMM (channel and spatial mixing module) component shown on the right. The CSMM employs different patch sizes to capture multi-scale features and uses depth-separable convolution to learn the correlation between spatial and channel dimensions.

It adopts parallel, decoupled branches that provide independent, separately optimized subnetworks for classification and regression. Here, the classification branch uses standard convolutions to enhance the representation of fine-grained semantic features such as stamens and petals; the regression branch primarily employs lightweight DWConv to efficiently model the object's geometric boundaries. This structural separation enables the two tasks to learn independently within their respective representation spaces, fundamentally avoiding feature competition and objective conflict, and significantly improving both training stability and final localization accuracy.

Building on this, Detect_SEAM integrates the Spatial Enhancement Attention Module (SEAM) [29,30] into both of the decoupled branches to strengthen feature representations under challenging conditions. The core of this enhancement lies in the SEAM module itself (Fig. 6), which is explicitly designed to tackle occlusion and multi-scale challenges. As illustrated in Fig. 6, SEAM first employs a multi-branch Channel and Spatial Mixing Module (CSMM). The input feature map is fed into parallel CSMM branches with different patch sizes (e.g., patch=6, 7, 8) to capture multi-scale contextual information. Inside each CSMM, DWConv are used to efficiently learn spatial and channel correlations, and a residual connection is added to facilitate gradient flow. The outputs from these multi-scale branches are fused and passed through a global average pooling layer, followed by a two-layer fully connected network (depicted as the pink and orange nodes in Fig. 6) to learn complex inter-channel dependencies. Its key innovation is the use of an exponential activation function (Channel exp). This function maps the learned channel-attention weights to a wider range of [1, e], non-linearly amplifying the response of critical channels while suppressing noise. Finally, the resulting attention map is multiplied element-wise with the original input feature map, enhancing responses in target regions.

2.3. Experimental platform and evaluation metrics

2.3.1. Experimental platform and parameter settings

The experimental environment ran on Windows 10, and the hardware configuration included an Intel Core i7–10,700 CPU (2.90 GHz), an NVIDIA GeForce RTX 3090 GPU with 24 GB of VRAM, and 32 GB of RAM. The deep learning framework was PyTorch 1.10.1 with CUDA 11.3. The detailed parameters are shown in Table 1. For model optimization, we utilized the default composite loss function of YOLOv11n, which comprises Box_Loss for spatial localization, Cls_Loss for target classification, and Dfl_Loss for fine-grained boundary refinement. Specifically, although an early stopping mechanism with a patience of 50 was set to prevent overfitting, it was not triggered during our experiment, and the model was completely trained for the full 300 epochs.

2.3.2. Model evaluation metrics

To provide a comprehensive and objective assessment of model performance, especially considering future real-time deployment, we

Table 1
Hyperparameters used for model training.

Hyperparameters	Values
Epochs	300
Batch size	32
Image size/(width × height, pixels)	640 × 640
Optimizer	SGD
Patience	50
IoU Threshold	0.7
Beta	0.937
Initial learning rate	0.01
Weight decay coefficient	0.0005

used several standard metrics. Detection accuracy was evaluated using Precision (P) and mean Average Precision (mAP) at an Intersection over Union (IoU) threshold of 0.5 (mAP@0.5). Moreover, to theoretically justify the model's feasibility on resource-constrained edge devices, its computational complexity and compactness were measured using GFLOPs and model size (MB), alongside traditional inference speed (FPS).

3. Results

3.1. Performance evaluation of different deep learning models for litchi flower recognition

To provide a comprehensive and objective assessment of the proposed FDMA-Net model, we selected six representative deep-learning object detectors for comparison. Based on their detection paradigm and application characteristics, these models are grouped into three categories: high-accuracy (two-stage) models, mainstream single-stage models, and lightweight (edge-deployment) models. Through a multi-dimensional comparison against these models, we systematically validate the advancement and practical utility of FDMA-Net in litchi-flower detection. YOLOv11n is adopted as the foundational model for this research, principally due to its distinction within the YOLO series as an architecture that masterfully balances detection accuracy, inference speed, and a lightweight design. Its comprehensive performance has been proven exemplary across numerous public datasets, thus providing a high-standard and compelling benchmark against which the enhancements proposed herein can be effectively evaluated. All experiments were conducted on the same software and hardware platform and dataset, with a fixed input resolution of 640 × 640 pixels, 300 training epochs, and a batch size of 32. Detailed comparative results are reported in Table 2.

As shown in Table 2, the two-stage detector Fast Region-based Convolutional Neural Network (Fast R-CNN) lags behind single-stage methods in both precision and mAP, while also suffering from massive computational overhead. Although Single Shot MultiBox Detector (SSD) offers some gains, its overall accuracy remains limited. Furthermore, while RT-DETR is a modern architecture, its high computational cost restricts its practical application on edge devices. In contrast, widely used single-stage detectors—YOLOv8 and the YOLOv11n baseline—deliver strong performance, in particular, the baseline YOLOv11n attains an mAP@0.5 of 89.9%. While the lightweight YOLOv5s achieves the highest inference speed, its detection accuracy remains inferior to that of the proposed model.

By comparison, the proposed FDMA-Net achieves the best results among the evaluated methods in both precision and overall detection performance: its precision reaches 89.4%, a 5.2-point improvement over YOLOv11n; its mAP@0.5 reaches 94.1%, outperforming the baselines and indicating that the introduced SimAM attention mechanism and architectural refinements contribute to improved litchi-flower detection. Crucially, FDMA-Net requires the lowest computational cost among all models, operating at merely 5.2 GFLOPs with a minimal model size of 4.3 MB. This superior computational efficiency underscores its exceptional feasibility for deployment in resource-constrained edge environments. In terms of inference speed (FPS), FDMA-Net also maintains efficiency comparable to widely used lightweight models, preserving detection accuracy without compromising deployment efficiency.

3.2. Ablation study results

To validate the effectiveness of each improved module in FDMA-Net, we conducted ablation studies under identical training conditions and hyperparameter settings, configuring nine ablation variants; results are reported in Table 3.

The experiments used YOLOv11n as the baseline model: its mAP@0.5 on the test set was 89.9%, the model size was 5.3 MB, the computational complexity was 6.3 GFLOPs, and the inference speed was 232.5 FPS. The first step was to replace the baseline's backbone with HGNetV2; as shown in Table 3, this change reduced the model size to 4.5 MB, decreased the computational complexity to 5.7 GFLOPs, and increased FPS to 248.9. Although mAP@0.5 exhibited a slight fluctuation, introducing HGNetV2 established a lighter and more computationally efficient foundation, providing headroom for integrating subsequent modules.

Building on this, we individually evaluate the effectiveness of the SimAM attention mechanism, the C3k2_Star feature-fusion module, and the Detect_SEAM detection head. Notably, SimAM delivers the largest gain in detection accuracy, raising mAP@0.5 to 92.5% and indicating its efficacy in strengthening critical feature representations, while notably reducing the computational complexity to 5.2 GFLOPs. Similarly, C3k2_Star and Detect_SEAM increase mAP@0.5 to 91.7% and 91.3%, respectively, along with reductions in model size and improvements in inference speed, all while keeping the computational complexity strictly below the baseline level.

Further experimental results demonstrate a positive synergistic effect from the combination of the proposed modules. The fully integrated FDMA-Net model achieves an mAP@0.5 of 94.1%, representing a 4.2% improvement over the baseline. Simultaneously, its model size is merely 4.3 MB (a reduction of 18.9% from the baseline), and its computational complexity drops significantly to 5.2 GFLOPs (a 17.5% reduction compared to the baseline's 6.3 GFLOPs). Its inference speed also

Table 3
Ablation study results.

Models	Precision	mAP@0.5	Modelsize	GFLOPs	FPS
YOLOv11n	84.2	89.9	5.3	6.3	232.5
HGNetV2	84.4	89.6	4.5	5.7	248.9
HGNetV2-C3k2_Star	87.7	91.7	4.4	5.8	256.1
HGNetV2-SimAM	87.9	92.5	6.3	5.2	**265.5**
HGNetV2-Detect_SEAM	86.1	91.3	4.3	5.4	256.1
HGNetV2-Detect_SEAM-C3k2_Star	88.1	92.7	4.3	5.5	254.7
HGNetV2-Detect_SEAM-SimAM	86.4	91.9	6.2	5.6	252.7
HGNetV2-C3k2_Star-SimAM	88.1	92.7	6.3	**5.1**	245.7
FDMA-Net	**89.4**	**94.1**	**4.3**	5.2	256.5

Table 2
Performance comparison of different detection models.

Category	Model	Precision	mAP@0.5	Modelsize	GFLOPs	FPS
Highaccuracy	Fast R-CNN	67.3	72.6	108.2	116.3	14.2
Mainstream	SSD	79.6	83.4	76.7	14.2	216.8
single-stage	YOLOv8	83.2	88.6	5.4	6.8	243.7
detectors	RT-DETR	84.6	78.8	66.2	103.5	47.8
	YOLOv11n(baseline)	84.2	89.9	5.3	6.3	232.5
Lightweight	YOLOv5s	85.1	89.8	4.5	5.8	**264.6**
ProposedModel	FDMA-Net	**89.4**	**94.1**	**4.3**	**5.2**	256.5

surpasses that of the baseline model. These ablation study results strongly prove that the improvements proposed in this study constitute a systematic and holistic optimization. Consequently, the final model exhibits a significant comprehensive advantage across the four key dimensions of detection accuracy, model size, computational complexity, and inference speed.

3.3. Comparison of attention mechanisms

To assess the effectiveness and state-of-the-art potential of the SimAM attention mechanism for litchi-flower detection, we designed a comparative study. Using the same model architecture and replacing only the attention module, we compared SimAM against three widely used attention mechanisms: Efficient Multi-Scale Attention (EMA) [31], Large Separable Kernel Attention (LSKA) [32], and Efficient Channel Attention (ECA) [33].

All experiments were conducted on the same software–hardware platform, dataset, and hyperparameter settings to ensure a fair comparison, with the results reported in Table 4.

As indicated by the results in Table 4, the SimAM attention mechanism delivers the strongest overall performance: it attains an mAP@0.5 of 94.1% and a precision of 89.4%, outperforming the other baselines on key accuracy metrics; this suggests that its neuron-importance-based feature enhancement is effective for recognizing fine-grained targets such as litchi flowers. By comparison, EMA and ECA also perform well, with mAP@0.5 of 93.4% and 93.2%, respectively, supporting the utility of multi-scale features and channel-wise interaction for this task. Although LSKA's large-kernel separable-convolution design affords a marginal edge in inference speed, it yields the weakest mAP among the compared mechanisms. In terms of compactness, the other three mechanisms all maintain a very small model size of 4.3 MB, whereas LSKA is slightly larger. Overall, SimAM attains the highest detection accuracy while keeping model size and inference throughput at competitive levels, indicating no efficiency penalty for accuracy. This balanced performance across accuracy, speed, and size makes SimAM a compelling choice for challenging small-object recognition tasks, such as litchi-flower detection.

3.4. Experiments on model robustness and resistance to interference

During real-world UAV operations, the acquired remote-sensing imagery often exhibits varying degrees of motion blur due to platform motion, gimbal shake, and ambient wind, posing a serious challenge to detection performance [34]. To assess the robustness of the proposed FDMA-Net in such scenarios, we designed a motion-blur robustness experiment. Using motion-blur kernels, we processed the original test set to generate three subsets with mild, moderate, and severe blur, and evaluated the performance of FDMA-Net and the baseline (YOLOv11n) on each subset. Comparative mAP@0.5 results across blur levels are reported in Table 5, and the degradation trends are illustrated in Fig. 7.

From Fig. 7, two key observations can be made. First, across all test conditions, FDMA-Net's mAP@0.5 is consistently higher than that of the baseline YOLOv11n; as blur severity increases, FDMA-Net maintains a clear performance advantage. Second, FDMA-Net exhibits stronger robustness, with a more gradual degradation curve. When the motion-blur level reaches "severe," the baseline model's performance drops sharply to 69.2%, a 20.6-point decrease relative to the no-blur condition. Under such severe motion blur, FDMA-Net's mAP@0.5 remains at 78.9%, with performance decreasing by only 14.8 percentage points—substantially less degradation than the baseline. This robustness primarily stems from FDMA-Net's advanced feature-extraction and fusion architecture: motion blur predominantly corrupts high-frequency edge details [35], whereas the strong multi-scale information aggregation in the HGNetV2 backbone and the C3k2_Star module allows the model to rely less on such fragile fine details and instead leverage relatively stable mid- to low-frequency semantic cues, such as shape contours and color distributions. Meanwhile, when features are degraded, the SimAM attention mechanism guides the model to adaptively focus on the most informative, core regions of the target, thereby mitigating interference induced by edge blurring.

Table 4
Performance comparison of different attention mechanisms.

Attn. mechanism	Precision	mAP@0.5	Modelsize	GFLOPs	FPS
SimAM	**89.4**	**94.1**	**4.3**	5.2	256.5
ECA	88.0	93.2	4.3	5.4	234.6
LSKA	87.9	93.0	4.4	**5.4**	**257.7**
EMA	88.7	93.4	4.3	5.2	254.6

Table 5
Impact of different blur levels on model detection performance.

Blur level	Blur kernel size	Baseline mAP	FDMA-Net mAP
0 (no blur)	0	89.8	93.7
1 (slight)	3	88.9	93.2
2 (moderate)	7–11	87.1	90.7
3 (severe)	17–23	69.2	78.9

3.5. Visualization and analysis of detection results

To provide an intuitive assessment of FDMA-Net's performance in complex agricultural environments, we selected representative images spanning various scenarios—incorporating simulated lighting variations, high object density, and complex backgrounds—for visual comparative analysis. We performed inference on these images using the baseline YOLOv11n and the proposed FDMA-Net, and the detection results are shown in Fig. 8.

As shown in Fig. 8, FDMA-Net exhibits clear advantages across multiple complex scenarios. In scenes with high-density flower clusters and complex backgrounds, the baseline YOLOv11n shows limitations, with numerous missed detections—particularly for instances that are partially occluded or tightly clustered in the center. By contrast, leveraging the Detect_SEAM decoupled detection head's ability to exploit contextual cues, FDMA-Net more accurately separates contiguous targets and successfully detects occluded flowers. In addition, in scenes with complex backgrounds, YOLOv11n is prone to false positives on branches and leaves whose colors or textures resemble those of flowers, whereas, benefiting from the SimAM attention mechanism, FDMA-Net more precisely attends to flower-specific salient features, effectively suppresses background clutter, and thus exhibits stronger robustness. The model's advantage is also evident in its tolerance to lighting variations: its HGNetV2 backbone and C3k2_Star feature-fusion module extract more discriminative features from overexposed or underexposed imagery, thereby maintaining more stable and reliable detection than the baseline under extreme lighting conditions.

4. Discussion

Accurate and efficient detection of litchi flowers is essential for regulating floral load and improving fruit quality, providing critical support for the transition of commercial orchards from extensive to precision management. The experimental results demonstrate that the architectural innovations of FDMA-Net effectively address the primary challenges of detecting small, densely clustered, and heavily occluded targets within litchi flower imagery. Specifically, the integration of the HGNetV2 backbone and the C3k2_Star module establishes a robust foundation for multi-scale feature extraction without introducing prohibitive computational overhead. Furthermore, the SimAM attention mechanism acts as an effective spatial filter to suppress complex background clutter, while the decoupled Detect_SEAM head resolves the inherent feature conflicts between bounding box regression and

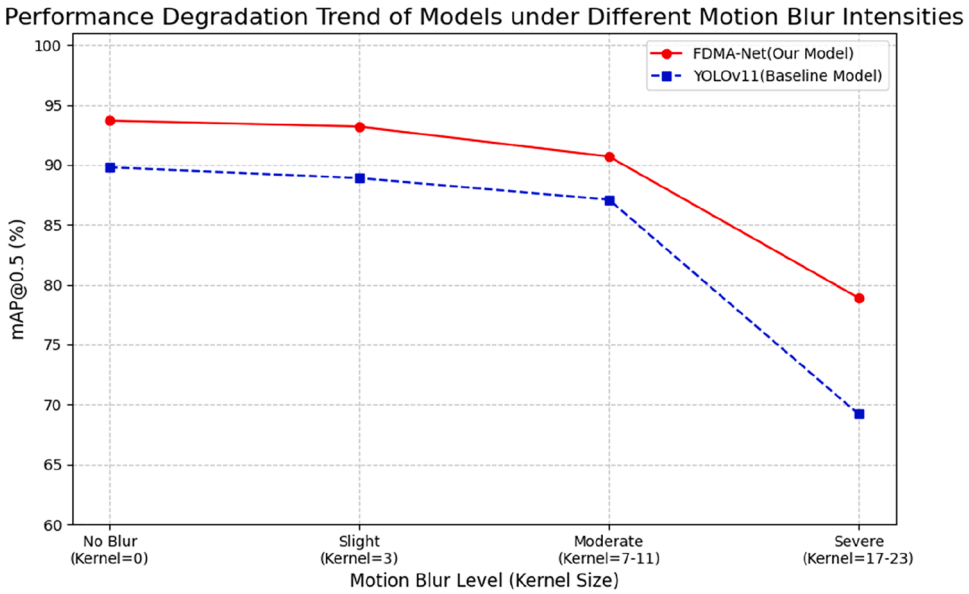

Fig. 7. Performance comparison of the model under different motion blur levels.

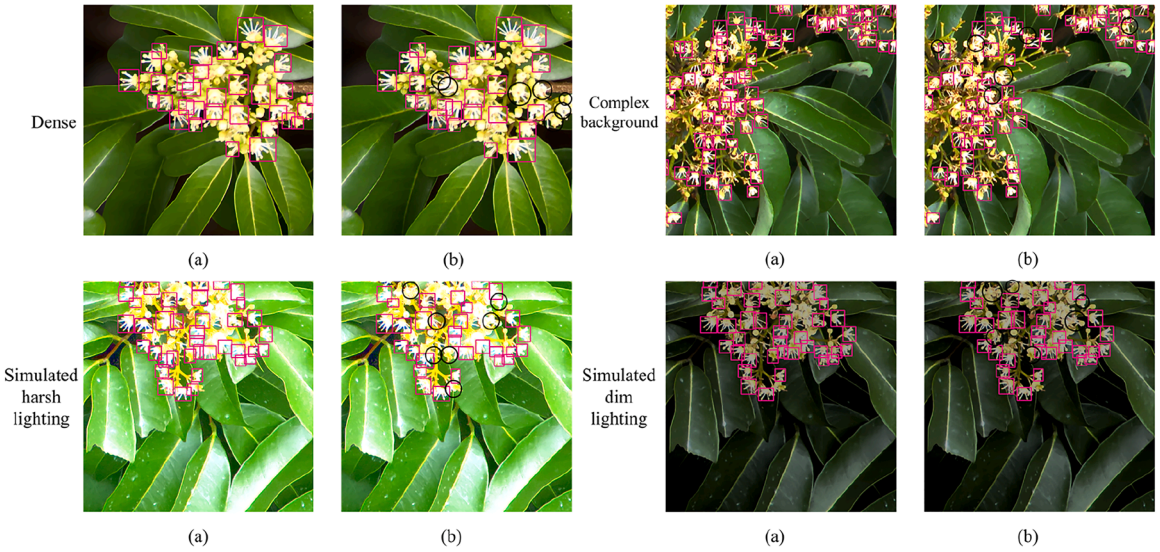

Fig. 8. Visual comparison between FDMA-Net and YOLOv11n. (a) and (b) show the detection results of FDMA-Net and YOLOv11n, respectively. The bounding boxes represent the detected litchi flowers, and the black circles highlight the missed detections.

classification tasks in highly dense scenes.

Despite the promising performance achieved on the litchi flower dataset, this study has several limitations. First, regarding the dataset, the practical constraints of agricultural production resulted in a limited sample size of images captured under natural extreme weather conditions. Although digital data augmentation was employed to simulate basic lighting variations, these artificial transformations cannot fully replicate the complex optical effects induced by genuine extreme weather. Second, the cross-scene generalization capability of the model remains limited. The current evaluation is restricted to specific litchi orchards, and its adaptability to entirely different dense flowers or fruits across diverse agricultural scenes has yet to be verified. Finally, owing to current resource constraints, the model has not been practically deployed or tested on physical UAV-mounted edge computing platforms. Its real-time feasibility for resource-constrained environments is currently substantiated solely by theoretical metrics, such as GFLOPs and model size.

To address these limitations, future work will focus on expanding the dataset to encompass a broader range of genuine extreme weather conditions, as well as multi-variety and multi-scene agricultural data, to rigorously validate and enhance the model's generalizability. Furthermore, we plan to actually deploy the model on UAV-mounted edge computing hardware and conduct closed-loop field tests to verify its operational reliability in real-world scenarios.

5. Conclusions

To address the challenges posed by litchi-flower targets—namely small object size, dense occlusions, and complex backgrounds—in UAV remote-sensing imagery, we propose a lightweight, high-accuracy YOLOv11n-based detector—FDMA-Net:

1. This study develops the FDMA-Net model by introducing HGNetV2 as the backbone to enable lightweight, efficient feature

extraction; integrates the parameter-free SimAM attention mechanism, markedly enhancing the model's discrimination of fine-grained litchi-flower features; employs the C3k2_Star module to strengthen multi-scale feature fusion and adapts a decoupled Detect_SEAM head to effectively mitigate localization–classification conflicts in dense scenes.

2. Experimental results show that FDMA-Net achieves strong performance. On our in-house UAV remote-sensing dataset of litchi flowers, it attains an mAP@0.5 of 94.1%-a 4.2-percentage-point gain over the YOLOv11n baseline—and it outperforms mainstream detectors including Faster R-CNN, SSD, and YOLOv8. The model size is 4.3 MB (18.9% smaller than the baseline) and inference throughput remains high, yielding a favorable overall trade-off across detection accuracy, model size, and inference efficiency.

3. Through the motion-blur robustness experiment and visual analyses across multiple complex scenarios, we show that FDMA-Net is robust. Even when image quality degrades or background clutter is severe, its performance deteriorates substantially less than that of the baseline model. Its balanced trade-off across accuracy, throughput, and compactness indicates strong potential for deployment on edge-computing platforms such as UAVs.

In summary, the FDMA-Net model proposed in this study offers an effective and robust solution specifically for litchi flower detection in UAV remote sensing imagery, effectively addressing the challenges of small object sizes and dense occlusions. The model provides essential technical support for smart agriculture applications such as automated flower thinning and yield prediction, demonstrating theoretical significance and broad application potential.

Ethical statement

This study does not involve any human participants or animal subjects. Therefore, ethical approval and informed consent were not required.

CRediT authorship contribution statement

Xiaoling Deng: Conceptualization, Methodology, Investigation, Writing – original draft, Funding acquisition. **Yiwei Wang:** Conceptualization, Methodology, Investigation, Writing – original draft, Visualization. **Heguang Sun:** Methodology, Investigation. **Zifan Shen:** Data curation, Software. **Wenqian Tang:** Data curation, Validation. **Fernando Auat Cheein:** Methodology, Investigation, Data curation. **Haibo Chen:** Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

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

Acknowledgements

This research was funded by the Guangdong Provincial Key R&D Program (Grant No. 2023B0202090001 and Grant No. 2025B0202090001), the National Natural Science Foundation of China General Program (Grant No. 32371984), and the National Key Research and Development Plan Project (Grant No. 2023YFD2000200). Support was also provided by the Guangdong Higher Education Key Area (Artificial Intelligence) Special Project (Grant No. 2019KZDZX1012), the '111 Center' (Grant No. D18019), and the China Agriculture Research System (Grant No. CARS-15–22).

Data availability

Data will be made available on request.

References

[1] W. Qi, H. Chen, T. Luo, F. Song, Development status, trend and suggestion of litchi industry in mainland China, Guangdong Agric. Sci. 46 (2019) 132–139.
[2] J. Lin, J. Li, Z. Yang, H. Lu, Y. Ding, H. Cui, Estimating litchi flower number using a multicolumn convolutional neural network based on a density map, Precis. Agric. 23 (2022) 1226–1247, https://doi.org/10.1007/s11119-022-09882-7.
[3] X.M. Huang, H.B. Chen, Studies on shoot, flower and fruit development in litchi and strategies for improved litchi production, in: IV International Symposium on Lychee, Longan and Other Sapindaceae Fruits 1029, 2012, pp. 127–136.
[4] Z. Li, P. Yuan, Y. Qiu, J. Li, C. Fan, The relationship between winter irrigation and spring flowering of guiwei litchi trees, Chin. J. Trop. Crops. 33 (2012) 402–407.
[5] R. Banjare, N. Nidhi, A. Sood, Physiological aspects of flowering, fruit setting, fruit development and fruit drop, regulation and their manipulation: a review, Int. J. Env. Clim. Chang. 13 (2023) 205–224.
[6] D. Hehnen, I. Hanrahan, K. Lewis, J. McFerson, M. Blanke, Mechanical flower thinning improves fruit quality of apples and promotes consistent bearing, Sci. Hortic. 134 (2012) 241–244.
[7] X. Lei, Q. Yuan, T. Xyu, Y. Qi, J. Zeng, K. Huang, Y. Sun, A. Herbst, X. Lyu, Technologies and equipment of mechanized blossom thinning in orchards: a review, Agronomy 13 (2023) 2753, https://doi.org/10.3390/agronomy13112753.
[8] Ren, P.P., Ma, X.S., Wei, B.W., Wang, H.C., 2021. Effects of foliar rare earth fertilizer on photosynthesis, flowering and fruiting in Litchi chinensis.
[9] G. Farjon, O. Krikeb, A.B. Hillel, V. Alchanatis, Detection and counting of flowers on apple trees for better chemical thinning decisions, Precis. Agric. 21 (2020) 503–521, https://doi.org/10.1007/s11119-019-09679-1.
[10] B. Martin-Gorriz, A. Torregrosa, J.G. Brunton, Feasibility of peach bloom thinning with hand-held mechanical devices, Sci. Hortic. 129 (2011) 91–97.
[11] F. Palacios, G. Bueno, J. Salido, M.P. Diago, I. Hernández, J. Tardaguila, Automated grapevine flower detection and quantification method based on computer vision and deep learning from on-the-go imaging using a mobile sensing platform under field conditions, Comput. Electron. Agric. 178 (2020) 105796.
[12] N. Aierken, B. Yang, Y. Li, P. Jiang, G. Pan, S. Li, A review of unmanned aerial vehicle based remote sensing and machine learning for cotton crop growth monitoring, Comput. Electron. Agric. 227 (2024) 109601.
[13] A. Barreto, P. Lottes, F.R.I. Yamati, S. Baumgarten, N.A. Wolf, C. Stachniss, A. K. Mahlein, S. Paulus, Automatic UAV-based counting of seedlings in sugar-beet field and extension to maize and strawberry, Comput. Electron. Agric. 191 (2021) 106493.
[14] J. Zhao, J. Yan, T. Xue, S. Wang, X. Qiu, X. Yao, Y. Tian, Y. Zhu, W. Cao, X. Zhang, A deep learning method for oriented and small wheat spike detection (OSWSDet) in UAV images, Comput. Electron. Agric. 198 (2022) 107087.
[15] R. Reedha, E. Dericquebourg, R. Canals, A. Hafiane, Transformer neural network for weed and crop classification of high resolution UAV images, Remote Sens. 14 (2022) 592.
[16] Y. Shang, X. Xu, Y. Jiao, Z. Wang, Z. Hua, H. Song, Using lightweight deep learning algorithm for real-time detection of apple flowers in natural environments, Comput. Electron. Agric. 207 (2023) 107765.
[17] X. Mu, L. He, P. Heinemann, J. Schupp, M. Karkee, Mask R-CNN based apple flower detection and king flower identification for precision pollination, Smart Agric. Technol. 4 (2023) 100151, https://doi.org/10.1016/j.atech.2022.100151.
[18] J.S. Estrada, J.P. Vasconez, L. Fu, F.A. Cheein, Deep Learning based flower detection and counting in highly populated images: a peach grove case study, J. Agric. Food Res. 15 (2024) 100930.
[19] X. Wang, A. Wang, J. Yi, Y. Song, A. Chehri, Small object detection based on deep learning for remote sensing: a comprehensive review, Remote Sens. 15 (2023) 3265.
[20] J. Terven, D.M. Córdova-Esparza, J.A. Romero-González, A comprehensive review of yolo architectures in computer vision: from yolov1 to yolov8 and yolo-nas, Mach. Learn. Knowl. Extr. 5 (2023) 1680–1716.
[21] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, J. Chen, Detrs beat yolos on real-time object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024 16965–16974.
[22] R. Zhang, F. Zhu, J. Liu, G. Liu, Depth-wise separable convolutions and multi-level pooling for an efficient spatial CNN-based steganalysis, IEEE Trans. Inf. Forens. Secur. 15 (2019) 1138–1150.
[23] L. Liang, Y. Zhang, S. Zhang, J. Li, A. Plaza, X. Kang, Fast hyperspectral image classification combining transformers and SimAM-based CNNs, IEEe Trans. Geosci. Remote Sens. 61 (2023) 1–19.
[24] L. Yang, R.Y. Zhang, L. Li, X. Xie, SimAM: a simple, parameter-free attention module for convolutional neural networks, PMLR (2021).
[25] X. Ma, X. Dai, Y. Bai, Y. Wang, Y. Fu, Rewrite the Stars, CoRR, 2024.
[26] L. Jin, W. Ding, S. Han, J. Wang, A real-time edge inference method for insulator contamination detection with YOLOv11-ssL, IEEe Trans. Instrum. Meas. (2025).
[27] J. Xu, Y. Hu, Z. Gou, Y. Lu, D. Cui, Automated non-invasive method for measuring chicken comb and wattle using multi-camera system and CW-measure-pose, Expert. Syst. Appl. (2025) 129420.
[28] Z. Yu, H. Huang, W. Chen, Y. Su, Y. Liu, X. Wang, Yolo-facev2: a scale and occlusion aware face detector, Pattern. Recognit. 155 (2024) 110714.

[29] Z. Qiu, X. Huang, S. Li, J. Wang, Stellar-YOLO: a graphite ore grade detection method based on improved YOLO11, Symmetry 17 (2025) 966.
[30] J. Xie, X. Xie, W. Xie, Q. Xie, An improved YOLOv8-based method for detecting pests and diseases on cucumber leaves in natural backgrounds, Sensors 25 (2025) 1551.
[31] D. Ouyang, S. He, G. Zhang, M. Luo, H. Guo, J. Zhan, Z. Huang, Efficient multi-scale attention module with cross-spatial learning, in: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2023, pp. 1–5.
[32] K.W. Lau, L.M. Po, Y.A.U. Rehman, Large separable kernel attention: rethinking the large kernel attention design in cnn, Expert. Syst. Appl. 236 (2024) 121352.
[33] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, ECA-net: efficient channel attention for deep convolutional neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020 11534–11542.
[34] J.H. Lee, G.H. Gwon, I.H. Kim, H.J. Jung, A motion deblurring network for enhancing UAV image quality in bridge inspection, Drones 7 (2023) 657.
[35] X. Luo, N.Z. Salamon, E. Eisemann, Controllable motion-blur effects in still images, IEEe Trans. Vis. Comput. Graph. 26 (2020) 2362–2372, https://doi.org/10.1109/TVCG.2018.2889485.