



OPEN ACCESS

EDITED BY

Bimlesh Kumar,
Indian Institute of Technology Guwahati, India

REVIEWED BY

Yang Lu,
Heilongjiang Bayi Agricultural University, China
Meena Pandey,
University of California, Davis, United States
Xiwang Xie,
Dalian Maritime University, China

*CORRESPONDENCE

Maocheng Zhao
✉ mczhao@njfu.edu.cn

RECEIVED 04 August 2025

REVISED 27 October 2025

ACCEPTED 10 November 2025

PUBLISHED 27 November 2025

CITATION

Wang J, Gu S and Zhao M (2025) PLFYNet-based edge-deployable detection system for Ginkgo biloba leaf diseases.
Front. Plant Sci. 16:1679455.
doi: 10.3389/fpls.2025.1679455

COPYRIGHT

© 2025 Wang, Gu and Zhao. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

PLFYNet-based edge-deployable detection system for Ginkgo biloba leaf diseases

Jun Wang¹, Siyuan Gu¹ and Maocheng Zhao^{2*}

¹College of Information Science and Technology & Artificial Intelligence, Nanjing Forestry University, Nanjing, China, ²College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing, China

Introduction: Target detection is a pivotal technology for precise monitoring of leaf-used Ginkgo biloba diseases in precision agriculture. However, complex plantation environments impose significant constraints on existing detection systems, manifesting as degraded detection accuracy, suboptimal efficiency, and prohibitive computational overhead for edge deployment. This study aims to develop a lightweight deep learning model tailored for real-time disease detection on resource-constrained embedded devices.

Methods: First, a comprehensive multi-class dataset was constructed, containing 7,158 augmented images covering three disease categories: chlorosis, insect pest, and physical damage. Five lightweight architectures were systematically evaluated, and an optimized reconstructed backbone network was adopted. To maintain architectural efficiency, attention mechanisms, an improved detection head, and efficient convolution techniques were integrated, along with a custom feature fusion module designed to address small target feature loss—forming the base model LCNet-FusionYOLO. Subsequently, Layer-Adaptive Magnitude-based Pruning (LAMP) was applied to reduce model scale while enhancing performance, yielding the final PLFYNet model.

Results: The PLFYNet model achieves 94.5% mAP@0.5 with only 3.0M parameters, surpassing the baseline YOLOv7-tiny by 4.8% while using merely half the parameters. Deployment on the Jetson Orin Nano embedded platform demonstrates real-time inference at 50.5 FPS, validating its practical applicability in field scenarios.

Discussion: This work establishes a paradigm for developing high-precision, computationally efficient disease detection systems. By balancing accuracy and resource efficiency, PLFYNet provides a practical edge-based monitoring solution for sustainable Ginkgo biloba cultivation, addressing the key deployment challenges of existing detection systems in complex agricultural environments.

KEYWORDS

leaf-used Ginkgo biloba, lightweight, disease detection, attention mechanism, LAMP

1 Introduction

Ginkgo biloba is a multifunctional agricultural commodity in China, serving ornamental, culinary, and pharmaceutical purposes—its medicinal value lies primarily in foliage containing flavonoids and terpenoid lactones. These bioactive compounds benefit cardiovascular health and cerebrovascular circulation, promote brain cell metabolism (Wang et al., 2022; Zhang et al., 2024), and exhibit antioxidant, anti-inflammatory, and other physiological benefits (Zhang et al., 2023a). This pharmaceutical leaf value makes foliar diseases (e.g., anthracnose caused by *Colletotrichum fructicola*) particularly damaging, as infections reduce both yield and medicinal compound content.

As the global leader in Ginkgo biloba cultivation, China's extensive subtropical plantations have positioned it as a crucial research center for advancing leaf disease management strategies (Sun et al., 2017). The intensification of global climate change and ecological disruptions has amplified disease pressures on leaf-harvested Ginkgo biloba (Zhang et al., 2023b). Insect infestations frequently induce leaf chlorosis and premature defoliation, significantly impeding plant development. Under conducive environmental conditions—particularly elevated temperature and humidity—pests such as *Scirtothrips* can proliferate rapidly throughout plantations, resulting in substantial economic losses (Mirab-balou et al., 2012). Furthermore, various stressors including natural disasters (hail, frost, excessive precipitation) and anthropogenic factors (mechanical injury, pesticide misapplication) continue to threaten sustainable Ginkgo biloba production.

Recent advances in deep learning, particularly the integration of convolutional neural networks (CNNs) and target detection algorithms, have revolutionized automated plant disease detection by enhancing accuracy and efficiency. They enable early identification of infections. This reduces manual monitoring efforts. It also supports timely pest control measures (Gkountakos et al., 2024; Jiang et al., 2020). However, Ginkgo biloba's unique characteristics—medicinally valuable foliage with fine-textured lesions and dense plantation canopies—create unmet needs that generic crop detection models fail to address.

Girshick et al.'s R-CNN established the foundational CNN-based detection framework (Girshick et al., 2014), but Ren et al. later noted that even the optimized Faster R-CNN retains excessive

computational complexity (Ren et al., 2017)—a critical limitation for edge deployment in remote Ginkgo plantations where high-performance servers are unavailable. The YOLO series addressed real-time demands, yet Omer et al.'s lightweight YOLOv5l, optimized for cucumber diseases, demonstrated that crop-specific customization is essential (Omer et al., 2024); their model achieved high accuracy for cucumber lesions but lacked adaptability to Ginkgo's tiny, irregular spots (e.g., early anthracnose lesions under 2mm). Even YOLOv7-tiny, a common edge-focused variant, struggles with Ginkgo-specific challenges: its unidirectional PANet fusion leads to high false-negative rates in dense canopies, and conventional CIoU loss fails to precisely localize morphologically diverse lesion. Compounding these issues, Gkountakos et al.'s review highlighted that existing models lose up to 15% accuracy in variable field lighting—a prevalent condition in Ginkgo plantations.

Although these vision-based algorithms enable rapid identification and localization of both emerging and established infections, facilitating timely targeted interventions such as precision fungicide applications, ultimately mitigating economic losses (Wang et al., 2025a; Zhu et al., 2023), several limitations still exist. To address these limitations, we propose PLFYNet, an enhanced lightweight architecture based on YOLOv7-tiny. This model targets the identified gaps through four key innovations. First, it integrates Bottleneck Transformer to capture long-range lesion dependencies (Xie et al., 2025a). This resolves the canopy occlusion issue highlighted. Second, it uses BiECaFusion for bidirectional cross-scale feature interaction. This fixes the small lesion loss problem in crop-specific model. Third, it adopts Shape-IoU loss for accurate bounding box optimization. This overcomes the localization inaccuracies of YOLOv7-tiny noted by Zhao et al. (2023). Fourth, it incorporates DyHead (Duan et al., 2024) to enable robust target representation under variable lighting. We selected the reconstructed PP-LCNet as the backbone after evaluating five lightweight architectures: MobileNetV3, GhostNet, ShuffleNetV2, PP-PicoDet, and PP-LCNet. This backbone significantly reduces computational complexity while maintaining detection performance. Additionally, the model integrates the Efficient Channel Attention (ECA) mechanism. This enhances feature discrimination between similar lesions, such as insect-induced chlorosis and abiotic yellowing. Comprehensive evaluations confirm that LCNET-FusionYOLO is superior to both the original and conventional models in terms of mAP@0.5, false-positive, and false-negative metrics. Further optimization through LAMP pruning, which involves adaptive layer-wise redundancy elimination, results in the final PLFYNet architecture. This architecture has reduced parameters (3.0M), improved real-time capability (50.5 FPS on Jetson Orin Nano), and accelerated inference. It directly addresses the deployment barrier identified in Ren et al. (2017) Faster R-CNN analysis and enables high-precision field detection in dense canopy environments.

The research methodology encompasses three primary phases:

1. Dataset Development: We acquired 3,600 high-resolution Ginkgo leaf images and applied comprehensive

Abbreviations: PP-LCNet, PaddlePaddle Lightweight Convolutional Neural Network; MobileNetV3, Mobile Neural Network Version 3; GhostNet, Ghost Convolutional Neural Network; ShuffleNetV2, ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design; PP-PicoDet, PaddlePaddle Pico Object Detector; Dyhead, Dynamic Head; BiECaFusion, Bidirectional ECA-enhanced Feature Fusion; SE, Squeeze and Excitation; CBAM, Convolutional Block Attention Module; ECA, Efficient Channel Attention; Shape-IoU, Shape-Intersection over Union; C-IoU, Complete IoU; CBL, Convolution-Batch Normalization-Leaky ReLU; MP, Max Pooling; ELAN, Enhanced Layer Aggregation Network; SPPCPC, Spatial Pyramid Pooling with Convolutional Pyramid Convolution; LAMP, Layer-Adaptive Magnitude-based Pruning (LAMP); mAP@0.5, mean Average Precision at IoU=0.5

augmentation strategies to expand the dataset, effectively mitigating overfitting while enhancing model robustness. Precise disease annotation was performed using Labelme software.

2. Architectural Innovation: We developed an optimized detection framework based on YOLOv7-tiny, reconstructing the backbone with PP-LCNet's DepthSepConv modules to minimize computational overhead. Systematic evaluation of five lightweight architectures (MobileNetV3, GhostNet, ShuffleNetV2, PP-PicoDet, and PP-LCNet) identified the optimal configuration. The selected Light-YOLO variant underwent comparative analysis with SE, CBAM, and ECA attention mechanisms. Performance enhancement was achieved through four integrated components: Bottleneck Transformer for global context modeling, BiECAFusion for multi-scale feature fusion, Shape-IoU for bounding box refinement, and DyHead for target representation. LAMP pruning provided final model compression. Evaluation metrics encompassed parameter count, precision, recall, mAP@0.5, and memory footprint.
3. Edge Deployment: The optimized model was successfully deployed on the Jetson Orin Nano embedded platform, achieving 50.5 FPS real-time detection while maintaining

accuracy. Comparative analysis validated superior inference speed and precision relative to baseline models.

2 Materials and methods

2.1 Preparation of the dataset

2.1.1 Source of experimental dataset

Data collection was conducted at two strategic locations: the primary site at Changrong Agricultural Development Co., Ltd. in Pizhou, Jiangsu Province (117°52'E, 34°37'N)—China's premier "Ginkgo Hometown" hosting the nation's largest commercial plantations—and supplementary sampling from Nanjing Forestry University. The principal research facility features 12 cultivation zones covering 117.07 hectares of leaf-oriented Ginkgo production, with spatial distribution presented in Figure 1.

2.1.2 Design of datasets

In this work, classifications were developed based on field research. We tracked the one-year growth cycle of Pizhou City's ginkgo plantations, and the dataset covers representative disease symptoms across annual growth cycles, with main diseases

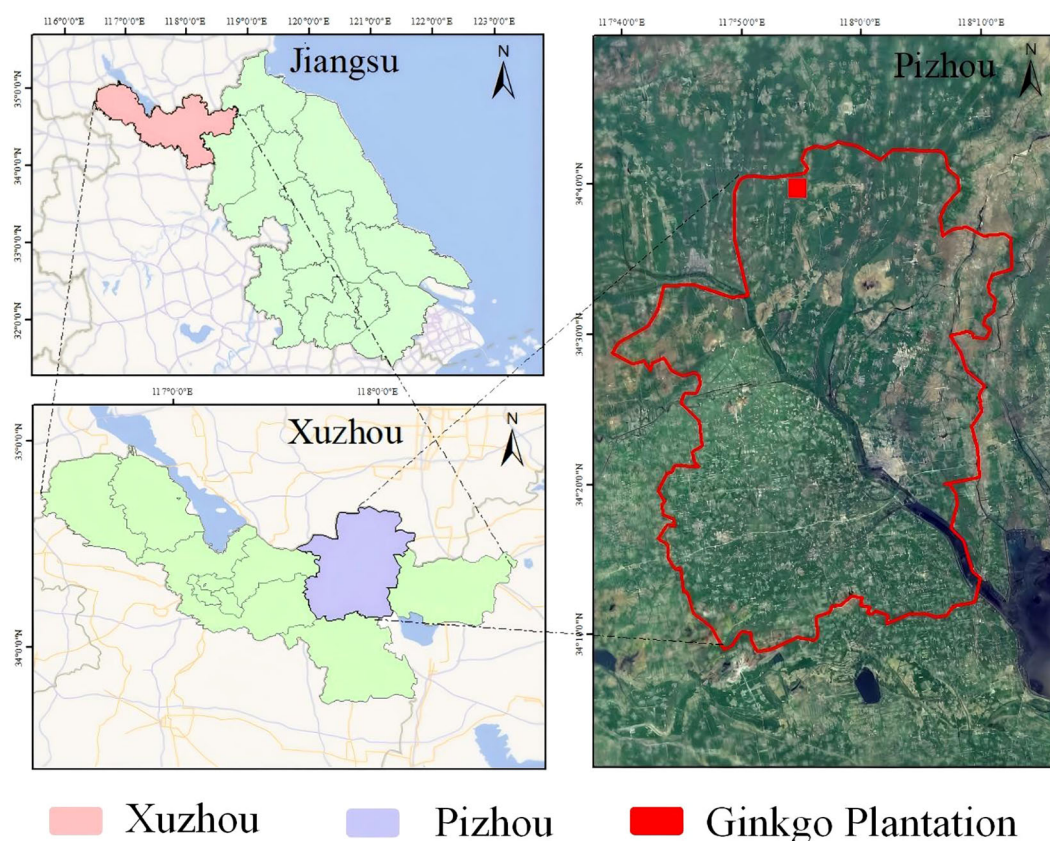


FIGURE 1
Main dataset collection sites.

distinguished by professional plantation disease managers. June–September is pest outbreak season; high temperatures bring tender shoots, and pests like *Scirtothrips* (short cycles, strong drug resistance) outbreak massively. They gnaw leaves rapidly, cutting plantation yields, and these pests are hard to detect (hide in leaves, have protective coloration) while gnawed leaves show obvious symptoms. Thus, such cases are classified as “insect_pest” by leaf characteristics. Physical damage (hailstorms, rain/snow, mechanical operations) has distinct features, so it is categorized as “physical_damage”. Chlorosis, a biotic stress disease, stems from fungi (*Alternaria* spp., etc.), yellowing is typical of such fungal infections on ginkgo leaves, and it is fully classified as “chlorosis” based on field research. A simplified labeling scheme was adopted: 0–chlorosis (symptomatic manifestation), 1–insect pest, 2–physical damage.

To strengthen model generalization capabilities and prevent overfitting, extensive data augmentation protocols were applied. Transformations included photometric variations (brightness, hue), contrast manipulations, and spatial rescaling, yielding an enriched dataset of 7,158 images. Dataset partitioning followed an 8:1:1 distribution for training, validation, and testing cohorts, respectively, ensuring methodological rigor in model evaluation. we have conducted generalization verification experiments using the public dataset PlantDoc. PlantDoc is an open-source public dataset for plant disease detection, containing 2,598 manually annotated images covering 13 plant species and 17 diseases, with an 85% training set and 15% validation set split (Singh et al., 2020).

LabelMe software facilitated precise annotation of the augmented dataset, generating 9,582 labeled instances

categorized as: chlorosis ($n=1,779$), insect pest ($n=5,260$), and physical damage ($n=2,543$). Visual representation of these categories appears in Figure 2, arranged hierarchically with insect-induced damage, physical trauma, and chlorosis symptoms displayed in successive rows. Class balancing was achieved through targeted augmentation of underrepresented categories, particularly enhancing insect pest samples to ensure equitable distribution.

Figure 3 a comprehensive analysis of dataset properties. (a) Class distribution of disease labels: The class distribution histogram displays three categories—chlorosis (red), insect pest (pink), and physical damage (orange)—with insect pest exhibiting the highest frequency due to morphological complexity, necessitating extensive training samples. Loss function weighting compensates for class imbalance. (b) Bounding box size distribution via center point aggregation: Concentrically arranged rectangles illustrate bounding box size distributions centered at coordinate origins. The radial expansion pattern demonstrates damage scale progression from minor peripheral lesions to substantial central deterioration, with intermediate sizes showing highest prevalence. (c) Spatial distribution of label centroid coordinates: Scatter plot analysis of normalized centroid coordinates (range: 0–1) reveals uniform spatial distribution across leaf surfaces, extending from periphery (x, y approaching 0 or 1) to center ($x, y \approx 0.5$), confirming comprehensive spatial representation without clustering artifacts. (d) Width-height distribution of bounding boxes for disease regions: Width-height correlation analysis exhibits strong diagonal alignment, indicating predominantly isometric damage patterns (characteristic of puncture wounds), while off-diagonal dispersal represents anisotropic lesions (typical of linear abrasions). The



FIGURE 2
Enhanced image of the dataset.

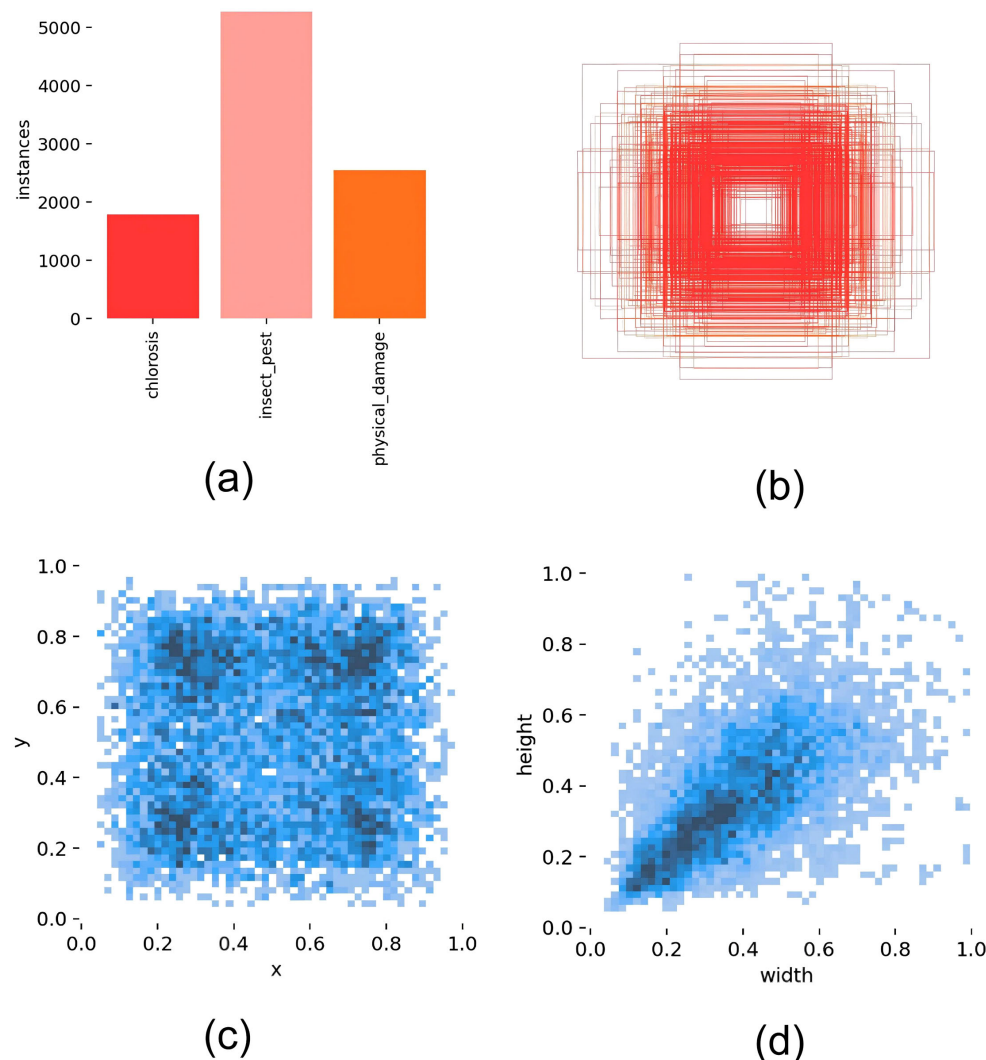


FIGURE 3

Dataset characteristics analysis: (a) Class distribution of disease labels; (b) Bounding box size distribution via center point aggregation; (c) Spatial distribution of label centroid coordinates; (d) Width-height distribution of bounding boxes for disease regions.

pronounced diagonal concentration confirms regular geometric damage patterns in *Ginkgo biloba* pathology.

2.2 YOLOv7 object detection model

The YOLOv7-tiny architecture offers an optimal balance between model compactness and detection speed, rendering it particularly well-suited for embedded applications on Jetson Orin Nano hardware, enabling real-time pathological assessment of *Ginkgo biloba* foliage (Zhao et al., 2023; Wang et al., 2023).

The YOLOv7-tiny architecture is structured into four hierarchical stages: initial input processing for image preparation, a feature extraction backbone, an intermediate neck for multi-scale aggregation, and a final detection module. The backbone employs a combination of CBL units (convolutional blocks with batch normalization and LeakyReLU activation) and ELAN modules,

which leverage multi-path architectures for enhanced feature representation, complemented by MP layers for progressive spatial reduction. Feature fusion occurs through SPPCSPC modules (Xu et al., 2024), which integrate Spatial Pyramid Pooling with Cross-Stage Partial connections, effectively expanding receptive fields while maintaining parameter efficiency across multiple scales.

The SPPCSPC module employs a bifurcated processing strategy for input features. One pathway directly processes features using a CBL block, while the alternate pathway implements a more complex structure: after initial CBL transformation, it diverges into four parallel streams comprising an identity mapping and three max-pooling operations with distinct kernel dimensions. This dual-pathway design achieves optimal balance between computational efficiency and detection performance via multi-level feature aggregation, as depicted in Figure 4, showing the complete YOLOv7 architecture.

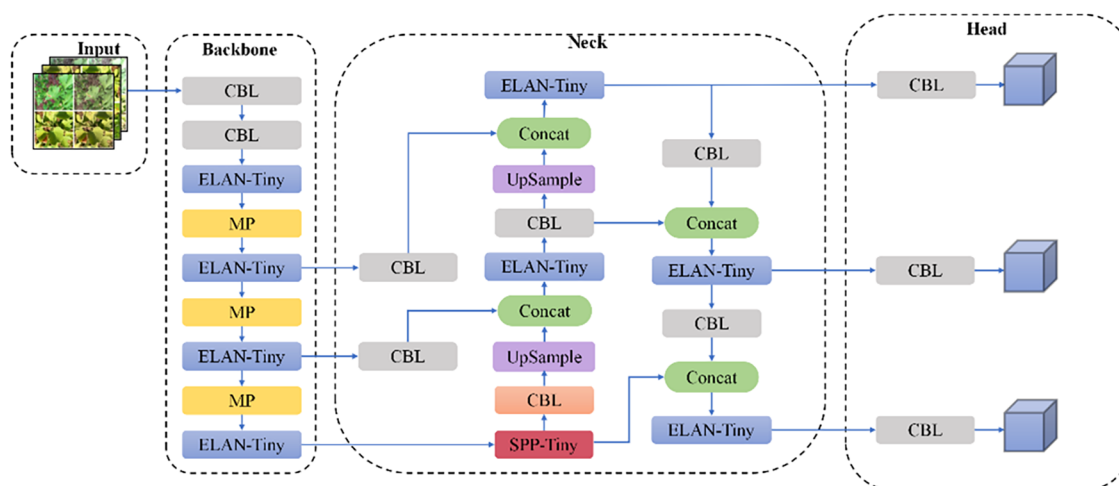


FIGURE 4
YOLOv7 network structure diagram.

2.3 PLFYNet object detection model

2.3.1 Network reconstruction and optimization

A systematic evaluation of five lightweight architectures—MobileNetV3, GhostNet, ShuffleNetV2, PP-PicoDet, and PP-LCNet—was conducted to identify the optimal backbone replacement for YOLOv7-tiny in Ginkgo disease detection applications (Wang et al., 2025b; Shen et al., 2023; Han and Yang, 2021; Lu et al., 2025a). Through comprehensive benchmarking, PP-LCNet emerged as the superior choice, demonstrating exceptional performance metrics.

As a CNN architecture specifically engineered for mobile deployment (Xue and Wang, 2025), PP-LCNet employs a streamlined design featuring sequential convolutional and pooling operations tailored for real-time inference. The architecture's foundation consists of DepthSepConv blocks, which utilize dual convolutional layers paired with H-Swish activation functions for computational efficiency (Huang et al., 2023). Strategic integration of Squeeze-and-Excitation mechanisms in select modules yields DepthSepConvSE variants with enhanced representational capacity.

A distinctive feature of PP-LCNet is its architectural departure from standard classification models through the incorporation of a 1280-channel 1×1 convolutional layer following global average pooling. This design innovation delivers improved classification performance without compromising inference speed. The overall structure of PP-LCNet is shown in Figure 5.

Figure 6 depicts the architectural diagram of the modified YOLOv7-tiny incorporating a PP-LCNet-inspired backbone reconstruction based on DepthSepConv modules. The redesigned backbone initiates with a conventional convolutional layer, subsequently progressing through a cascade of DepthSepConv blocks featuring dual kernel configurations: 3×3 and 5×5 convolutions. This heterogeneous kernel design facilitates multi-scale feature extraction by capturing information across diverse receptive field dimensions.

Table 1 delineates the architectural specifications and parametric configurations of the modified backbone network through five essential components: (i) repetition factor (n), indicating the iteration count for each module; (ii) parameter volume (Params), quantifying trainable weights per stage; (iii) module designation, specifying the computational blocks utilized; (iv) configuration details, encompassing comprehensive parametric settings; and (v) output dimensionality, representing spatial resolution and channel depth ($H \times W \times C$).

Parametric distinctions between modules are evident: Standard Convolution operates with a five-tuple specification (input channels, output channels, kernel dimensions, stride, padding), whereas DepthSepConv employs a reduced four-parameter scheme (input/output channels, kernel size, stride), notably omitting padding specifications and SE attention integration.

2.3.2 The light-YOLO network model integrating the attention mechanism

A comprehensive benchmarking study was conducted on the Ginkgo biloba pathology dataset to assess three prominent attention mechanisms—SE, CBAM, and ECA (Ai et al., 2025; Fareed et al., 2025; Waghumbare et al., 2024)—within the Light-YOLO framework under controlled experimental protocols. Performance quantification employed a four-metric evaluation suite comprising parameter efficiency, detection precision, recall rate, and mAP0.5. Empirical results demonstrated ECA's dominant performance across all assessed dimensions, warranting its selection for the proposed architecture.

2.3.3 Neck network improvement

The YOLOv7-tiny architecture prioritizes computational efficiency through reduced network depth and parameter count, facilitating deployment on resource-constrained platforms. This architectural parsimony, however, compromises detection performance, manifesting as elevated false negative rates in

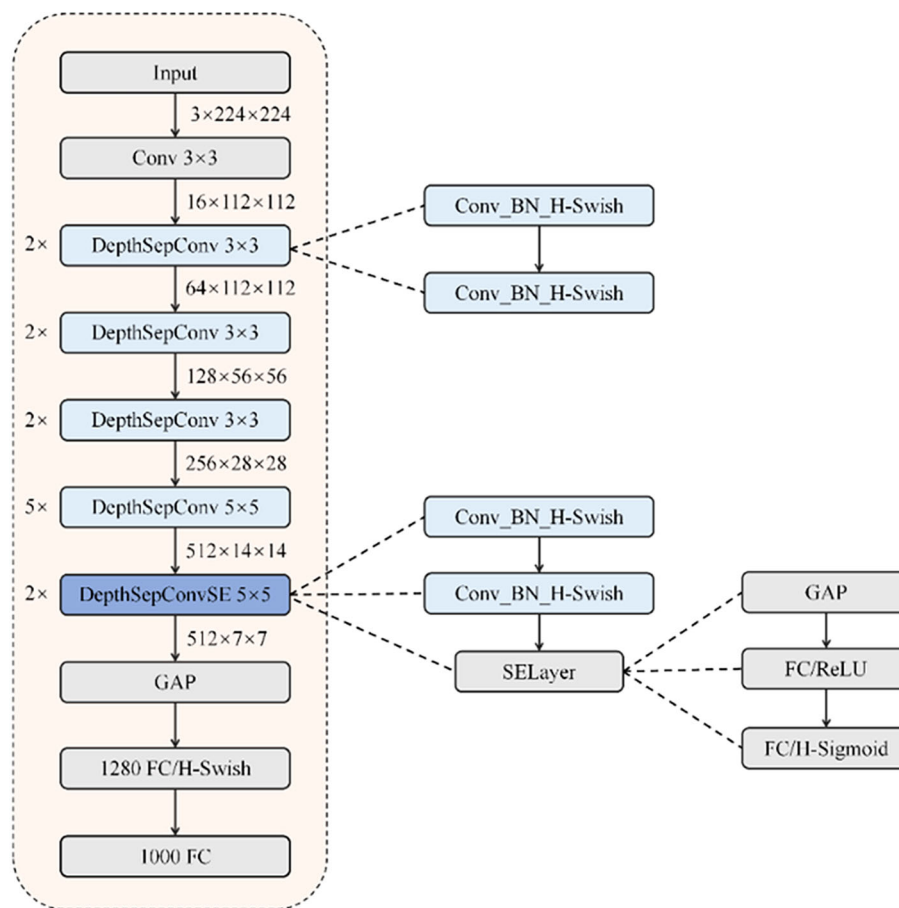


FIGURE 5
PP-LCNet network structure.

complex visual environments. Addressing these limitations, we introduce three targeted architectural modifications tailored for Ginkgo biloba pathology detection, emphasizing enhanced discrimination of diminutive lesions within visually cluttered canopy environments:

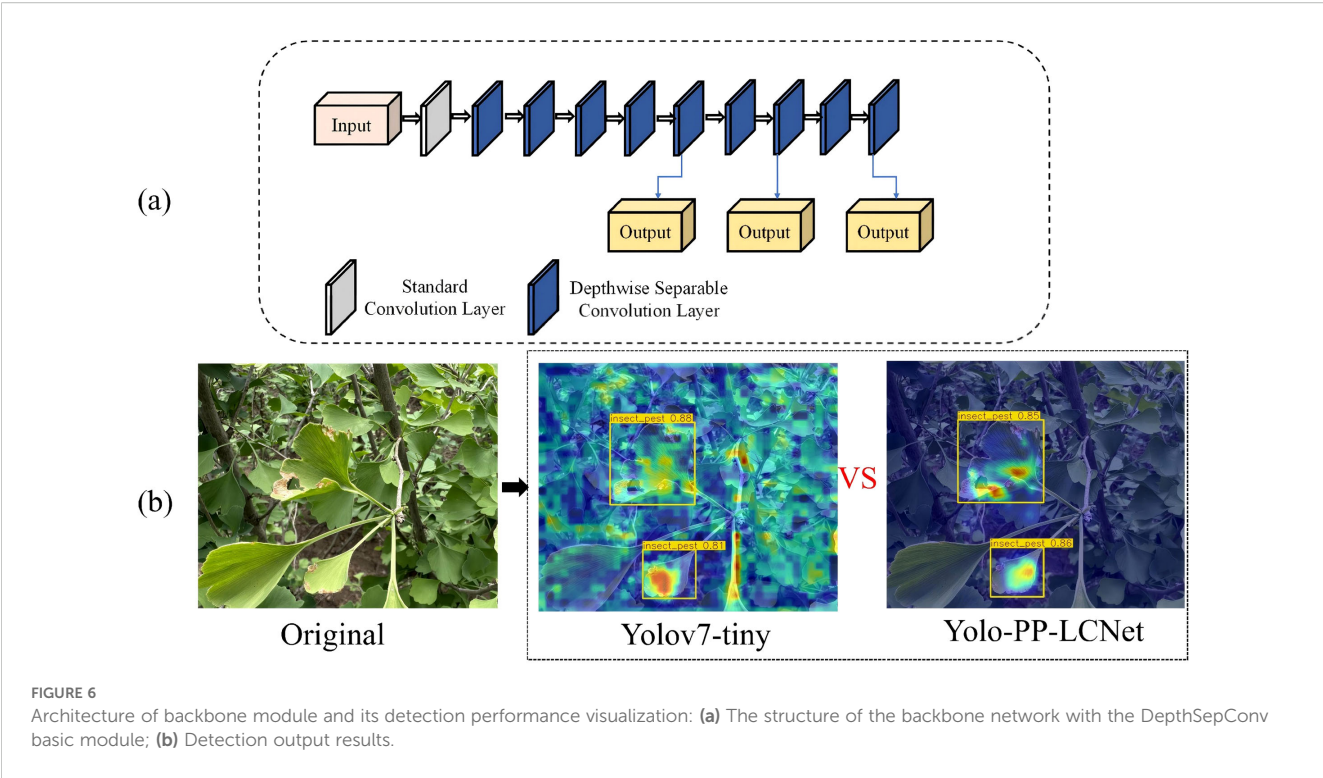
1. Bottleneck Transformer (BoT) Module: Incorporates self-attention mechanisms to model long-range spatial dependencies, augmenting global context awareness. This integration substantially enhances detection sensitivity and localization accuracy for small-scale pathological features.
2. Bidirectional ECA-enhanced Feature Fusion (BiECAFusion): Implements bidirectional feature propagation across multiple scales while preserving fine-grained spatial details. This dual-pathway architecture synergistically combines high-level semantic representations with low-level textural features, yielding superior detection robustness.
3. Shape-IoU Loss Formulation: Supersedes conventional CIoU with a geometry-aware loss function (Zhao et al., 2024), achieving improved alignment between predicted and ground-truth bounding boxes. This refinement

particularly benefits the localization of morphologically diverse disease manifestations.

The structure of the CBL, ELAN, and SPPCPC modules is shown in Figure 7.

2.3.4 Fused bottleneck transformer module

The Bottleneck Transformer architecture substitutes ResNet's conventional 3×3 convolutions with a hybrid attention-convolution design (Srinivas et al., 2021). The module architecture consists of two principal components: a multi-head self-attention (MHSA) mechanism and a nonlinear projection block. The MHSA component performs dimensional decomposition of input features, enabling parallelized computation while establishing global spatial dependencies (Lu et al., 2025b). Following attention computation, the nonlinear projection block employs dual linear transformations to introduce essential nonlinearities. Hierarchical stacking of these modules, as illustrated in Figure 8, yields a deep architecture that synergistically leverages convolutional and transformer paradigms (Yang et al., 2025; Yu and Zhou, 2023). This design logic aligned with the discriminative feature modeling philosophy in DFPNet, which enhances fine-grained target



representation via structured feature pyramid (Xie et al., 2024). This design philosophy achieves enhanced feature representation through global context modeling while maintaining computational efficiency. Within the YOLOv7 framework, this integration yields simultaneous improvements in detection accuracy and computational performance, facilitating real-time inference with reduced deployment overhead.

2.3.5 Fused BiECAFusion structure

The original YOLOv7 architecture employs PANet (Path Aggregation Network) for feature aggregation through unidirectional pathways (Zhang et al., 2023c). This approach exhibits inherent limitations when applied to Ginkgo biloba

pathology detection, particularly in preserving discriminative feature representations across scales, resulting in suboptimal learning dynamics. To mitigate these deficiencies, we propose BiECAFusion (Bidirectional ECA-enhanced Feature Fusion), a novel feature aggregation module that facilitates bidirectional information flow while incorporating channel-wise attention mechanisms. The architectural design is detailed in Figure 9.

BiECAFusion is a fundamental redesign of feature fusion strategies. It is specifically developed to tackle the challenges of small lesion detection in complex foliar environments. The module operates in three stages. First, 1×1 convolutions are used for channel dimensionality harmonization. This ensures compatible feature representations across different hierarchical levels. Second, the

TABLE 1 Backbone network structure and parameterization.

Serial number	n	Params	Module	Configuration	Output size
0	1	464	Conv	[3, 16, 3, 2, 1]	16×320×320
1	1	752	DepthSepConv	[16, 32, 3, 1]	32×320×320
2	1	2528	DepthSepConv	[32, 64, 3, 2]	64×160×160
3	1	4928	DepthSepConv	[64, 64, 3, 1]	64×160×6
4	1	9152	DepthSepConv	[64, 128, 3, 2]	128×80×80
5	1	18048	DepthSepConv	[128, 128, 3, 1]	128×80×80
6	1	34688	DepthSepConv	[128, 256, 3, 2]	256×40×40
7	5	364800	DepthSepConv	[256, 256, 5, 1]	256×40×40
8	1	139008	DepthSepConv	[256, 512, 5, 2]	512×20×20
9	1	276992	DepthSepConv	[512, 512, 5, 1]	512×20×20

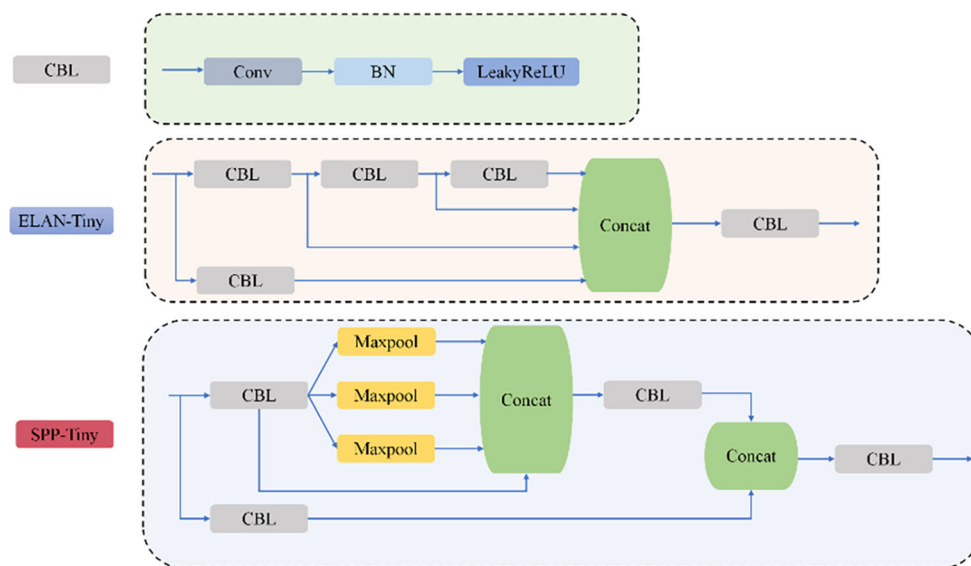


FIGURE 7
CBL, ELAN, and SPPCPC modules.

Efficient Channel Attention (ECA) mechanism is integrated. It dynamically recalibrates channel responses, highlighting discriminative features and suppressing background noise. This is crucial for detecting tiny pathological indicators with minimal spatial coverage. Finally, the bidirectional fusion paradigm retains both fine-grained spatial details and high-level semantic information. It achieves this through reciprocal feature enhancement (Tang et al., 2025; Xie et al., 2025b).

The BiECAFusion Module differs significantly from the traditional concat fusion in FPN/PAN in terms of fusion logic and attention mechanism. Traditional FPN/PAN achieves fusion by simply concatenating high-level (semantic-rich, low-resolution) and low-level (detail-rich, high-resolution) features, where all feature channels are treated equally without adaptive weighting. BiECAFusion replaces the original SE attention with ECA attention and adopts a dual-branch interaction mechanism: it first adjusts the channel dimensions of input features via 1×1 convolution to ensure compatibility, concatenates the features, applies ECA attention to generate channel-wise weights, splits the weights into branches corresponding to input features, and then fuses each branch feature with the weighted feature of the other branch (e.g., $x_0 + x_1_weight$, $x_1 + x_0_weight$), realizing bidirectional mutual enhancement of cross-level features. Besides, traditional concat lacks an attention

mechanism and tends to retain redundant or irrelevant features (e.g., mixing non-lesion leaf textures with critical lesion edges), while BiECAFusion's ECA attention can adaptively highlight effective channels, focus on high-frequency features like lesion edges, and suppress noise such as irrelevant leaf veins, solving the "feature confusion" problem in traditional fusion.

In this study, BiECAFusion demonstrates three key advantages. First, it enhances lesion edge discrimination: lesion edges are crucial for detecting ginkgo leaf diseases (e.g., yellow spots, brown blight) but easily confused with dense leaf veins, and its ECA attention prioritizes high-frequency edge features, while dual-branch interaction strengthens the correlation between semantic (lesion category) and geometric (edge shape) information, facilitating small lesion recognition. Second, it enables efficient cross-level feature interaction: unlike the one-way fusion of FPN/PAN (high-to-low or low-to-high), BiECAFusion allows x_0 to benefit from x_1 's details and x_1 from x_0 's semantics, adapting to ginkgo scenarios where lesions vary greatly in size (from tiny spots to large patches) and require balanced use of multi-scale features. Third, it has a lightweight design: ECA attention removes fully connected layers, which aligns with the study's goal of developing edge-deployable ginkgo disease detection models, ensuring fusion efficiency without increasing computational burden.



FIGURE 8
Bottleneck transformer module.

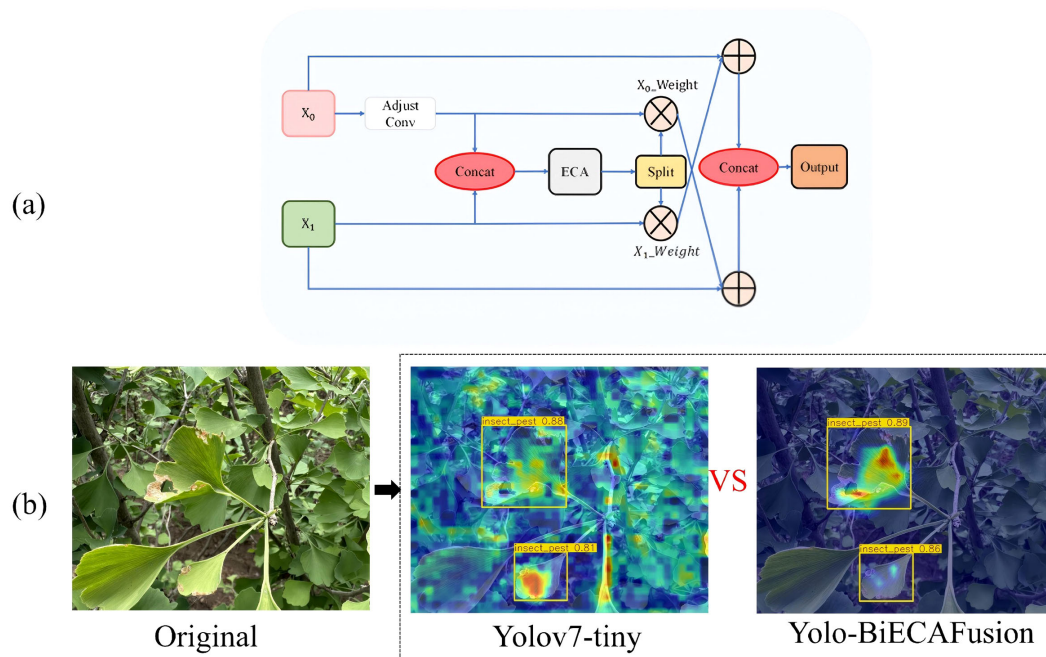


FIGURE 9

Architecture of BiECAFusion module and its detection performance visualization: (a) Bidirectional ECA-enhanced feature fusion structure; (b) Detection output results.

- Dimensional Harmonization:** Employing 1×1 convolutions for channel normalization facilitates seamless cross-scale integration while minimizing computational complexity.
- Adaptive Channel Recalibration:** The ECA mechanism generates channel-specific attention weights with minimal overhead, selectively amplifying features relevant to small-scale pathological patterns.
- Bidirectional Feature Synthesis:** The reciprocal enhancement strategy ($x_0 + w_1x_1$, $x_1 + w_0x_0$) establishes complementary interactions between spatially-rich shallow features and semantically-rich deep representations.

2.3.6 Loss function improvement

The YOLOv7 architecture uses a multi-component loss. It covers confidence, localization and classification objectives. For bounding box regression, the baseline uses Complete Intersection over Union (CIoU). CIoU combines overlap ratio, centroid distance and aspect ratio consistency. But CIoU has flaws. It performs poorly when predicted and ground truth aspect ratios are similar. This causes suboptimal convergence.

We adopt Shape-IoU for ginkgo leaf disease detection. It explicitly models geometric links between predicted and target boxes. For irregular ginkgo lesions (e.g., leaf spots), it optimizes shape-aware localization. This fixes CIoU's flaws and boosts localization accuracy for ginkgo scenarios.

The Shape-IoU loss function is calculated as follows (Equations 1-6):

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \quad (1)$$

$$ww = \frac{2 \times (w^{gt})^{scale}}{(w^{gt})^{scale} + (h^{gt})^{scale}} \quad (2)$$

$$hh = \frac{2 \times (h^{gt})^{scale}}{(w^{gt})^{scale} + (h^{gt})^{scale}} \quad (3)$$

$$distance^{shape} = hh \times (x_c - x_c^{gt})^2 / c^2 + ww \times (y_c - y_c^{gt})^2 / c^2 \quad (4)$$

$$\Omega^{shape} = \sum_{t=w,h} (1 - e^{-\omega_t})^\theta, \theta = 4 \quad (5)$$

$$\begin{cases} \omega_w = hh \times \frac{|w - w^{gt}|}{\max(w, w^{gt})} \\ \omega_h = ww \times \frac{|h - h^{gt}|}{\max(h, h^{gt})} \end{cases} \quad (6)$$

The scale parameter is empirically calibrated based on the object size distribution within the training corpus. Directional weight coefficients, denoted as ww and hh for horizontal and vertical axes respectively, are dynamically computed from the geometric properties of ground truth annotations. The comprehensive Shape-IoU loss integrates IoU, shape distance, and shape penalty, as shown in Equation 7.

$$L_{Shape-IoU} = 1 - IoU + distance^{shape} + 0.5 \times \Omega^{Shape} \quad (7)$$

2.3.7 Detection head improvement

DyHead constitutes an innovative detection head architecture that employs multi-dimensional self-attention to enhance feature discrimination across scale, spatial, and semantic dimensions (Dai et al., 2021). The framework's distinctive characteristic lies in its unified attention mechanism that operates orthogonally across feature pyramid levels (L), spatial locations (S), and channel dimensions (C), facilitating holistic feature refinement. Through this tripartite attention strategy, DyHead augments the detection head's representational power without imposing computational penalties, thereby achieving an optimal balance between detection performance and computational efficiency (Gong et al., 2024). Formally, for a feature tensor $F \in \mathbb{R}^{(L \times S \times C)}$, the self-attention mechanism is expressed as in Equation 8:

$$W(F) = \pi(F) \cdot F \quad (8)$$

In this formulation, $\pi(\cdot)$ represents the attention transformation function. Although fully-connected architectures could theoretically model such high-dimensional interactions, the computational complexity of simultaneously learning across all tensor dimensions renders this approach intractable. Consequently, we adopt a factorized attention strategy, decomposing the operation into three consecutive transformations, with each targeting a specific dimensional axis independently.

$$W(F) = \pi_C(\pi_S(\pi_L(F) \cdot F) \cdot F) \cdot F \quad (9)$$

$\pi_L(\cdot)$, $\pi_S(\cdot)$ and $\pi_C(\cdot)$ represent dimension-specific attention transformations corresponding to level, spatial, and channel axes, respectively. Sequential execution of these operators ensures computational tractability while preserving inter-dimensional dependencies. The formulation in Equation 9 inherently supports recursive composition, facilitating the construction of deep

architectures through cascaded π_L , π_S , and π_C modules, as depicted in the architectural diagram of Figure 10.

2.3.8 Improved network structure

The culmination of these architectural innovations yields LCNET-FusionYOLO, whose comprehensive topology is delineated in Figure 11.

2.3.9 LAMP model pruning

Following model convergence, we implement LAMP (Layer-Adaptive Magnitude-based Pruning) to further compress the architecture for edge deployment (Lee et al., 2020). High-resolution disease image computational demands present significant constraints for resource-limited embedded platforms. To address these challenges, LAMP uses adaptive layer-wise sparsification, where connection importance is quantified by normalized weight magnitudes. The algorithm computes relative significance scores by normalizing squared weight values against the aggregate magnitude of retained parameters within each layer, enabling automatic derivation of layer-specific pruning ratios. This approach ensures optimal model compression while maintaining detection fidelity for embedded applications (Yuan et al., 2025).

The methodology fundamentally balances sparsity optimization with performance preservation through adaptive global pruning. Weight salience is determined via magnitude-based scoring coupled with ℓ_2 distortion minimization at the network level. The algorithmic pipeline comprises:

Sorting according to the magnitude of weights: Parameters within each network layer are arranged in descending order based on their absolute values, establishing a magnitude-based hierarchy. LAMP Score Calculation: The algorithm computes normalized importance metrics by evaluating the squared magnitude of each

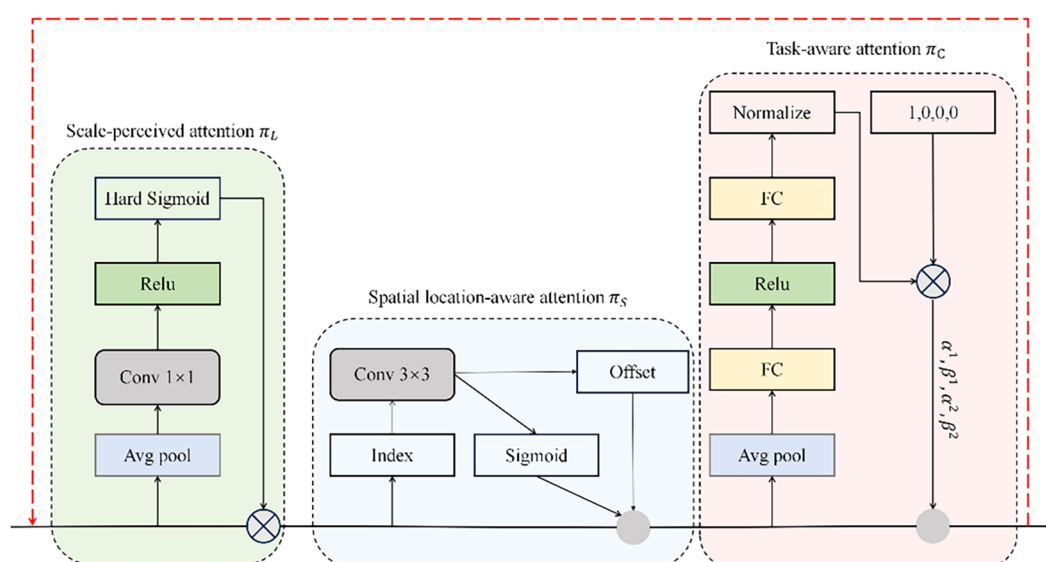


FIGURE 10
DyHead structure.

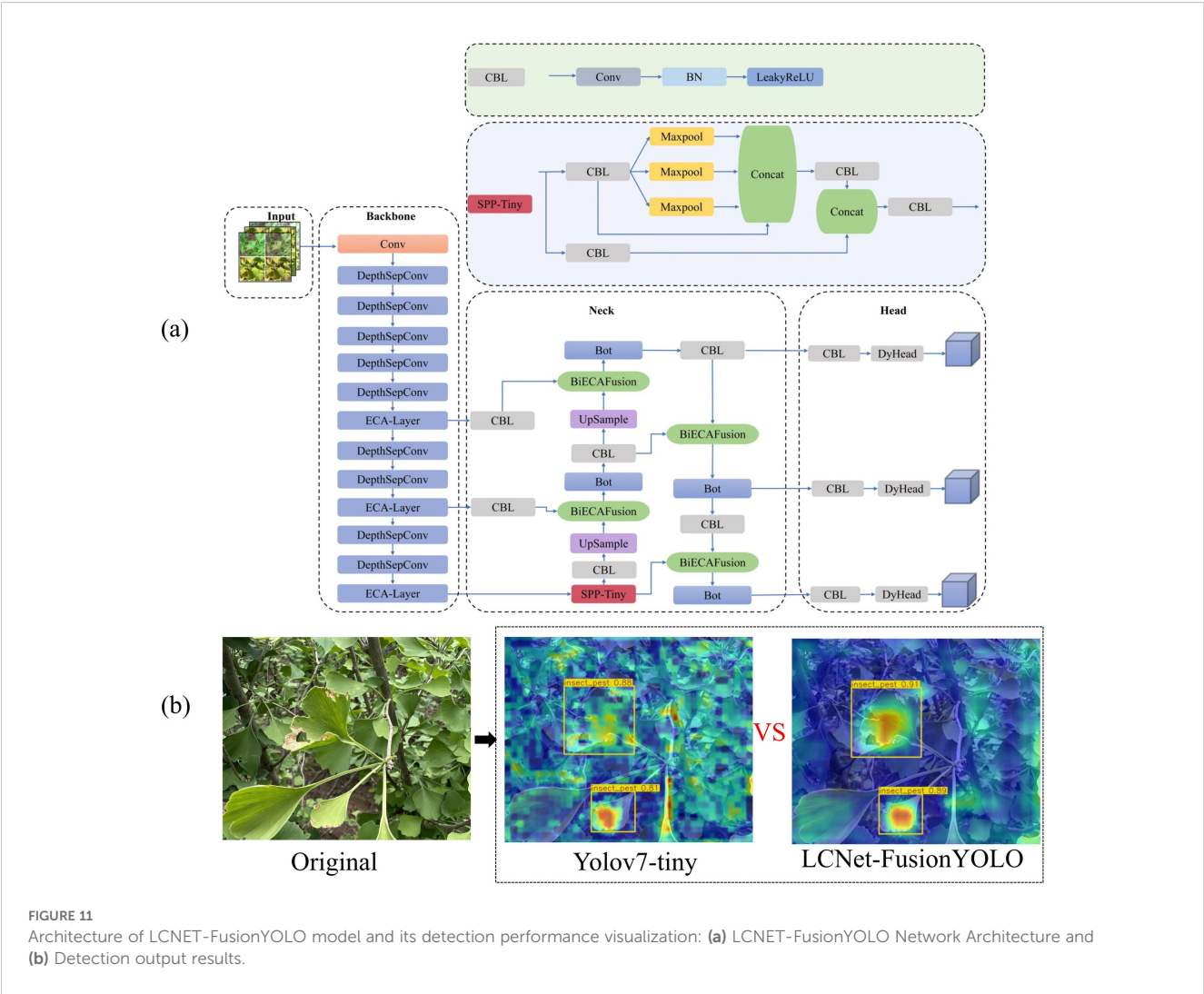


FIGURE 11 Architecture of LCNET-FusionYOLO model and its detection performance visualization: (a) LCNET-FusionYOLO Network Architecture and (b) Detection output results.

weight relative to the layer’s weight distribution, yielding calibrated significance scores. Global Pooling and Pruning: Layer-specific scores undergo aggregation into a unified importance repository, followed by comprehensive ranking and systematic parameter elimination based on global thresholds. This methodology achieves optimal compression while preserving essential model capabilities.

2.4 Evaluation indicators

Model selection criteria encompass five comprehensive evaluation metrics: detection precision (P), sensitivity/recall (R),

mean average precision (mAP@0.5), computational burden quantified through GFLOPS, and temporal efficiency measured via frames per second (FPS). The mAP@0.5 metric specifically denotes the averaged precision values computed across all disease categories at an intersection-over-union threshold of 0.5.

$$P = \frac{TP}{TP + FP}$$
 (10)

$$R = \frac{TP}{TP + FN}$$
 (11)

$$AP = \int_0^1 P(R) dR$$
 (12)

$$mAP = \frac{\sum_{k=1}^N AP_k}{N}$$
 (13)

The core evaluation metrics for model performance—Precision, Recall, AP, and mAP—are calculated using Equations 10–13. Within these formulations, TP represents true positive detections where diseased instances are correctly identified, FP indicates false

TABLE 3 Lightweighting results for different backbone network models.

Basic model	MobileNetV3	GhostNet	ShuffleNetV2	PP-PicoDet	PP-LCNet	Parameter (million)	Memory (MB)	Precision (%)	Recall (%)	mAP@0.5 (%)
YOLOv7-tiny	✓					6.0	12.4	89.9	83.7	89.7
						1.5	3.2	88.6	83.5	89.9
		✓				7.2	14.1	90.6	87.6	91.9
			✓			4.3	8.5	91.1	87.2	91.1
				✓		4.3	8.7	82.6	84.7	87.7
					✓	4.4	8.6	91.4	87.7	91.2

positives arising from misclassification of healthy samples as diseased, FN denotes false negatives occurring when pathological cases are erroneously classified as healthy, and N signifies the total number of disease categories under consideration.

2.5 Experimental environment

The algorithmic model experiments in this paper were conducted on the Ubuntu operating system, with a CPU of 12 vCPU Intel(R) Xeon(R) Platinum 8352V CPU @ 2.10GHz, and GPUs of NVIDIA RTX 3080x2 (20GB). The Python programming language was adopted, and the PyTorch 3.8, CUDA 11.8, and CUDNN deep learning framework was used for model training and inference. The main parameter settings shown in Table 2. All baseline models (YOLOv7-tiny, SSD, Faster R-CNN, etc.) were trained under the same conditions. and the dataset comprised 7, 158 images collected in this study.

3 Results

3.1 Comparison of ablation experiments

3.1.1 Analysis of network model lightweighting results

Table 3 presents a comprehensive performance analysis of YOLOv7-tiny variants reconstructed with five alternative backbone architectures: MobileNetV3, GhostNet, ShuffleNetV2, PP-PicoDet, and PP-LCNet. The comparative evaluation encompasses computational metrics (parameter count and memory footprint) alongside detection performance indicators (precision, recall, and mAP@0.5).

Experimental analysis reveals a consistent trade-off between parameter efficiency and detection performance across all evaluated architectures. While MobileNetV3 achieves substantial parameter reduction (74.5%), its marginal mAP@0.5 gain (0.2%) coupled with notable precision degradation (1.3%) renders it suboptimal for this application. GhostNet exhibits counterproductive behavior, expanding parameters by 56% relative to baseline, thus eliminating it from consideration. The remaining architectures—ShuffleNetV2, PP-PicoDet, and PP-LCNet—demonstrate comparable parameter compression (approximately 28%), with PP-LCNet emerging as the superior variant, achieving 91.2% mAP@0.5. This configuration yields a 26.8% parameter reduction while enhancing recall by 4.0% and preserving precision levels, substantiating the architectural choice.

3.1.2 Analysis of light-YOLO network models incorporating attention mechanisms

Building upon the Light-YOLO architecture, we conducted systematic ablation studies incorporating three prominent attention mechanisms—SE, CBAM, and ECA—to assess their respective contributions to detection performance. Table 4 presents a comprehensive comparative analysis of these attention-augmented variants.

TABLE 4 Modelling results incorporating different attention mechanisms.

Basic model	PP-LCNet	SE	CBAM	ECA	Parameter (million)	Precision (%)	Recall (%)	mAP@0.5 (%)
YOLOv7-tiny	✓				4.4	91.4	87.7	91.2
	✓	✓			4.8	91.3	87.6	91.4
	✓		✓		4.5	87.5	89.3	90.6
	✓			✓	4.4	88.8	88.8	93.1

Empirical analysis reveals distinct performance characteristics across the evaluated attention mechanisms. SE module integration incurs an 8.6% parameter overhead while yielding marginal performance shifts—a 0.1% precision reduction offset by equivalent recall improvement and 0.2% mAP@0.5 gain. CBAM demonstrates inferior performance, with mAP@0.5 falling 0.6% below SE-augmented models, suggesting suboptimal feature recalibration for this application. Conversely, ECA exhibits superior efficiency, maintaining parameter parity with the baseline while achieving substantial performance gains: 1.1% recall enhancement and 1.9% mAP@0.5 improvement. These results establish ECA as the optimal attention mechanism for Ginkgo pathology detection, warranting its integration into the PP-LCNet backbone for enhanced detection fidelity.

3.1.3 Performance comparison of fusion improvement modules

Following the empirical validation of ECA as the optimal attention mechanism for the PP-LCNet backbone, we conducted systematic ablation studies to assess individual and combined contributions of the proposed enhancements. The evaluation encompassed baseline YOLOv7-tiny alongside variants incorporating BoT, BiECAFusion, Shape-IoU, and DyHead modifications, culminating in the fully integrated BBSD-YOLO architecture. Performance quantification employed a comprehensive metric suite—parameter count, memory footprint, precision, recall, and mAP@0.5—with comparative results tabulated in Table 5.

Table 5 demonstrates consistent performance improvements across all proposed architectural modifications relative to baseline YOLOv7-tiny. The BoT module, while increasing parameters by 48.2% (4.8MB memory overhead), yields substantial gains: 2.6% precision, 7.0% recall, and 3.4% mAP@0.5. BiECAFusion achieves parameter-efficient enhancement, maintaining computational complexity while delivering incremental improvements (0.48% precision, 2.6% recall, 0.4% mAP@0.5). Shape-IoU optimization significantly enhances localization accuracy, contributing 2.5% precision, 7.9% recall, and 3.5% mAP@0.5 gains. DyHead simultaneously reduces parameters while improving detection metrics (1.3% precision, 7.1% recall, 1.5% mAP@0.5). Synergistic integration of these components in BBSD-YOLO achieves optimal performance: 92.8% precision, 90.2% recall, and 93.5% mAP@0.5—

representing a 4.9 percentage point improvement over baseline. The complete LCNET-FusionYOLO architecture further elevates performance to 94.1% precision, 91.8% recall, and 94.1% mAP@0.5.

Figure 12 presents convergence analysis comparing baseline YOLOv7-tiny with LCNET-FusionYOLO across training iterations. The visualization reveals LCNET-FusionYOLO’s superior optimization characteristics: accelerated convergence, enhanced asymptotic performance, and improved training stability across all metrics, with particularly pronounced advantages in mAP@0.5 convergence dynamics. YOLOv7-tiny has an unstable Jaccard index when using the AdaDelta optimizer, and this is caused by five flaws. (a) Rigid pruning leads to poor depth-width coupling, no adaptive modules, unbalanced gradient propagation and high-frequency oscillations. LCNet-FusionYOLO (with PP-LCNet and four core modules) uses ECA, BiECAFusion and DyHead to stabilize the Jaccard index. (b) CIoU loss has sharp curvature, which causes gradient spikes and coordinate shifts. LCNet-FusionYOLO’s Shape-IoU ensures gradual regression. (c) Unidirectional PANet dilutes small-target features. BiECAFusion enables bidirectional feature flow to retain these features. (d) The static detection head is sensitive to semantic-scale drift. DyHead’s real-time reparameterization avoids step changes in the Jaccard index. (e) YOLOv7-tiny’s parameter manifold is non-convex, making AdaDelta oscillate. LCNet-FusionYOLO’s implicit regularization smooths gradient updates, resulting in stable Jaccard index curves.

3.2 Comparison of pruning ablation experiments

Post-pruning evaluation employed the speed_up metric, defined as the computational ratio between pruned and unpruned architectures. This metric quantifies efficiency gains, where speed_up = 1.6 corresponds to a 37.5% computational reduction (1 - 1/1.6). The pruning protocol systematically identifies and eliminates redundant structures while preserving essential architectural components. Subsequently, iterative fine-tuning recovers potential performance degradation induced by sparsification.

After LAMP pruning, the model was fine-tuned. It used the same SGD optimizer as pre-pruning training. Key hyperparameters were adjusted to suit the sparse architecture. This ensured stable

TABLE 5 Performance comparison of fusion improvement modules.

Basic model	BoT	BiECAFusion	Shape-IoU	DyHead	PP-LCNet	ECA	Parameter (million)	Memory (MB)	Precision (%)	Recall (%)	mAP@0.5 (%)
YOLOv7-tiny							6.0	12.4	89.9	83.7	89.7
	✓						8.9	17.2	92.5	90.7	93.1
		✓					6.0	8.6	89.4	86.3	90.1
			✓				6.0	11.6	92.4	91.6	93.2
				✓			5.9	10.5	91.2	90.8	91.2
	✓		✓	✓			7.8	12.7	92.8	90.2	93.5
	✓	✓	✓	✓	✓	✓	6.2	10.2	94.1	91.8	94.1

convergence and performance retention. Core settings are as follows: initial learning rate (lr0) = 0.001, which was reduced from pre-pruning 0.01 to avoid gradient explosion in the sparse model; final learning rate (lrf) = 0.01, which is consistent with pre-pruning and calculated as $lr0 \times lrf$; momentum = 0.937; weight decay = 0.0005; warmup epochs = 3.0; warmup momentum = 0.8; warmup bias lr = 0.1.

Fine-tuning lasted 200 epochs, which is the same as pre-pruning. The learning rate decayed linearly from lr0 to lrf×lr0. We selected lr0 = 0.001 instead of 0.01. The pruned model is more sensitive to high learning rates. Tests showed lr0 = 0.01 caused unstable convergence. lr0 = 0.001 balanced convergence speed and detection performance.

To determine the optimal speed_up, we tested targets from 1.0 to 2.43 with a layer-adaptive magnitude pruning strategy. At speed_up = 2.2, PLFYNet hits 94.5% mAP@0.5 (0.4% higher than pre-pruning LCNET-FusionYOLO). Its parameters drop to 2.98M (50.5% reduction), enabling 50.5 FPS on Jetson Orin Nano (meets <50ms/frame for edge use). Though speed_up = 2.43 is the upper limit, it causes over-pruning: mAP@0.5 falls to 94.2%. Thus, speed_up = 2.2 balances efficiency and performance best. Data supports this in Table 6.

3.3 Comparison with mainstream object detection models

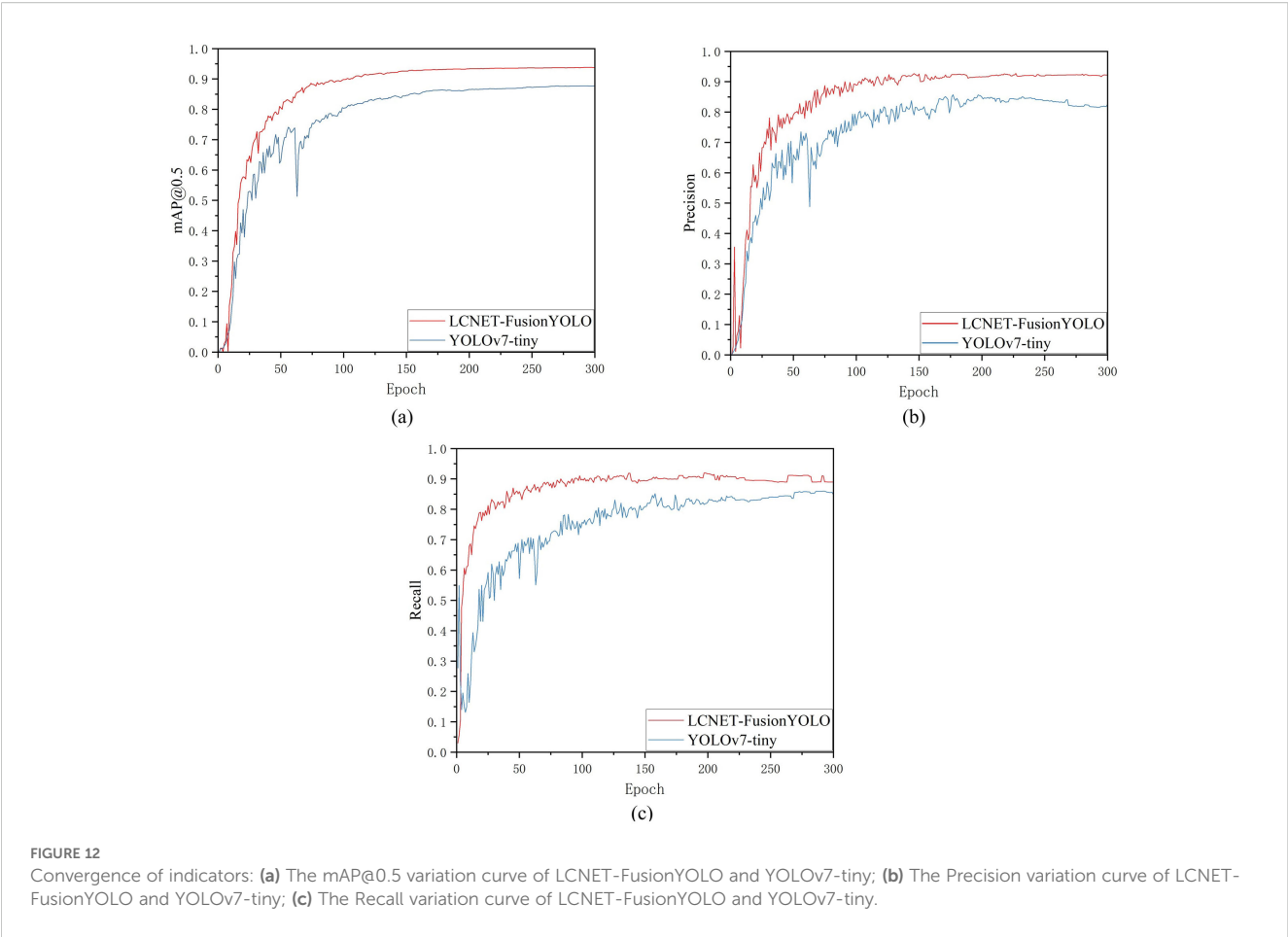
In this subsection, the final improved model LCNET-FusionYOLO is experimentally compared with eight other common models. The specific experimental results are shown in Table 7.

The LCNET-FusionYOLO and PLFYNet models were experimentally compared with eight other common object detection models in terms of their mAP@0.5 metrics.

The optimization pipeline employed LAMP pruning for systematic weight elimination, followed by structured sparsification and 200-epoch fine-tuning to mitigate performance degradation. This dual-phase optimization yielded PLFYNet, which demonstrates a 0.4% mAP@0.5 improvement over LCNET-FusionYOLO post-refinement, as illustrated in Figure 13.

Comparative analysis against established architectures reveals PLFYNet’s superior efficiency-performance trade-off. Two-stage detectors exemplified by Faster R-CNN achieve competitive recall (90.1%) but suffer from computational intractability (137M parameters) and inferior precision (83.2%) relative to single-stage alternatives. SSD’s multi-scale prediction paradigm reduces complexity (26.3M parameters) yet exhibits suboptimal detection metrics (mAP@0.5: 90.3%, recall: 88.6%). DEIM-n’s architectural constraint—utilizing only P4/P5 feature levels—inherently limits small object detection capability, yielding mAP@0.5 of 90.5%. PLFYNet finally surpasses RT-DETR-r18 by 1.5% in Precision, 3% in Recall, and 1.6% in mAP@0.5, demonstrating superior performance.

Within the YOLO lineage, evolutionary progression demonstrates continuous refinement. YOLOv3’s architectural simplicity (61.53M parameters) correlates with inadequate



performance (recall: 73.7%, mAP@0.5: 84.6%). Contemporary variants YOLOv5s and YOLOv8n leverage depthwise separable convolutions to achieve substantial compression (7.22M and 3.16M parameters respectively), though performance remains inferior to PLFYNet. Our proposed architecture, with merely 3.0M parameters, achieves state-of-the-art metrics (precision: 94.2%, recall: 92.1%, mAP@0.5: 94.5%)—surpassing Faster R-CNN by 7.6 and SSD by 4.2 percentage points while utilizing

2.2%–21.8% of their parameters. This validates the efficacy of integrated pruning and architectural optimization in achieving unprecedented efficiency without compromising detection fidelity.

The normalized confusion matrices show that YOLOv7-tiny achieves recall rates of 0.89, 0.85, and 0.88 for chlorosis, insect_pest, and physical_damage, respectively, with a background recall of 0.67 and a main false-positive rate of background misclassified as chlorosis (0.17). In contrast, PLFYNet significantly improves the recall of the three disease classes to 0.96, 0.93, and 0.94, markedly reducing missed detections of insect_pest and physical_damage; however, background recall drops to 0.78. When the model training incorporates over-augmented background noise, the chlorosis false-positive rate increases to 0.12. Overall, PLFYNet markedly enhances disease detection accuracy by strengthening foreground feature extraction at the cost of diminished background discriminability, exhibiting a biased improvement of “more accurate foreground, more confused background,” making it suitable for applications demanding high sensitivity to diseases. It is shown in [Figure 14](#).

TABLE 6 Results of different speed_up data.

Speed_up	Parameter (number)	Memory (MB)	mAP@0.5 (%)
1.0	6.2	10.2	94.1
1.2	5.2	9.3	94.2
1.4	4.4	8.6	94.2
1.6	3.9	7.6	94.3
1.8	3.4	6.7	94.4
2.0	3.1	6.2	94.4
2.2	3.0	5.9	94.5
2.4	2.9	5.7	94.2

3.4 External validation of the public dataset

[Table 8](#) compares the core performance metrics of two models in the relevant task (inferred as plant disease detection based on

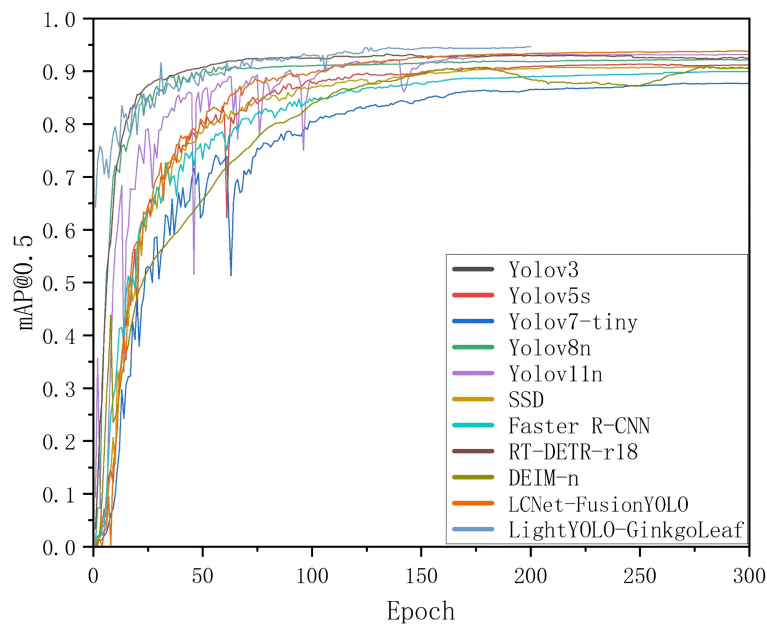


FIGURE 13
Comparison of mAP0.5 Among LCNET-FusionYOLO, PLFYNet, and eight other common models.

data characteristics). Among them, PLFYNet performs better, with its precision (56.1%), recall (55.9%), and mAP@0.5 (55.7%) all higher than those of YOLOv7-tiny, indicating that the former is superior in recognition accuracy, missed detection control, and comprehensive detection capability. Our Ginkgo-focused dataset has unique traits and high-quality data, but PlantDoc covers 13 plants and 17 diseases with low-quality data. PLFYNet, pruned for edge use, lacks generalization and feature learning, causing low mAP@0.5. PlantDoc’s uneven distribution of different types of labels leads to a significant reduction in learning outcomes. However, the results are still improved compared to the original YOLOv7-tiny model, which is more suitable for marginalised deployments.

3.5 Models deployed on Jetson Orin Nano

3.5.1 Hardware and environment

The Jetson Orin Nano hardware configuration is shown in Table 9 below:

Jetson Orin Nano Environment Setup: Install Ubuntu 20.04 operating system, configure the model runtime environment with JetPack 5.1, Python 3.8, PyTorch 1.10.0, TorchVision 0.11.0, CUDA 11.8, and CUDNN 8.6. Integrate a CSI camera for hardware acceleration.

3.5.2 Hardware modules and configurations

The hardware deployment configuration centers on the Jetson Orin Nano platform, integrated with essential peripherals including

TABLE 7 Results of comparison with mainstream object detection models.

Model	Parameter (million)	Precision (%)	Recall (%)	mAP@0.5 (%)
Faster R-CNN	41.1	83.2	90.1	86.9
SSD	26.3	87.6	88.6	90.3
DEIM-n	3.6	89.3	90.2	90.5
YOLOv3	61.5	87.8	73.7	84.6
YOLOv5s	7.2	91.1	90.2	92.1
YOLOv8n	3.2	92.4	89.7	92.6
YOLOv11n	2.6	92.6	91.2	93.0
RT-DETR-r18	20.1	92.7	89.1	92.9
LCNET-FusionYOLO	6.2	94.1	91.8	94.1
PLFYNet	3.0	94.2	92.1	94.5

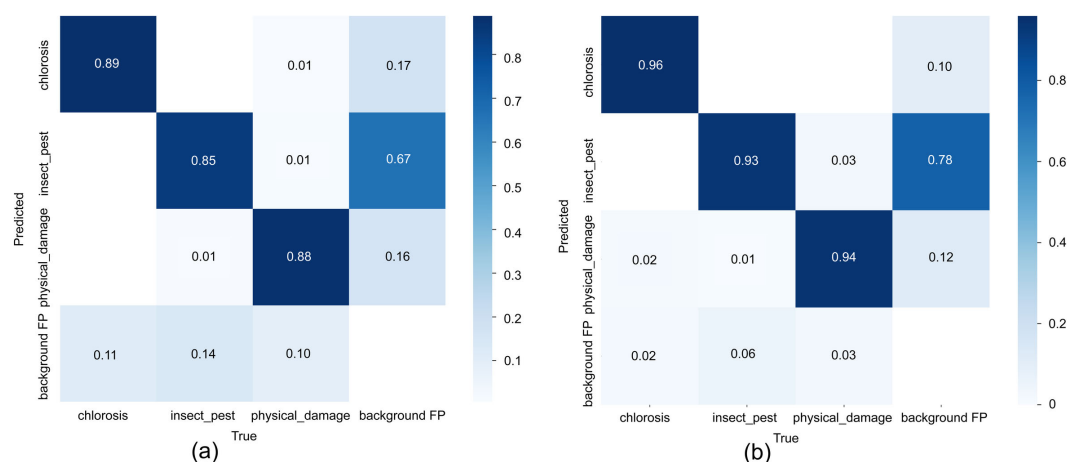


FIGURE 14 Confusion matrix : (a) Confusion matrix of YOLOv7-tiny; (b) Confusion matrix of PLFYNet.

autonomous power management, visual acquisition system, and wireless communication interface for field operations. Mobile deployment utilizes a ROS-enabled robotic platform manufactured by Helloblock, featuring compact form factor, modular architecture, and versatile hardware integration capabilities for dynamic agricultural monitoring applications.

3.5.3 comparison deployed on Jetson Orin Nano

For on-site power supply, we used a 12V/5A battery. During continuous inference, the model's average power consumption was 9.8W—17.6% lower than the baseline YOLOv7-tiny model (11.9W). To verify robustness, tests were conducted under extreme backlight, backlighting, or canopy occlusion conditions, with an average inference time of 14.6ms (9.5ms in the laboratory environment), which meets the <50ms threshold required for real-time accurate decision-making.

Table 10 presents the real-time detection speed comparison of the three models on Jetson Orin Nano. Figure 15 shows the specific detection results of YOLOv7-tiny (Figure 15a) and PLFYNet (Figure 15b) after final deployment, with the two subfigures comparing their display effects. They respectively demonstrate the detection performance of the three labels (insect pest, physical damage, and chlorosis), and clearly, PLFYNet achieves better detection accuracy.

4 Conclusions

Addressing the critical requirements for real-time Ginkgo biloba leaf disease detection in edge computing scenarios, this

study presents PLFYNet, a lightweight deep learning model that resolves the deployment challenges of high-precision disease detection systems on resource-constrained embedded devices while maintaining detection accuracy comparable to computationally intensive models.

This research makes three main contributions. First, we systematically evaluated five lightweight backbone architectures: MobileNetV3, GhostNet, ShuffleNetV2, PP-PicoDet, and PP-LCNet. We found that the reconstructed PP-LCNet is the optimal backbone for Ginkgo disease detection. Compared to YOLOv7-tiny, it reduces parameters by 26.8% while maintaining competitive accuracy. Second, we developed BiECaFusion to replace PANet. BiECaFusion includes 1×1 convolutions for channel alignment, ECA attention for dynamic weighting, and bidirectional feature interaction ($x_0 + w_1x_1$, $x_1 + w_0x_0$). This addresses the small target feature loss inherent in unidirectional fusion. We combined BiECaFusion with Bottleneck Transformer, Shape-IoU loss, and DyHead to create LcNet-FusionYOLO. This model achieves 94.1% mAP@0.5 with 6.21M parameters. Third, we used Layer-Adaptive Magnitude-based Pruning (LAMP) to further compress the model. The model's parameters were reduced to 3.0M (a 50.5% reduction), and its mAP@0.5 was improved to 94.5%. This shows that strategic pruning boosts both model efficiency and accuracy, ultimately resulting in the PLFYNet model.

Deployment on Jetson Orin Nano validated practical applicability: the model achieved 50.5 FPS inference speed (22.6% improvement over YOLOv7-tiny's 41.2 FPS) with 94.2% precision

TABLE 8 External validation results of the public dataset PlantDoc.

Models	Precision (%)	Recall	mAP@0.5
YOLOv7-tiny	51.8%	55.7%	52.4%
PLFYNet	56.1%	55.9%	55.7%

TABLE 9 Jetson Orin Nano hardware configuration.

Names	Specifications
CPU	6-core Arm® Cortex®-A78AE v8.2 64-bit CPU 1.5MB L2 + 4MB L3
GPU	NVIDIA Ampere architecture with 1024 CUDA cores and 32 tensor cores
Memory	8GB 128-bit LPDDR5

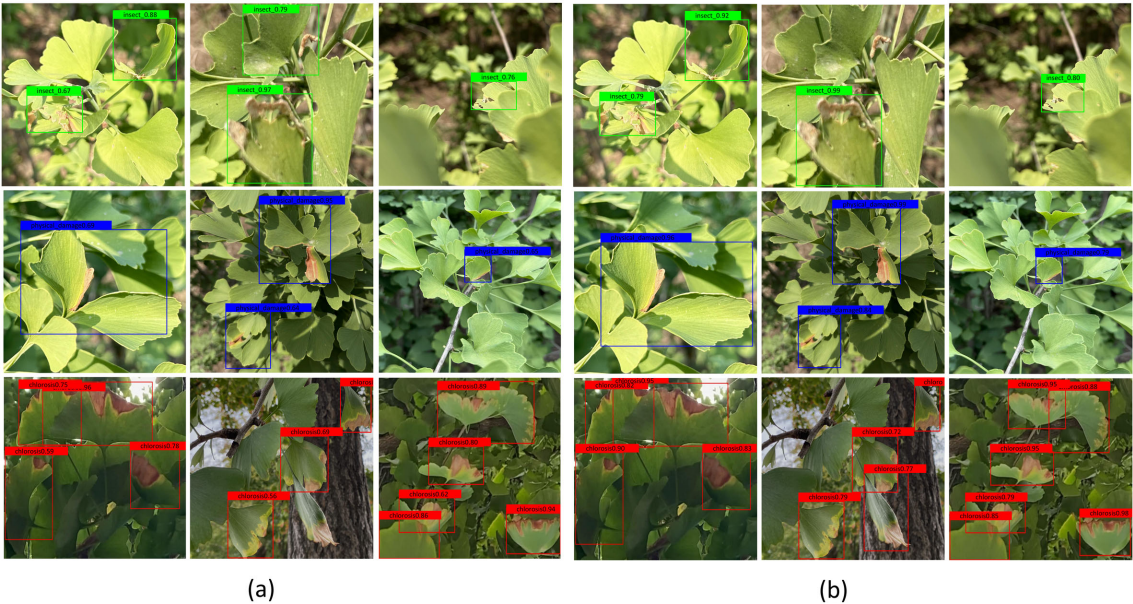


FIGURE 15 The detection results of the YOLOv7-tiny and PLFYNet on Jetson Orin Nano: (a) YOLOv7-tiny; (b) PLFYNet.

TABLE 10 Speed comparison of real-time detection speed of three models on Jetson Orin Nano.

Model	GFLOPS	Inference (ms)	NMS (ms)	FPS
YOLOv7-tiny	13.2	13.1	2.2	41.2
LCNET-FusionYOLO	13.6	13.5	2.6	40.3
PLFYNet	6.2	9.5	1.9	50.5

and 92.1% recall across three disease categories (chlorosis, insect pest, physical damage). Comparative analysis against eight mainstream detectors revealed superior mAP@0.5 performance using only 2.2%–21.8% of traditional two-stage detector parameters.

Current limitations include: (1) dataset geographical constraints to Jiangsu Province, China, potentially limiting generalization; (2) unexplored performance under extreme weather and illumination conditions; (3) pruning strategy requiring extensive fine-tuning, suggesting opportunities for more efficient compression methods.

Future work encompasses: This study lays a foundation for practically edge-deployable disease detection in precision agriculture and has broad implications for AI-driven sustainable farming practices. Future research will cover multiple dimensions: in terms of data, expand datasets across diverse geographical and environmental conditions, dynamically adjust augmentation strategies based on natural sample distribution, standardize annotation processes via cross-regional collaboration, introduce an augmentation-validation feedback loop, and verify annotation consistency with Kappa coefficient (target ≥ 0.90); in terms of model & hardware, explore knowledge distillation to reduce complexity without accuracy loss, develop adaptive pruning frameworks for

hardware-specific optimization, and integrate multi-spectral imaging to enhance disease characterization; in terms of deployment, develop an offline-first mode by integrating LoRa modules. These measures aim to fully improve technical implementation feasibility.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

JW: Conceptualization, Funding acquisition, Methodology, Resources, Visualization, Writing – review & editing. SG: Data curation, Formal analysis, Investigation, Resources, Software, Validation, Writing – original draft. MZ: Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This research was funded by the Jiangsu Forestry Science and Technology Innovation and Promotion Project (LYKJ (2024)05), the Nanjing Forestry University Postgraduate Joint Training Demonstration Base Program (2023204), and Jiangsu Province Graduate Workstation Program (2022103).

Acknowledgments

All data generated or presented in this study are available upon request from the first author. Furthermore, the models and code used during this study cannot be shared as the data also form part of an ongoing study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Ai, H., Zhu, X., Han, Y., Ma, S., Wang, Y., Ma, Y., et al. (2025). Extraction of levees from paddy fields based on the SE-CBAM UNet model and remote sensing images. *Remote Sensing*. 17, 1871–1871. doi: 10.3390/rs17111871
- Dai, X., Chen, Y., Xiao, B., Chen, D., Liu, M., Yuan, L., et al. (2021). “Dynamic head: unifying object detection heads with attentions,” in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (Piscataway, USA: Institute of Electrical and Electronics Engineers (IEEE)) 7369–7378. doi: 10.1109/CVPR46437.2021.00729
- Duan, Z., Xie, T., Wang, L., Chen, Y., and Wu, J. (2024). Microalgae detection based on improved YOLOv5. *IET Image Processing*. 18, 2602–2613. doi: 10.1049/ipr2.13119
- Fareed, K., Khan, A., Alhussein, M., Aurangzeb, K., Shahzad, A., and Islam, M. (2025). CBAM attention gate-based lightweight deep neural network model for improved retinal vessel segmentation. *Int. J. Imaging Syst. Technol.* 35, 1–21. doi: 10.1002/ima.70031
- Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition* (Piscataway, USA: Institute of Electrical and Electronics Engineers (IEEE)) 580–587. doi: 10.1109/cvpr.2014.81
- Gkoutakos, K., Ioannidis, K., Demestichas, K., Vrochidis, S., and Kompatsiaris, I. (2024). A comprehensive review of deep learning-based anomaly detection methods for precision agriculture. *IEEE Access*. 12, 197715–197733. doi: 10.1109/access.2024.3522248
- Gong, R., Fan, X., Cai, D., and Lu, Y. (2024). Sec-CLOCs: multimodal back-end fusion-based object detection algorithm in snowy scenes. *Sensors* 24, 7401. doi: 10.3390/s24227401
- Han, J., and Yang, Y. (2021). L-Net: lightweight and fast object detector-based ShuffleNetV2. *J. Real-Time Image Processing*. 18, 2527–2538. doi: 10.1007/s11554-021-01145-4
- Huang, T., Zhu, J., Liu, Y., and Tan, Y. (2023). UAV aerial image target detection based on BLUR-YOLO. *Remote Sens. Letters*. 14, 186–196. doi: 10.1080/2150704x.2023.2174385
- Jiang, F., Lu, Y., Chen, Y., Cai, D., and Li, G. (2020). Image recognition of four rice leaf diseases based on deep learning and support vector machine. *Comput. Electron. Agriculture*. 179, 105824. doi: 10.1016/j.compag.2020.105824
- Lee, J., Park, S., Mo, S., Ahn, S., and Shin, J. (2020). Layer-adaptive sparsity for the magnitude-based pruning. *arXiv [preprint]*. 1–19. doi: 10.48550/arXiv.2010.07611
- Lu, Y., Zhou, H., Wang, P., Wang, E., Li, G., and Yu, T. (2025a). IMobileTransformer: A fusionbased lightweight model for rice disease identification. *Eng. Appl. Artif. Intelligence*. 161, 112271. doi: 10.1016/j.engappai.2025.112271
- Lu, Y., Li, P., Wang, P., Li, T., and Li, G. (2025b). A method of rice yield prediction based on the QRBILSTM-MHSA network and hyperspectral image. *Comput. Electron. Agriculture*. 239, 110884. doi: 10.1016/j.compag.2025.110884
- Mirab-balou, M., Tong, X., and Chen, X. (2012). A new species of *scirtothrips* infesting *Ginkgo biloba* in eastern China. *J. Insect Science*. 12, 1–7. doi: 10.1673/031.012.11701
- Omer, S. M., Ghafoor, K. Z., and Askar, S. K. (2024). Lightweight improved yolov5 model for cucumber leaf disease and pest detection based on deep learning. *Signal Image Video Processing*. 18, 1329–1342. doi: 10.1007/s11760023028659
- Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intelligence*. 39, 1137–1149. doi: 10.1109/tpami.2016.2577031
- Shen, X., Wang, H., Wei, B., and Cao, J. (2023). Real-time scene classification of unmanned aerial vehicles remote sensing image based on Modified GhostNet. *PLoS One* 18, e0286873–e0286873. doi: 10.1371/journal.pone.0286873
- Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., and Batra, N. (2020). “PlantDoc: A dataset for visual plant disease detection,” in *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*. (New York, USA: Association for Computing Machinery (ACM)) 249–253. doi: 10.1145/3371158.3371196
- Srinivas, A., Lin, T.-Y., Parmar, N., Shlens, J., Abbeel, P., and Vaswani, A. (2021). “Bottleneck transformers for visual recognition,” in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (Nashville, TN, USA: Institute of Electrical and Electronics Engineers (IEEE)) 16514–16524. doi: 10.1109/CVPR46437.2021.01625
- Sun, Y., Cao, F., Wei, X., Welham, C., Chen, L., Pelz, D., et al. (2017). An ecologically based system for sustainable agroforestry in sub-tropical and tropical forests. *Forests* 8, 102. doi: 10.3390/f8040102

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2025.1679455/full#supplementary-material>

- Tang, L., Yan, Q., Xiang, X., Fang, L., and Ma, J. (2025). C2RF: bridging multimodal image registration and fusion via commonality mining and contrastive learning. *Int. J. Comput. Vision*. 133, 5262–5280. doi: 10.1007/s11263-025-02427-1
- Waghumbare, A., Singh, U., and Kasera, S. (2024). DIAT-DSCNN-ECA-Net: separable convolutional neural network-based classification of galaxy morphology. *Astrophysics Space Science*. 369, 38. doi: 10.1007/s10509-024-04302-w
- Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2023). “YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, , 17–24 JUN 2023. 7464–7475 (Vancouver, Canada: Institute of Electrical and Electronics Engineers (IEEE)). doi: 10.1109/cvpr52729.2023.00721
- Wang, Q., Hua, Y., Lou, Q., and Kan, X. (2025a). SWMD-YOLO: A lightweight model for tomato detection in greenhouse environments. *Agronomy* 15, 1593–1593. doi: 10.3390/agronomy15071593
- Wang, J., Guo, J., Wang, R., Zhang, Z., Fu, L., and Ye, Q. (2025b). Parameter disentanglement for diverse representations. *Big Data Min. Analytics*. 8, 606–623. doi: 10.26599/bdms.2024.9020087
- Wang, H., Shi, M., Cao, F., and Su, E. (2022). Ginkgo biloba seed exocarp: A waste resource with abundant active substances and other components for potential applications. *Food Res. Int.* 160, 111637. doi: 10.1016/j.foodres.2022.111637
- Xie, X., Xie, L., Li, G., Guo, H., Zhang, W., Shao, F., et al. (2024). Discriminative features pyramid network for medical image segmentation. *Biocybernetics Biomed. Engineering*. 44, 327–340. doi: 10.1016/j.bbe.2024.04.001
- Xie, X., Xie, L., Pan, X., Wang, Y., Yu, Z., Shao, F., et al. (2025a). Local and long-range progressive fusion network for knee joint segmentation. *Biomed. Signal Process. Control* 112, 108624. doi: 10.1016/j.bspc.2025.108624
- Xie, X., Xie, L., Pan, X., Shao, F., Zhao, W., and An, J. (2025b). PIF-Net: A parallel interweave fusion network for knee joint segmentation. *Biomed. Signal Process. Control*. 109, 107967. doi: 10.1016/j.bspc.2025.107967
- Xu, H., Xue, R., Zhang, Z., and Hua, S. (2024). MGDE-YOLO: an improved lightweight algorithm for personnel departure detection based on YOLOv7. *IEEE Access*. 12, 150592–150603. doi: 10.1109/access.2024.3480040
- Xue, R., and Wang, L. (2025). Research on lightweight citrus leaf pest and disease detection based on PEW-YOLO. *Processes* 13, 1365–1365. doi: 10.3390/pr13051365
- Yang, N., Li, G., Wang, S., Wei, Z., Ren, H., Zhang, X., et al. (2025). SS-YOLO: A lightweight deep learning model focused on side-scan sonar target detection. *J. Mar. Sci. Engineering*. 13, 66–66. doi: 10.3390/jmse13010066
- Yu, G., and Zhou, X. (2023). An improved YOLOv5 crack detection method combined with a bottleneck transformer. *Mathematics* 11, 2377–2377. doi: 10.3390/math11102377
- Yuan, Z., Wang, S., Wang, C., Zong, Z., Zhang, C., Su, L., et al. (2025). Research on calf behavior recognition based on improved lightweight YOLOv8 in farming scenarios. *Animals* 15, 898. doi: 10.3390/ani15060898
- Zhang, L., Fang, X., Sun, J., Su, E., Cao, F., and Zhao, L. (2023a). Study on synergistic anti-inflammatory effect of typical functional components of extracts of ginkgo biloba leaves. *Molecules* 28, 1377–1377. doi: 10.3390/molecules28031377
- Zhang, Y., Zhang, J., Tian, L., Huang, Y., and Shao, C. (2023b). The ginkgo biloba L. in China: current distribution and possible future habitat. *Forests* 14, 2284–2284. doi: 10.3390/f14122284
- Zhang, J., Zhang, H., Liu, B., Qu, G., Wang, F., Zhang, H., et al. (2023c). Small object intelligent detection method based on adaptive recursive feature pyramid. *Heliyon* 9, e17730. doi: 10.1016/j.heliyon.2023.e17730
- Zhang, S., Wang, Q., and Gao, S. (2024). Pharmacokinetics and pharmacodynamics provide insights into the possible action mechanism of ginkgo flavonoids in cardiovascular disease treatment. *Int. J. Pharmacol.* 20, 926–935. doi: 10.3923/ijp.2024.926.935
- Zhao, Z., Ma, X., Shi, Y., and Yang, X. (2024). Multi-scale defect detection for plaid fabrics using scale sequence feature fusion and triple encoding. *Visual Computer*. 41, 5205–5221. doi: 10.1007/s00371-024-03716-5
- Zhao, K., Zhao, L., Zhao, Y., and Deng, H. (2023). Study on lightweight model of maize seedling object detection based on YOLOv7. *Appl. Sci.* 13, 7731–7731. doi: 10.3390/app13137731
- Zhu, S., Ma, W., Wang, J., Yang, M., Wang, Y., and Wang, C. (2023). EADD-YOLO: An efficient and accurate disease detector for apple leaf using improved lightweight YOLOv5. *Front. Plant science*. 14. doi: 10.3389/fpls.2023.1120724