Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Lightweight YOLOv11n-OBB-Based Method for Oriented Detection of Luosifen Outer Packages

Guojian Liang

liangguojian@lzpu.edu.cn

Liuzhou Polytechnic University

Shuwen Zhou

Liuzhou Polytechnic University

Yinhai Li

Liuzhou Polytechnic University

Article

Keywords:

Posted Date: July 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10112780/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Lightweight YOLOv11n-OBB-Based Method for Oriented Detection of Luosifen Outer Packages

Guojian Liang[1,*], **Shuwen Zhou**[2], **and Yinhai Li**[2]

[1]School of Mechanical and Electrical Engineering, Liuzhou Polytechnic University, Liuzhou 545000, China
[2]School of Electronic Information Engineering, Liuzhou Polytechnic University Liuzhou 545000, China
[*]liangguojian@lzpu.edu.cn

ABSTRACT

The automated boxing of Luosifen (river-snail rice noodle) outer packages requires simultaneous multi-category detection, rotation-angle estimation, and real-time inference under edge-computing constraints. However, existing methods face three major limitations: conventional template-matching approaches are poorly suited to flexible production scenarios involving multiple object poses; horizontal bounding box detectors are inherently unable to estimate object rotation angles; and standard oriented bounding box (OBB) models remain difficult to deploy under edge-computing constraints due to excessive redundancy in their backbone networks. To address these challenges, we constructed an oriented bounding box dataset covering four categories of river-snail rice noodle outer packages, containing 1,603 images, which were divided into training, validation, and test sets at a ratio of 7:2:1, comprising 1,122, 320, and 161 images, respectively. We further propose a lightweight oriented object detection method based on a key-layer sensitivity-guided replacement strategy, whose core components include GhostConv_OBB, a lightweight feature extraction unit integrating dual-path angle-aware enhancement (AAE) and progressive efficient channel attention (PECA); C3k2_GhostConv_OBB, a multi-scale feature fusion module; and the optimal deployment of the Neck-to-Head transition layer, specifically detection head 11, determined through key-layer sensitivity analysis. Experimental results show that the proposed model achieves 97.79% mAP@0.5 and 93.95% mAP@0.5:0.95 on the test set with only 2.95M parameters and 3.45 GFLOPs, while maintaining an inference latency of 67.73 ms (14.76 FPS). Compared with the YOLOv11n-OBB baseline, the proposed model improves mAP@0.5:0.95 by 32.27 percentage points and mAP@0.5 by 4.26 percentage points. These results demonstrate that the proposed method achieves localization accuracy comparable to medium-scale models under an extremely low parameter budget, and further reveal the complementary synergy between AAE and PECA in improving recall and localization precision, providing a lightweight design paradigm for edge-oriented rotated object detection in industrial scenarios. The dataset and improved modules are publicly available at https://github.com/TPNS-007/luosifen-obb-detection.

Introduction

The detection of Luosifen outer packages is a critical step in automated boxing production lines, where the localization accuracy of packaging bags directly determines the success rate of robotic grasping. Since the terminal stage of the production line is deployed on edge computing devices, the detection algorithm must achieve high-precision real-time detection under limited computational resources[1,2]. However, although existing oriented bounding box detection models can describe rotational poses, they commonly suffer from redundant feature extraction networks and high computational overhead in angle prediction branches, making them difficult to adapt to edge-computing constraints. Meanwhile, rotation, tilting, and complex illumination variations during conveyor-belt transportation can easily cause localization deviations, resulting in missed or insufficient grasping[3]. Therefore, lightweight and high-precision oriented detection methods for outer packages that are suitable for edge deployment urgently require further investigation.

Current research on rotated object detection mainly focuses on two aspects: improvements to horizontal bounding box detection and lightweight designs for oriented bounding box detection. Existing horizontal bounding box detection methods[4–6] improve general detection accuracy by modifying YOLO-series networks. However, horizontal bounding boxes cannot accurately conform to the contours of tilted packaging bags and tend to introduce substantial background interference, leading to deviations in the calculation of robotic grasping points. For rotated objects, Yao et al.[7] systematically analyzed the representational ambiguity of oriented bounding boxes and proposed an improved encoding strategy. Li et al.[8] and Huang et al.[9] embedded rotated detection modules into lightweight backbones to reduce the number of parameters, enabling real-time detection in dense scenes and remote-sensing images. Nevertheless, existing OBB methods exhibit an inherent trade-off between representational accuracy and computational efficiency. Although the representation optimization proposed by Yao et al.[7] alleviates angular ambiguity, its two-stage regression and rotated-box refinement introduce additional computational

overhead. The multi-scale attention fusion module proposed by Li et al.[8] improves detection accuracy, but still incurs relatively high computational cost. Huang et al.[9] reduced the number of parameters using depthwise separable convolution, but their method still faces challenges in preserving fine-grained spatial information and maintaining stable rotation-pose estimation. Other studies[10–12] have introduced channel attention mechanisms to enhance feature discrimination under complex industrial backgrounds. However, existing SE or ECA modules either introduce a large number of parameters or rely on a single receptive field, making it difficult to achieve fine-grained foreground–background separation on resource-constrained devices. In addition, most existing lightweight modules adopt fixed residual connections, which cannot adaptively adjust the fusion ratio according to feature importance. They also lack fault-tolerant mechanisms for numerical anomalies on edge-oriented deployment, resulting in limited stability during continuous industrial operation.

Based on the above considerations, this study proposes an oriented detection method for Luosifen outer packages tailored to edge deployment. To address the high computational complexity of existing OBB models, a lightweight feature extraction network based on GhostConv_OBB and C3k2_GhostConv_OBB is constructed. By expanding the receptive field through cost-efficient operations, the proposed network reduces the inference overhead on edge-oriented deployment. To overcome insufficient feature extraction for rotated objects, a dual-path angle-aware enhancement module is designed to integrate global semantic information with local spatial geometric cues, thereby explicitly strengthening the representation of package edges and corner features. To mitigate complex background interference and the excessive complexity of conventional attention mechanisms, a progressive efficient channel attention module is proposed, which enables progressive learning of channel weights through a lightweight two-stage convolutional structure and finely suppresses background noise. Furthermore, to address the rigidity of residual fusion and insufficient numerical stability, learnable residual fusion coefficients and a safe fallback mechanism are introduced, improving the robustness of the model on edge-oriented deployment.

The main contributions of this study can be summarized as follows:

1. A lightweight oriented detection architecture tailored to edge deployment is proposed. By constructing the GhostConv_OBB unit and the C3k2_GhostConv_OBB multi-scale fusion module, and by introducing a key-layer sensitivity-guided replacement strategy, the optimal deployment position is identified as the Neck-to-Head transition layer through key-layer sensitivity analysis. This design substantially reduces the number of parameters and computational cost while preserving the macro-level geometric information required for rotated object detection.

2. A dual-path feature enhancement mechanism is proposed to improve the stability of rotation-angle estimation. A collaborative dual-branch enhancement scheme is constructed by combining angle-aware enhancement (AAE) and progressive efficient channel attention (PECA). Specifically, the AAE module integrates global semantic context with local spatial geometric cues, explicitly strengthening feature responses along edges and at corner regions. In synergy with AAE, the PECA mechanism performs progressive learning and adaptive recalibration of channel weights through lightweight two-stage convolution, thereby enabling fine-grained suppression of background noise.

3. A numerically stable residual learning scheme is proposed to ensure robust and continuous operation on edge-oriented deployment. The rigid identity mapping is replaced with learnable residual fusion coefficients, allowing the network to dynamically balance the contributions of skip connections and convolutional features. In addition, the embedded safe fallback mechanism automatically restores the identity mapping when non-finite numerical anomalies occur, thereby improving fault tolerance during long-term edge inference.

Related Work

In automated Luosifen boxing production lines, multiple categories of outer packages are transported by conveyor belts into the visual inspection area and subsequently grasped by robotic manipulators. During transportation, packaging bags are prone to rotation, tilting, stacking-induced occlusion, and extrusion deformation. Therefore, the detection algorithm must satisfy two requirements simultaneously: accurately estimating the rotation angle to provide the grasping pose, and maintaining real-time inference capability on edge computing devices.

Early production lines commonly adopted template matching for single-category localization. This method performs sliding-window similarity comparison using predefined templates and is easy to implement in scenarios with fixed object poses[13,14]. However, when facing multi-category and multi-pose requirements, a template library containing different rotation angles and deformation states must be established for each target category, followed by full-image convolution and back-end fusion for each template. This not only causes the computational cost to increase linearly with the number of templates, but also makes the method highly sensitive to wrinkles, specular reflection, and occlusion. Moreover, template matching cannot directly output rotation angles, making it difficult to meet the requirements of flexible production[15].

With the development of deep learning, horizontal bounding box detectors, such as YOLOv5 and YOLOv8, have been introduced into industrial vision applications. These methods automatically learn semantic features in a data-driven manner,

eliminating the need to maintain template libraries and enabling parallel multi-category detection. However, in the present task, horizontal bounding boxes include excessive background regions for tilted packaging bags, incorporating a large amount of background interference. This not only increases unnecessary computation, but also leads to insufficient localization accuracy. Moreover, horizontal bounding box detectors cannot estimate rotation angles, which directly constrains the accurate calculation of robotic grasping poses[16].

To overcome the geometric limitations of horizontal bounding boxes, rotated object detection based on oriented bounding boxes (OBBs) has attracted increasing research attention in recent years. Models such as YOLOv8-OBB introduce an angle prediction branch, enabling the detection boxes to better conform to the actual contours of targets and substantially reducing background interference. Nevertheless, angle learning in OBB detection is more challenging than that in horizontal bounding box detection. Issues such as boundary discontinuity and angular periodicity can easily lead to unstable training[17]. In addition, to ensure angle estimation accuracy, backbone networks are often designed with relatively high complexity, resulting in real-time inference bottlenecks when these models are directly deployed on edge-oriented deployment in production lines[18].

To address the high complexity of OBB models and the difficulty of edge deployment, recent studies have shifted toward lightweight OBB design. Shan et al.[19] used YOLOv8n-OBB as the baseline and introduced a large-kernel separable attention mechanism together with a dual-path feature refinement module, thereby enhancing the global perception capability for rotated features while maintaining a lightweight architecture. Kuang et al.[20] adopted YOLOv11n-OBB as the baseline and achieved high-frame-rate inference under substantially reduced parameter and computational costs through a lightweight direction-aware detection head, adaptive downsampling, and a multi-scale adaptive feature pyramid. These studies indicate that model compression within the OBB framework while maintaining detection accuracy has become an important direction in industrial detection.

However, most of the above studies focus on remote-sensing or general-purpose scenarios. For specific industrial scenarios such as Luosifen outer-package detection, where production-line illumination is complex and strict edge-computing constraints must be satisfied, existing lightweight OBB modules still exhibit several limitations. First, angle prediction branches do not sufficiently exploit edge and corner features of packaging bags. Second, channel attention mechanisms struggle to balance lightweight design with fine-grained suppression of background interference. Third, they lack fault-tolerant designs for numerical stability on edge-oriented deployment[21,22]. Therefore, this study adopts YOLOv11n-OBB as the baseline and designs corresponding enhancement modules for three core issues: rotation-angle estimation, complex-background suppression, and lightweight edge deployment. Ablation experiments are further conducted to achieve accurate and stable oriented detection of Luosifen outer packages under limited computational resources.

YOLOv11n-OBB Model

Dataset Construction and Preprocessing

To ensure sample diversity, four commercially available Luosifen outer packages, denoted as A–D, were selected as the research objects in this study. Image acquisition was conducted in a simulated conveyor-belt environment of an automated boxing machine, covering two operating conditions: a static state, which was used to obtain clear reference images without motion blur, and a dynamic running state, which was used to reproduce the motion blur and vibration characteristics of real production lines. Images were captured by invoking the SDK provided by the industrial camera manufacturer through Python scripts. Each package category was sampled 60–70 times, yielding a total of 266 original images.

Considering the complex disturbances that occur in actual production lines, including rotation, tilting, illumination variation, occlusion, and extrusion deformation, the diversity of the original samples was insufficient to support the generalization requirements of the model. Therefore, an OpenCV-Python-based data augmentation strategy was adopted, including rotation, perspective transformation, illumination perturbation, simulated occlusion, and scale resizing [23]. As a result, the dataset was expanded to 1,603 images, with approximately 400 images for each category.

The dataset was divided into training, validation, and test sets at a ratio of 7:2:1, consisting of 1,122 training images, 320 validation images, and 161 test images, respectively. For annotation, oriented bounding boxes (OBBs) were manually labeled using a parametric format, with the rotation annotations designed to cover arbitrary in-plane orientations of the targets[24].

Improved YOLOv11n-OBB Detection Model

In this study, YOLOv11-OBB was selected as the baseline model. As a recent oriented detection variant of the YOLO series, YOLOv11-OBB inherits the efficient inference capability of the one-stage architecture and enables end-to-end rotated object detection by introducing an angle prediction branch. Compared with conventional horizontal bounding box detectors, YOLOv11-OBB can output oriented bounding boxes that closely conform to the actual contours of targets, thereby effectively suppressing background redundancy. This characteristic has demonstrated good applicability in remote-sensing and industrial detection scenarios.

However, the standard YOLOv11-OBB model is designed for general-purpose scenarios, and its direct transfer to Luosifen outer-package detection faces two major challenges. First, as the most lightweight variant, YOLOv11n-OBB contains only 2.66 M parameters, with a computational cost of 3.36 GFLOPs. Its measured latency in the experimental environment is 65.88 ms, corresponding to 15.18 FPS. This only marginally satisfies the production-line cycle requirement for robotic grasping synchronization, namely $\geq$14 FPS, or $\leq$71 ms per frame, leaving almost no margin for system jitter, preprocessing overhead, or concurrent edge-side workloads[25,26]. Therefore, improving accuracy through conventional convolution stacking would immediately exceed the strict computational budget of edge-oriented deployment. Second, the existing angle prediction branch does not sufficiently exploit fine-grained geometric features such as packaging-bag edges and corners. Under complex operating conditions, including tilting caused by conveyor-belt vibration, stacking-induced occlusion, and extrusion deformation, angle regression errors and localization drift are likely to occur. As a result, the strict localization metric mAP@0.5:0.95 reaches only 61.68%[27]. Therefore, this study adopts YOLOv11n-OBB as the baseline and improves the network around three core objectives: precise lightweight optimization to recover real-time inference margins, angle-aware enhancement to stabilize rotation estimation, and progressive background suppression to improve contour fidelity[28,29]. The overall architecture of the proposed model is shown in Fig.1.

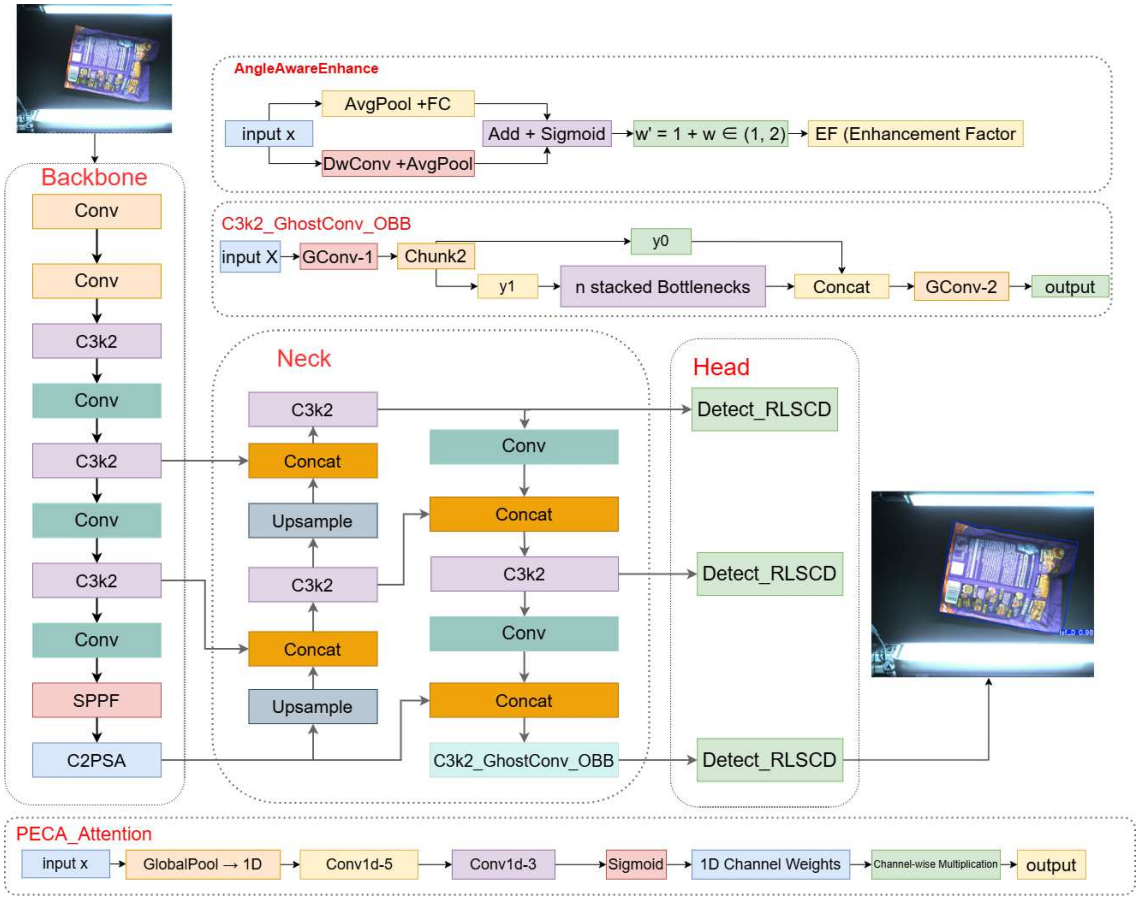


Figure 1. Improved network architecture diagram.

First, to construct a lightweight backbone while improving rotated object detection performance, standard convolutions are replaced with the improved GhostConv_OBB module. Built upon the redundant feature generation mechanism of GhostConv through cost-efficient operations, this module embeds dual-path angle-aware enhancement and progressive efficient channel attention. It effectively improves the contour-fitting accuracy of the model for rotated and tilted packaging bags, while substantially reducing the computational overhead on edge-oriented deployment through adaptive residual fusion[30,31].

Second, to address the high computational cost of multi-scale fusion modules in the Neck and the redundancy of cross-layer information aggregation, the original C3k2 module is replaced with C3k2_GhostConv_OBB. Owing to the lightweight characteristics of GhostConv_OBB, the proposed module significantly reduces the number of parameters during feature extraction and concatenation. Meanwhile, it preserves the cross-layer shortcut mechanism of the Bottleneck structure, maintaining feature flow for small and occluded targets. This further improves the feature fusion efficiency of the model and its

adaptability to edge deployment[32,33].

Finally, considering that fixed residual connections cannot adaptively adjust fusion ratios during training and that long-term operation on edge-oriented deployment may involve the risk of numerical overflow, learnable residual coefficients and a safe fallback mechanism are introduced to replace the original identity mapping[34]. This mechanism enables the network to dynamically balance the contributions of identity mapping and convolutional features according to the feature importance of rotated targets. From the perspective of numerical stability, it ensures continuous and reliable operation of the model in edge environments with limited computational resources, thereby enhancing its interference resistance under complex production-line conditions.

GhostConv_OBB Lightweight Adaptive Feature Extraction Unit

In the Luosifen outer-package detection task, backbone networks constructed with standard convolutions usually involve a large number of parameters and high computational cost, making it difficult to satisfy the real-time requirements of production lines when directly deployed on edge-oriented deployment. In addition, residual connections with fixed weights cannot adaptively adjust the fusion ratio according to feature importance, while numerical anomalies such as overflow may occur during long-term low-precision inference on edge-oriented deployment. To address these issues, this study proposes the GhostConv_OBB unit as the fundamental building block of the backbone network. Its structure is shown in Fig.2.

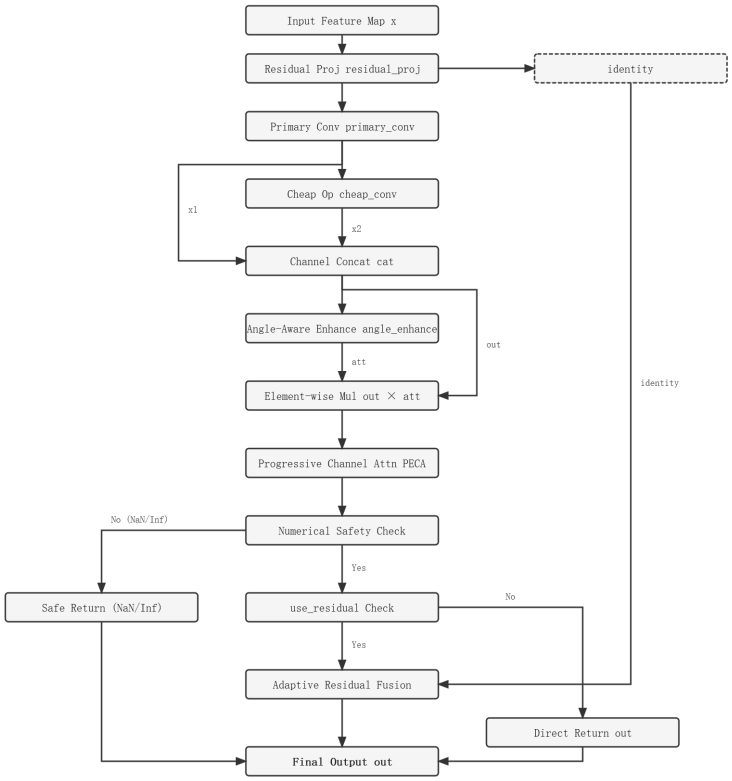


Figure 2. Architecture of the proposed GhostConv_OBB module.

The proposed module is built upon the low-cost feature generation mechanism of GhostConv. Given an input feature map ($\mathbf{x} \in \mathbb{R}^{C_{in} \times H \times W}$), the primary convolution branch first generates intrinsic feature maps:

$$\mathbf{x}_{\mathrm{p}} = \mathrm{Conv}_{k \times k}(\mathbf{x}) \in \mathbb{R}^{C' \times H' \times W'} \tag{1}$$

Subsequently, a lightweight (1×1) convolution branch is applied to the intrinsic feature maps to generate supplementary ghost features:

$$\mathbf{x}_{\mathrm{c}} = \mathrm{Conv}_{1 \times 1}(\mathbf{x}_{\mathrm{p}}) \in \mathbb{R}^{(C_{\mathrm{out}} - C') \times H' \times W'} \tag{2}$$

After concatenation along the channel dimension, the base feature representation is obtained as ($\mathbf{f} = [\mathbf{x}_p; \mathbf{x}_c] \in \mathbb{R}^{C_{out} \times H' \times W'}$), where $C' = eC_{out}$, and e denotes the channel allocation ratio of the primary convolution branch, with stride set to $s = 1$. Compared with the computational cost of standard convolution, i.e., $C_{in}C_{out}K^2HW$, the computational cost of GhostConv_OBB is approximately $C_{in}C'K^2HW + C'(C_{out} - C')H'W'$. When $e = 0.5$, the primary convolution branch generates only half of the output channels, while the remaining channels are produced by the low-cost (1×1) convolution branch, thereby effectively reducing the inference overhead on edge-oriented deployment.

After obtaining the basic feature representation, the module further processes it through embedded feature enhancement components to obtain a refined version of the feature. To avoid the rigidity caused by fixed identity-related residual connections, a learnable residual fusion coefficient λ is introduced, with an initial value of 0.1. The residual fusion process is defined as follows:

$$\mathbf{y} = (1 - \lambda) \cdot \mathbf{x}_{res} + \lambda \cdot \mathbf{f}_{enh} \tag{3}$$

where $\mathbf{x}_{res}$ denotes the residual path after projection, which is used to match the dimensionality of the enhanced feature $\mathbf{f}_{enh}$. This design enables the network to adaptively balance the contributions of residual information and enhanced convolutional features according to feature importance. Meanwhile, a numerical safety check is embedded into the module. When non-finite values occur in $\mathbf{f}_{enh}$, the output automatically falls back to $\mathbf{y} = \mathbf{x}_{res}$, thereby ensuring continuous and stable operation on edge-oriented deployment.

In addition, GhostConv_OBB reserves a low-cost feature enhancement interface. The base feature ($\mathbf{f}$) is subsequently modulated by local–global edge enhancement and progressive channel attention, improving the discriminability of rotated-object features without substantially increasing the computational burden.

Local–Global Edge Enhancement and Progressive Channel Attention

Although lightweight convolution reduces computational cost, its feature representation capability remains limited. Luosifen outer packages are typically made of aluminum–plastic laminated films. During conveyor-belt transportation, local geometric disturbances, including wrinkles, specular reflections, and extrusion deformation, frequently occur along the bag edges. These disturbances can distort the apparent contour orientation, causing boundary ambiguity and corner drift in the OBB angle regression branch. Without explicit edge enhancement, the model may confuse true package boundaries with background clutter, resulting in unstable rotation estimation under strict localization criteria[35,36].

To address this issue, two low-cost enhancement modules are embedded into GhostConv_OBB: dual-path angle-aware enhancement (AAE) and progressive efficient channel attention (PECA)[37,38]. AAE explicitly strengthens geometric feature responses along package edges and at corner regions, whereas PECA progressively suppresses background noise in the channel dimension. These two components work collaboratively to restore contour visibility and, more importantly, to correct subtle misalignments between predicted boxes and ground-truth oriented boxes under high-overlap thresholds[39].

It should be noted that PECA targets a major source of error in this industrial task: localization inaccuracy under high-IoU conditions, rather than missed detections. Through a lightweight two-stage convolution, PECA progressively learns channel weights and recalibrates feature responses contaminated by complex production-line illumination and conveyor-belt textures. As a result, the predicted OBBs can fit the true package contours more tightly. As confirmed by the ablation experiments presented later, PECA contributes substantially to the improvement in the strict localization metric mAP@0.5:0.95, indicating that its primary role is to reduce localization drift under stringent IoU criteria rather than merely improving detection recall[40,41].

Dual-Path Edge Enhancement (AngleAwareEnhance)

The proposed module is designed to strengthen the responses of package contour and corner features, thereby providing a more reliable geometric prior for angle estimation rather than directly encoding the rotation angle. Its structure is shown in Fig.3. The module adopts a complementary global–local dual-path architecture.

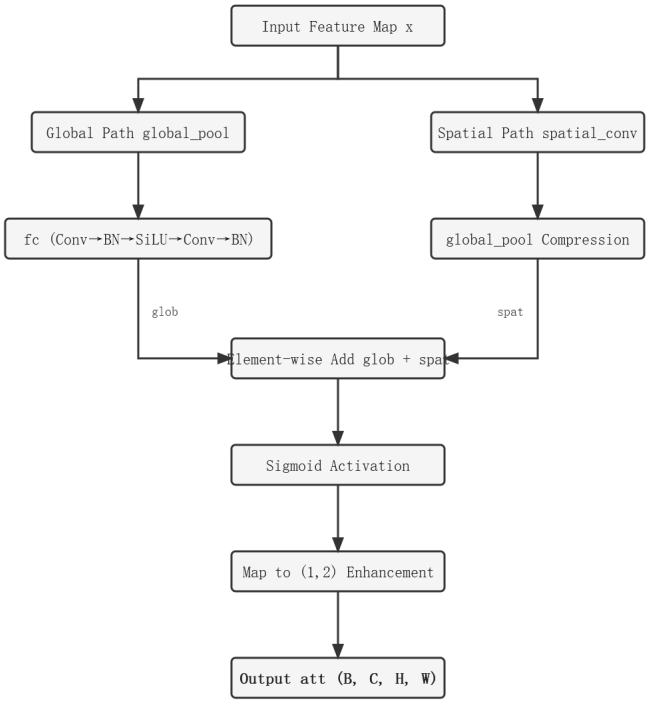


Figure 3. Dual-path edge enhancement module.

In the global path, global average pooling is first used to compress the spatial dimensions, and the resulting feature is passed through $(FC_{1\times1})$, which denotes two successive (1×1) convolutional layers, to learn high-level semantic weights:

$$g = \mathrm{FC}_{1\times1}\big(\mathrm{GAP}(f)\big) \in \mathbb{R}^{C_{\mathrm{out}}\times1\times1} \tag{4}$$

In the local path, depthwise separable convolution with a (3×3) kernel and groups $= C_{\mathrm{out}}$ is used to capture spatial geometric information. Global average pooling is then applied to compress the feature into channel weights.

$$s = \mathrm{GAP}\big(\mathrm{DWConv}_{3\times3}(f)\big) \in \mathbb{R}^{C_{\mathrm{out}}\times1\times1} \tag{5}$$

After fusing the two paths, Sigmoid normalization is applied to map the attention coefficients to the range of $(1,2)$, thereby achieving explicit feature enhancement:

$$w = \sigma(g+s),\ f_{\mathrm{edge}} = f\odot(1+w) \tag{6}$$

where $\odot$ denotes element-wise multiplication. This mechanism preferentially amplifies feature responses in edge and contour regions, thereby indirectly improving the stability of subsequent angle regression.

Progressive Efficient Channel Attention (PECA)

Complex production-line illumination and conveyor-belt backgrounds introduce persistent interference. Conventional SE modules rely on fully connected layers and therefore introduce a large number of parameters. Although ECA replaces fully connected layers with lightweight one-dimensional convolution, its single-stage fixed-kernel convolution struggles to capture both coarse-grained and fine-grained channel dependencies. To address this limitation, PECA is proposed in this study. It removes fully connected layers and cascades two one-dimensional convolutions with kernel sizes of 5 and 3, enabling progressive learning of channel weights. The main workflow of PECA is illustrated in Fig.4

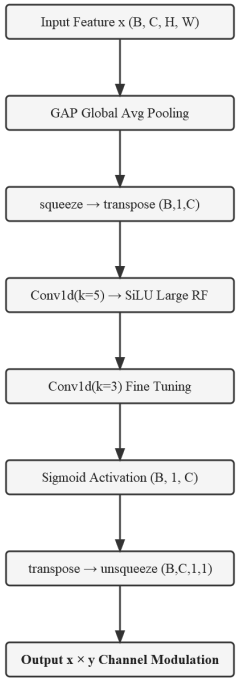


Figure 4. Workflow of the PECA module.

Given the input feature $\mathbf{f}_{edge}$, global average pooling is first applied to obtain the channel descriptor $\mathbf{z} \in \mathbb{R}^{1\times C_{\text{out}}}$. The descriptor is then progressively processed by two cascaded one-dimensional convolutions:

$$y_1 = \text{SiLU}\left(W_{k=5} * z\right) \tag{7}$$

$$y_2 = W_{k=3} * y_1 \tag{8}$$

where $*$ denotes one-dimensional convolution along the channel dimension. $\mathbf{W}_{k=5}$ is responsible for coarse-grained channel dependency modeling, whereas $\mathbf{W}_{k=3}$ performs fine-grained channel refinement. The channel attention weights are then generated via the Sigmoid function:

$$\alpha = \sigma(\mathbf{y}_2) \in \mathbb{R}^{1\times C_{\text{out}}} \tag{9}$$

The output feature is obtained by recalibrating $\mathbf{f}_{edge}$ with the broadcasted channel weights:

$$\mathbf{f}_{att} = \mathbf{f}_{edge} \odot \hat{\alpha} \tag{10}$$

where $\hat{\alpha} \in \mathbb{R}^{C_{\text{out}}\times 1\times 1}$ denotes the reshaped channel attention weights broadcast over the spatial dimensions, and $\odot$ represents element-wise multiplication. Compared with single-stage ECA, PECA progressively expands the effective receptive field for channel interaction and enables more refined inter-channel dependency modeling with only negligible additional parameters. Consequently, it more effectively suppresses complex background noise in production-line scenarios while satisfying the lightweight constraints of edge-oriented deployment.

C3k2_GhostConv_OBB **Lightweight Multi-Scale Fusion Module**

The C3k2 module in the Neck of YOLO-based detectors is responsible for cross-layer multi-scale feature aggregation. However, its input and output convolutions introduce considerable computational cost, and direct stacking may lead to additional computational bottlenecks. To address this issue, the standard convolutions in C3k2 are replaced with GhostConv_OBB, forming the C3k2_GhostConv_OBB module for lightweight feature fusion throughout the Neck. The workflow of this module is shown in Fig.5.

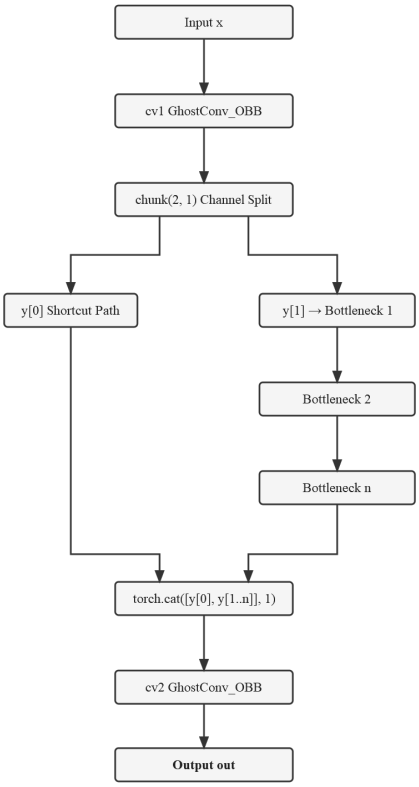


Figure 5. Workflow of the C3k2_GhostConv_OBB module.

Specifically, given an input feature map $\mathbf{x}$, channel expansion and splitting are first performed through GhostConv_OBB:

$$\mathbf{y}_0, \mathbf{y}_1 = \mathrm{Split}\big(\mathrm{GhostConv_OBB}_{\mathrm{in}}(\mathbf{x})\big) \tag{11}$$

where $\mathbf{y}_0$ directly participates in the final concatenation, whereas $\mathbf{y}_1$ sequentially passes through n Bottleneck modules for deep feature extraction:

$$\mathbf{y}_{k+1} = \mathrm{Bottleneck}(\mathbf{y}_k), \quad k = 1, \ldots, n \tag{12}$$

Finally, features from different stages are concatenated along the channel dimension and fused by the output GhostConv_OBB convolution:

$$\mathbf{y}_{\mathrm{out}} = \mathrm{GhostConv_OBB}_{\mathrm{out}}\big(\mathrm{Concat}(\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{n+1})\big) \tag{13}$$

This design preserves cross-layer shortcut connections and multi-scale information aggregation while substantially reducing the parameter count and computational cost of the Neck. It also helps prevent the degradation of features from small and occluded targets during feature compression.

GhostBottleneck_OBB Deep Residual Propagation Unit

To further increase the network depth while maintaining training stability, this study constructs a GhostBottleneck_OBB residual unit based on GhostConv_OBB. The unit forms a bottleneck structure by stacking two GhostConv_OBB subunits, as shown in Eq. (14), and introduces a learnable residual scaling coefficient γ, which is initialized to 0.1. The residual propagation process is defined as follows:

$$y = x + \gamma \cdot \mathrm{GhostConv_OBB}_2\big(\mathrm{GhostConv_OBB}_1(x)\big) \tag{14}$$

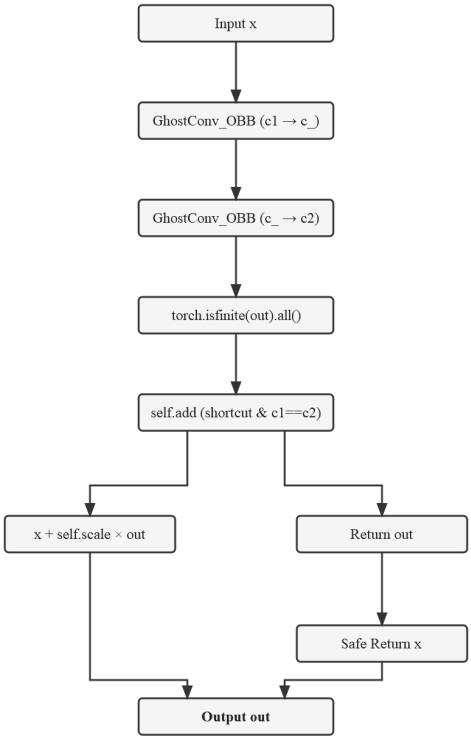


Figure 6. GhostBottleneck_OBB Residual Unit Flowchart

When numerical anomalies, such as NaN or Inf values, occur in deep features, the unit automatically falls back to the identity mapping. This mechanism integrates lightweight feature extraction, feature enhancement, and numerical safety into each network layer, thereby ensuring stable inference during long-term operation on edge-oriented deployment. Eq. (14) holds when $c_1 = c_2$ and the shortcut connection is enabled. If the feature dimensions do not match, a projection operation should be introduced to align the residual path before feature fusion.

Experiments and Results Analysis

Experimental Environment

All experiments are conducted in the environment shown in Table 1.

Operating System	**Ubuntu 24.04 LTS**
Hardware Platform	GIGABYTE B760M GAMING AC Platform
GPU	NVIDIA GeForce RTX 4090 D
VRAM	24GiB
GPU Driver Version	591.86
CUDA Toolkit Version	CUDA 11.8
PyTorch Version	PyTorch 2.5.0+cu118
torchvision Version	torchvision 0.20.0+cu118
torchaudio Version	torchaudio 2.5.0+cu118
Python Version	Python 3.11.14

Table 1. Experimental environment configuration.

To ensure experimental fairness and comparability, all models were trained under the same hardware, software, and hyperparameter settings. All models used the same input resolution of (640×640) and were initialized with their corresponding officially released pretrained weights for transfer learning, and followed the same training schedule of 200 epochs with an early-stopping patience of 50 epochs. Data augmentation was performed offline before training. Specifically, the original

266 images were expanded to 1,603 images using OpenCV-Python through rotation, perspective transformation, illumination perturbation, simulated occlusion, and scale resizing. Online augmentation was disabled during training to eliminate differences in augmentation strategies across experiments. Due to the nondeterministic operations introduced by the PyTorch data loader and CUDA kernel optimization, random seeds were not explicitly fixed. Nevertheless, all models were trained under the same stochastic initialization and data perturbation protocol, ensuring the relative fairness of cross-model evaluation. The complete training hyperparameters are listed in Table 2.

Hyperparameter	**Value**
Input resolution (imgsz)	640×640
Batch size	4
Workers	2
Optimizer	AdamW
Initial learning rate (lr0)	0.001
Weight decay	1×10^{-4}
Epochs	200
Early stopping	50
Warmup epochs	3

Table 2. Training hyperparameters.

Evaluation Metrics

To comprehensively evaluate the detection performance of the improved model for Luosifen outer packages in intelligent boxing scenarios, the following metrics were adopted: number of parameters (Params), floating-point operations (GFLOPs), single-image inference latency (ms), throughput rate (FPS), precision (P), recall (R), and mean average precision (mAP)[42]. For the OBB task, mAP was calculated based on rotated intersection over union (Rotated IoU), which penalizes angular deviations between the predicted boxes and the ground-truth boxes. The evaluation thresholds ranged from 0.5 to 0.95 with a step size of 0.05.

For a standard convolutional layer, the total number of learnable parameters is defined as:

$$\text{Params} = \left(C_{\text{in}} \times k^2 + b\right) \times C_{\text{out}} \tag{15}$$

where C_{in} and C_{out} denote the numbers of input and output channels, respectively; K denotes the convolution kernel size; and $b \in \{0,1\}$ indicates whether a bias term is used. If batch normalization (BN) is included, an additional $2C_{\text{out}}$ parameters are introduced, corresponding to the learnable scale parameter γ and shift parameter β.

Floating-point operations are used to measure the computational cost of model forward propagation. For a convolutional layer, the FLOPs are defined as:

$$\text{FLOPs} = 2 \times H_{\text{out}} \times W_{\text{out}} \times C_{\text{in}} \times k^2 \times C_{\text{out}} \tag{16}$$

where the factor 2 indicates that one multiply–accumulate (MAC) operation is counted as two floating-point operations, corresponding to one multiplication and one addition. H_{out} and W_{out} denote the height and width of the output feature map, respectively.

Precision and recall are fundamental metrics for evaluating detection performance and are defined as follows:

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{17}$$

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{18}$$

where TP, FP, and FN denote the numbers of true positives, false positives, and false negatives, respectively. The mean average precision (mAP) is calculated as the mean of average precision (AP) across all categories and is used to evaluate the overall detection performance. In this study, mAP@0.5 denotes mAP at an IoU threshold of 0.5, whereas mAP@0.5:0.95 denotes mAP averaged over IoU thresholds from 0.5 to 0.95 with a step size of 0.05.

Ablation Study

To evaluate the effectiveness of the proposed lightweight modules and identify their optimal deployment positions, this study avoided a global replacement strategy. Instead, key-layer sensitivity ablation and internal component ablation were performed to progressively examine the individual contributions of the GhostConv_OBB cost-efficient operation, dual-path edge enhancement, PECA attention, and the C3k2_GhostConv_OBB multi-scale fusion module to detection accuracy and edge-side inference efficiency. The experimental results are shown in Table3.

Group	Backbone Repl.	Head Repl.	Params (M)	GFLOPs (G)	Latency (ms)	FPS	mAP@0.5 (%)	mAP@0.5:0.95 (%)
Baseline	/	/	2.66	3.36	65.88	15.18	93.53	61.68
Exp.1	GC_OBB(0)	/	2.66	3.35	71.46	13.99	96.31	68.85
Exp.2	GC_OBB(3)	/	2.65	3.27	65.74	15.21	91.39	74.08
Exp.3	/	C3k2_GC(2)	2.65	3.32	75.19	14.58	83.10	51.06
Exp.4	/	C3k2_GC(5)	2.66	3.33	72.00	14.45	92.60	70.88
Exp.5	/	C3k2_GC(8)	2.64	3.30	74.63	14.64	88.72	60.25
Exp.6 (Ours)	/	C3k2_GC(11)	2.95	3.45	67.73	14.76	97.79	93.95
Exp.7	GC_OBB(0)	C3k2_GC(11)	2.95	3.45	75.07	13.32	94.92	91.04

Table 3. Key-layer sensitivity ablation study.

As shown in Table 3, deploying C3k2_GhostConv_OBB at different stages, namely detection heads 2, 5, 8, and 11, leads to substantial differences in mAP. Among them, detection head 11, located at the Neck-to-Head transition, achieves the best performance among the tested configurations. This result can be explained by three factors. First, detection head 11 provides better receptive-field matching. At this stage, the feature map has a relatively small spatial resolution of (20×20), and its receptive field covers most of the overall package region in the original image. Therefore, lightweight re-parameterized fusion at this stage can preserve macro-level geometric information while avoiding the fine-grained texture noise introduced by early high-resolution feature layers, such as Stage 2 with a resolution of (160×160). Second, detection head 11 provides a more balanced gradient flow. As this stage is located at the end of the Neck and is only one step away from the classification and regression heads, gradient attenuation is relatively limited. Consequently, the numerical errors introduced by the lightweight module are less likely to be amplified through multiple downsampling layers. Third, detection head 11 corresponds to the final step of multi-scale feature fusion. At this stage, YOLOv11 has already aggregated features from three FPN levels, resulting in rich semantic and spatial information. Embedding the discriminative enhancement modules, namely AAE and PECA, at this location can achieve a larger performance gain with limited additional computational cost. In contrast, replacing modules too early, such as at Stage 2, may disrupt low-level edge feature extraction, whereas replacing modules too late, such as inside the Head, provides limited opportunity to correct already consolidated features. Therefore, detection head 11 offers a favorable trade-off between detection accuracy and computational efficiency.

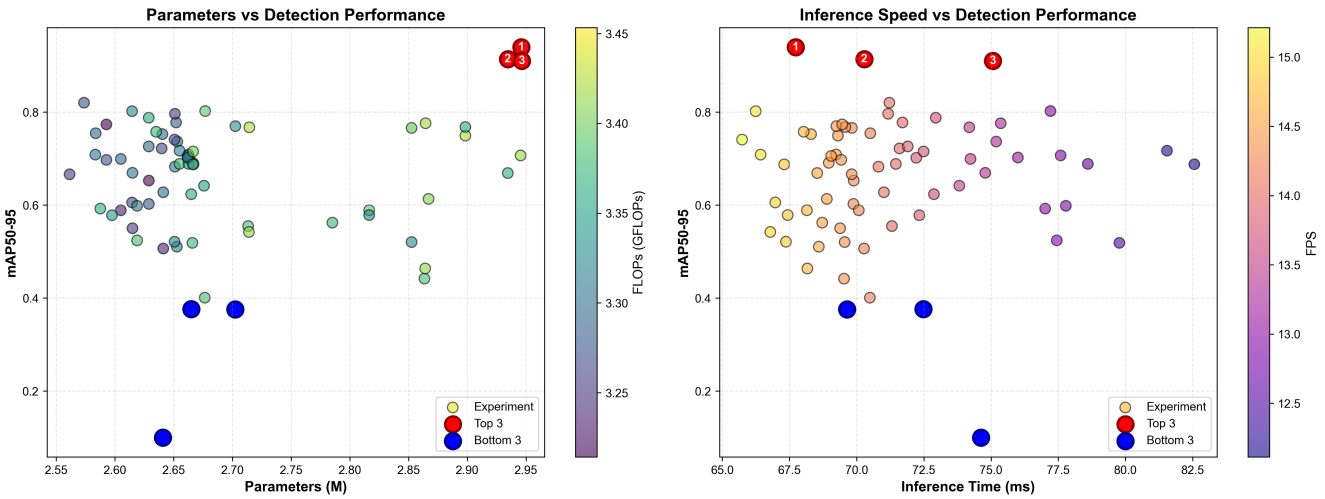


Figure 7. Ablation: Model complexity vs. detection performance.

To systematically evaluate the effect of different layer-placement combinations on the accuracy–efficiency trade-off, this study further expanded the ablation experiments from the seven representative configurations in Table3 to 71 configurations. These experiments covered the full backbone layers, Neck detection heads $\{2,5,6,8,9,11\}$, and cross-layer combinations through an exhaustive search. Fig.7 shows the two-dimensional scatter distributions of Params–mAP@0.5:0.95 and inference latency–mAP@0.5:0.95, where the color mappings represent GFLOPs and FPS, respectively. Table 4 lists six representative configurations with the best and worst performance.

As shown in Fig.7, most experimental points are densely distributed in the low-accuracy region with mAP@0.5:0.95 below 0.8, forming a clear ineffective search region. In contrast, high-accuracy solutions with mAP@0.5:0.95 above 0.9 are sparsely distributed within a narrow band in the upper-right region, corresponding to approximately 2.9–3.0 M parameters and 67–75 ms inference latency. This distribution indicates that the deployment position of the lightweight module is highly sensitive to model performance. Blind global replacement can easily lead to suboptimal configurations, whereas precise modification at specific layers is required to approach the accuracy–efficiency frontier.

The following six representative configurations were selected from the complete search space of 71 ablation experiments, including the three best-performing and three worst-performing configurations according to mAP@0.5:0.95. The number of parameters is provided only for cross-configuration comparison and is not used for ranking.

Rank	Experiment name	mAP@0.5:0.95 (%)	Params (M)	Type
1	Head 11 C3k2_GhostConv replacement	94.0	2.95	Best
2	Head 6 GhostConv replacement + Head 11 C3k2_GhostConv replacement	91.4	2.93	Best
3	Backbone 0 GhostConv replacement + Head 11 C3k2_GhostConv replacement	91.0	2.95	Best
4	Backbone 0 GhostConv replacement + Head 8 C3k2_GhostConv replacement	10.0	2.64	Worst
5	Head 6 GhostConv replacement + Backbone 6 C3k2_GhostConv replacement	37.5	2.70	Worst
6	Backbone 3 GhostConv replacement + Backbone 4 C3k2_GhostConv replacement	37.6	2.66	Worst

Table 4. Representative best- and worst-performing configurations selected from the 71 ablation experiments.

Table 4 further reveals the structural patterns of the best- and worst-performing configurations. The three best-performing configurations, with mAP@0.5:0.95 values of 94.0%, 91.4%, and 91.0%, all include the replacement of C3k2 with C3k2_GhostConv_OBB at detection head 11. Notably, replacing only detection head 11 achieves the best overall performance, whereas further combining this replacement with backbone-layer or additional head-layer replacement does not provide additional gains. In contrast, the three worst-performing configurations, with mAP@0.5:0.95 values ranging from 10.0% to 37.6%, share a common pattern: GhostConv_OBB is introduced into early backbone layers, or C3k2_GhostConv_OBB is deployed at non-critical backbone or head positions, such as backbone layers 4, 6, and 8 or detection head 8. These results suggest three key observations. First, lightweight replacement in the backbone may lead to the loss of fine-grained rotated-edge details. Second, detection head 11 is the best-performing deployment position among the tested configurations. Third, simultaneous replacement across multiple layers may instead lead to accuracy degradation due to possible feature interference, as shown in Table 3, Exp. 7.

The systematic search over 71 configurations reveals, from a statistical perspective, the high sensitivity of lightweight-module deployment position. Precise replacement only at detection head 11 is sufficient to approach the best accuracy–efficiency frontier, whereas blindly extending the replacement scope to backbone layers or non-critical detection heads leads to substantial accuracy degradation. Based on this finding, subsequent internal component ablation experiments were conducted at this optimal layer position to further disentangle the independent contributions of AAE and PECA within GhostConv_OBB.

After identifying detection head 11 as the optimal deployment position, ablation experiments were further performed on the internal components of C3k2_GhostConv_OBB to isolate the independent contributions and collaborative mechanism of dual-path edge enhancement, namely AAE, and progressive efficient channel attention, namely PECA. The results are shown in Table5.

Table 5. Ablation experiment results.

YOLOv11-OBB	AAE	PECA	Params (M)	GFLOPs (G)	Latency (ms)	FPS	mAP@0.5 (%)	mAP@0.5:0.95 (%)
✓	×	×	2.66	3.346	65.88	15.18	93.53	61.68
✓	✓	×	2.95	3.446	67.57	14.80	95.99	65.33
✓	×	✓	2.95	3.446	68.35	14.63	95.41	89.88
✓	✓	✓	2.95	3.445	67.73	14.76	97.79	93.95

To further reveal the internal optimization mechanisms of the proposed modules, Fig. 8 presents the convergence curves of Box Loss, Cls Loss, DFL Loss, and Angle Loss during training, while Table 6 reports the localization accuracy, convergence efficiency, and robustness metrics under high-IoU thresholds.

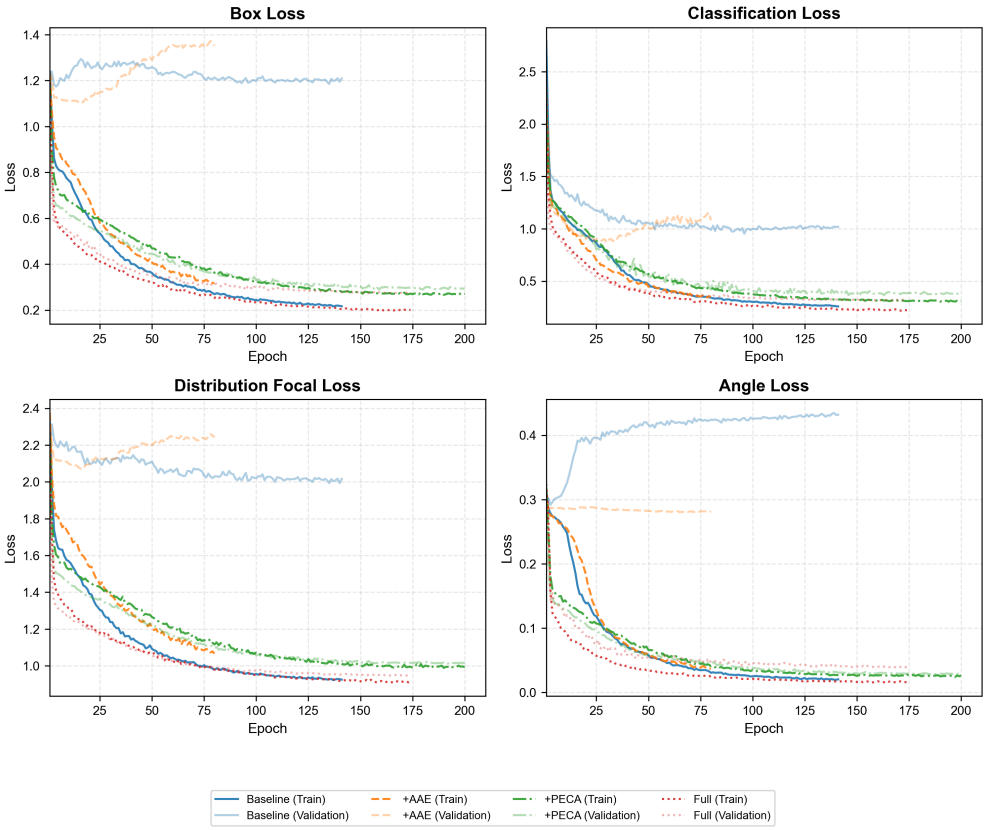


Figure 8. Comparison of training loss convergence curves in ablation experiments.

Group	IoU 0.75	IoU 0.85	IoU 0.90	P (%)	R (%)	Epoch	Time (s)	Noise Rob.	Aug. Rob.	Illum. Rob.
Baseline	0.7674	0.7211	0.6903	89.96	92.36	71.0	1476	0.9336	0.9271	0.8928
+AAE	0.7709	0.7230	0.6911	91.50	95.90	40.5	844	0.8988	0.8974	0.8561
+PECA	0.9699	0.9655	0.9625	91.42	91.96	100.5	2113	0.8061	0.8898	0.8628
Full	0.9778	0.9742	0.9717	95.96	96.05	87.5	1918	0.8272	0.9717	0.9254

Table 6. Effect of module components on positioning accuracy, convergence, and robustness.

From the perspectives of loss evolution and high-IoU performance, enabling PECA alone increased the accuracy at IoU = 0.90 from 0.6903 in the baseline model to 0.9625, while reducing the angle loss to 0.0543. In contrast, when AAE was used alone, the validation angle loss remained above 0.28 for an extended period, as shown in Fig. 8. Moreover, the validation curves

of box loss, classification loss, and DFL loss showed clear overfitting trends. These results suggest that solely enhancing edge features may cause the model to fit geometric noise in the training set, leading to reduced stability in angle prediction on the validation set. The Full configuration achieved the lowest values for all loss terms at 87.5 epochs, with box loss, classification loss, DFL loss, and angle loss reaching 0.2951, 0.3826, 1.0329, and 0.0362, respectively. It also achieved augmentation and illumination robustness scores of 0.9717 and 0.9254, respectively, as shown in Table 6. These results indicate the complementary roles of AAE and PECA: AAE enhances early edge visibility, whereas PECA provides channel-level background suppression that helps mitigate overfitting. Their combination enables a better balance between detection recall and localization precision.

The experimental results further show that PECA contributes most strongly to strict localization accuracy among the tested components. When only PECA was enabled, mAP@0.5:0.95 increased from 61.68% in the baseline model to 89.88%, corresponding to an improvement of 28.20 percentage points. This indicates that progressive channel modulation effectively suppresses complex background interference in production-line scenarios and improves the geometric alignment between predicted boxes and true package contours. In comparison, when AAE was enabled alone, mAP@0.5 increased from 93.53% to 95.99%, corresponding to an improvement of 2.46 percentage points, whereas mAP@0.5:0.95 increased only to 65.33%, corresponding to an improvement of 3.65 percentage points. This suggests that AAE helps reduce missed detections by strengthening edge and corner feature responses, but provides limited improvement for fine-grained rotation-angle regression when used alone.

Notably, when AAE and PECA were enabled simultaneously, the Full configuration achieved an mAP@0.5:0.95 of 93.95%, outperforming the AAE-only and PECA-only configurations by 28.62 and 4.07 percentage points, respectively. This result suggests a complementary effect between the two modules. Specifically, AAE improves the contour visibility of rotated objects and provides geometric priors for subsequent localization, while PECA further suppresses background noise and performs channel-level feature recalibration. Their joint use therefore enables the model to better satisfy the dual requirements of high detection recall and high-precision localization.

In summary, the proposed method adopts a key-layer sensitivity-guided deployment strategy, placing C3k2_GhostConv_OBB at detection head 11 and integrating the two enhancement components, AAE and PECA. Compared with the original model, the proposed method improves mAP@0.5 by 4.26 percentage points and mAP@0.5:0.95 by 32.27 percentage points, while increasing the inference latency by only 1.85 ms. These results demonstrate that the proposed method achieves a favorable trade-off between rotated object detection accuracy and inference efficiency under stringent edge-computing constraints.

Comparative Experiments

To comprehensively evaluate the effectiveness of the improved model, comparative experiments were conducted with representative horizontal bounding box detectors, including YOLOv8, YOLOv9, YOLOv10, YOLOv11, YOLOv12, and YOLOv26, as well as oriented bounding box variants, including YOLOv8-OBB, YOLOv11-OBB, and YOLOv26-OBB. To ensure a fair comparison, all models were trained on the same offline-augmented dataset, using an identical training/validation/test split of 7:2:1, corresponding to 1,122, 320, and 161 images, respectively. The same input resolution of (640×640) and training schedule of 200 epochs with an early-stopping patience of 50 epochs were used for all models. Each model was initialized with its corresponding officially released pretrained weights for transfer learning, and all non-architecture-related hyperparameters, including the optimizer, batch size, learning rate, and weight decay, were kept strictly consistent. The experimental results are shown in Table 7.

Model	P (%)	R (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Params (M)	GFLOPs (G)	FPS
YOLOv11n	89.42	93.39	89.30	79.98	2.59	3.22	15.89
YOLOv10n	87.45	89.18	92.00	87.30	2.66	3.34	15.18
YOLOv11n-OBB	89.96	92.36	93.53	61.68	2.66	3.35	15.35
YOLOv8n-OBB	93.35	98.68	94.73	89.86	3.08	4.22	17.43
YOLOv26n-OBB	85.28	90.23	89.38	86.99	2.65	3.14	12.05
YOLOv12n	93.06	95.27	93.60	87.58	2.57	3.24	12.01
YOLOv11s-OBB	95.19	97.66	96.47	93.26	9.72	11.27	8.84
YOLOv10x	97.03	96.03	97.60	93.66	31.66	85.51	2.15
YOLOv11x-OBB	94.54	96.30	95.67	92.76	58.79	101.89	1.85
Ours	**95.96**	**96.05**	**97.79**	**93.95**	**2.95**	**3.45**	**14.76**

Table 7. Comparison of experimental results between the proposed model and other models.

Overall, mainstream lightweight horizontal bounding box detectors, such as YOLOv11n and YOLOv10n, show certain

detection capability in general scenarios, achieving mAP@0.5 values of 89.30% and 92.00%, respectively. However, their performance under the stricter localization metric mAP@0.5:0.95 is limited to 79.98% and 87.30%, respectively. This indicates that horizontal bounding boxes cannot accurately conform to the contours of Luosifen outer packages under rotation, tilting, and extrusion deformation. As a result, substantial background interference is introduced, making it difficult to meet the requirements for accurate robotic grasping pose calculation.

After introducing the oriented bounding box (OBB) mechanism, the rotated object modeling capability of YOLOv8n-OBB, YOLOv26n-OBB, and YOLOv11n-OBB is improved to some extent. Specifically, YOLOv8n-OBB achieves an mAP@0.5:0.95 of 89.86%, while YOLOv26n-OBB reaches 86.99%, indicating that the angle prediction branch is beneficial for tilted-object localization. However, YOLOv11n-OBB, used as the baseline within the same model series, achieves an mAP@0.5 of 93.53% but only 61.68% mAP@0.5:0.95. This suggests that the standard OBB design has insufficient angle regression stability under complex production-line illumination, occlusion, and extrusion deformation, leading to poor geometric alignment between rotated boxes and true object contours. In addition, YOLOv12n, as a representative recent architecture, achieves an mAP@0.5:0.95 of 87.58%, which still remains below the 90% threshold under the strict localization metric.

To further improve detection accuracy, medium- and large-scale models such as YOLOv11s-OBB, YOLOv10x, and YOLOv11x-OBB increase network capacity and achieve higher detection performance, with mAP@0.5:0.95 values of 93.26%, 93.66%, and 92.76%, respectively. However, these models require substantially larger model sizes and computational costs, with parameter counts ranging from 9.72 M to 58.79 M and computational costs ranging from 11.27 to 101.89 GFLOPs. Their inference speeds range from 1.85 to 8.84 FPS, making them difficult to deploy on edge computing devices under storage, computational, and real-time constraints.

In contrast, the proposed model achieves the best overall performance among the tested models. Specifically, it obtains a precision of 95.96%, a recall of 96.05%, and an mAP@0.5 of 97.79%, corresponding to absolute improvements of 4.26 and 3.06 percentage points over YOLOv11n-OBB and YOLOv8n-OBB, respectively. Under the stricter mAP@0.5:0.95 metric, the proposed model reaches 93.95%, exceeding the strongest medium-scale competing models, YOLOv11s-OBB and YOLOv10x, by 0.69 and 0.29 percentage points, respectively. Moreover, the proposed model contains only 2.95 M parameters and requires 3.45 GFLOPs, which are comparable to those of lightweight baselines. Its inference speed of 14.76 FPS is substantially higher than those of the medium- and large-scale alternatives, demonstrating a favorable balance between detection accuracy and inference efficiency under edge-computing constraints.

Conclusion and Discussion

This study addresses the complex detection requirements of automated Luosifen outer-package boxing production lines, where multiple categories, rotation, tilting, and extrusion deformation must be handled simultaneously. To overcome the limitations of conventional template matching in adapting to multi-pose samples and the computational redundancy of existing OBB models that hinders edge deployment, a lightweight oriented detection method based on a key-layer sensitivity-guided replacement strategy is proposed.

The main contributions of this study are summarized as follows:

(1) **Dataset construction.** An oriented bounding box dataset for Luosifen outer packages was constructed under both static and dynamic conveyor-belt conditions. The dataset contains 1,603 images and was divided using a stratified 7:2:1 training/validation/test split, corresponding to 1,122, 320, and 161 images, respectively. This dataset provides a reproducible training and evaluation benchmark for real production-line scenarios.

(2) **Lightweight dual-path feature enhancement.** A GhostConv_OBB unit was proposed by integrating dual-path angle-aware enhancement (AAE) and progressive efficient channel attention (PECA). The ablation experiments show that AAE improves recall by enhancing edge and corner feature responses, whereas PECA independently improves mAP@0.5:0.95 by 28.20 percentage points through fine-grained background suppression. Their combination achieves 93.95% mAP@0.5:0.95 with only 2.95 M parameters and 3.45 GFLOPs.

(3) **Key-layer sensitivity-guided deployment.** A C3k2_GhostConv_OBB multi-scale fusion module was designed, and a systematic key-layer sensitivity analysis involving 71 configurations was conducted to identify the Neck-to-Head transition layer, namely detection head 11, as the best-performing deployment position. This selective replacement strategy avoids indiscriminate global lightweight modification and achieves the most favorable accuracy–efficiency trade-off among the tested configurations.

Experimental results show that the proposed method achieves 97.79% mAP@0.5 and 93.95% mAP@0.5:0.95. Compared with the YOLOv11n-OBB baseline, the proposed model increases the number of parameters by only 0.29 M, from 2.66 M to 2.95 M, the computational cost by only 0.09 GFLOPs, from 3.36 to 3.45 GFLOPs, and the inference latency by only 1.85 ms, from 65.88 ms to 67.73 ms. With this small additional cost, it achieves absolute improvements of 4.26 percentage points in mAP@0.5 and 32.27 percentage points in mAP@0.5:0.95, establishing a favorable accuracy–complexity trade-off. The model also outperforms both lightweight and medium-scale competing models in strict localization accuracy, empirically

demonstrating the complementary roles of AAE and PECA in improving recall and localization precision. This work provides a lightweight design reference for edge-oriented rotated object detection.

Future work will further investigate feature decoupling under severe stacking and occlusion scenarios, as well as INT8 quantization and deployment optimization for edge inference frameworks such as TensorRT.

References

1. Zhu, W. & Chen, K. Real-time object detection for unmanned aerial vehicles based on vision transformer and edge computing. *Sci. Rep.* **16**, 6814, DOI: 10.1038/s41598-026-37938-5 (2026).
2. Liu, Z., He, G., Dong, L., Jing, D. & Zhang, H. Task-sensitive efficient feature extraction network for oriented object detection in remote sensing images. *Remote. Sens.* **16**, 2271, DOI: 10.3390/rs16132271 (2024).
3. Zhao, B. & Chen, Y. Joint inference of image enhancement and object detection via cross-domain fusion transformer. *Computers* **15**, 43, DOI: 10.3390/computers15010043 (2026).
4. Qu, Z., Gao, L. Y., Wang, S. Y., Yin, H. N. & Yi, T. M. An improved YOLOv5 method for large objects detection with multi-scale feature cross-layer fusion network. *Image Vis. Comput.* **125**, 104518, DOI: 10.1016/j.imavis.2022.104518 (2022).
5. Zhang, J., Chen, Z., Yan, G., Wang, Y. & Hu, B. Faster and lightweight: An improved YOLOv5 object detector for remote sensing images. *Remote. Sens.* **15**, 4974, DOI: 10.3390/rs15204974 (2023).
6. Sun, Y. *et al.* PBA-YOLOv7: An object detection method based on an improved YOLOv7 network. *Appl. Sci.* **13**, 10436, DOI: 10.3390/app131810436 (2023).
7. Yao, Y. *et al.* On improving bounding box representations for oriented object detection. *IEEE Trans. Geosci. Remote. Sens.* **61**, 5600111, DOI: 10.1109/TGRS.2022.3231340 (2023).
8. Li, Z., Pang, C., Dong, C. & Zeng, X. R-YOLOv5: A lightweight rotational object detection algorithm for real-time detection of vehicles in dense scenes. *IEEE Access* **11**, 61546–61559, DOI: 10.1109/ACCESS.2023.3262601 (2023).
9. Huang, Z. *et al.* LO-Det: Lightweight oriented object detection in remote sensing images. *IEEE Trans. Geosci. Remote. Sens.* **60**, 5603515, DOI: 10.1109/TGRS.2021.3067470 (2022).
10. Su, B. Y. *et al.* Deep learning-based solar-cell manufacturing defect detection with complementary attention network. *IEEE Trans. Ind. Inform.* **17**, 4084–4095, DOI: 10.1109/TII.2020.3008021 (2020).
11. Wan, D. *et al.* Mixed local channel attention for object detection. *Eng. Appl. Artif. Intell.* **123**, 106442, DOI: 10.1016/j.engappai.2023.106442 (2023).
12. Wang, Q. *et al.* ECA-Net: Efficient channel attention for deep convolutional neural networks. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 11531–11539, DOI: 10.1109/CVPR42600.2020.01155 (2020).
13. Kong, Q., Wu, Z. & Song, Y. Online detection of external thread surface defects based on an improved template matching algorithm. *Measurement* **195**, 111087, DOI: 10.1016/j.measurement.2022.111087 (2022).
14. Liu, X., Li, Y., Guo, Y. & Zhou, L. Printing defect detection based on scale-adaptive template matching and image alignment. *Sensors* **23**, 4414, DOI: 10.3390/s23094414 (2023).
15. Wang, Y. *et al.* A fast template matching method based on improved ring projection transformation and local dynamic time warping. *Optik* **216**, 164954, DOI: 10.1016/j.ijleo.2020.164954 (2020).
16. Teng, M. Y. *et al.* An efficient and lightweight rotating target detection model for industrial scenarios. *Digit. Signal Process.* **168**, 105631, DOI: 10.1016/j.dsp.2025.105631 (2026).
17. Zhao, Z. & Li, S. ABFL: Angular boundary discontinuity free loss for arbitrary oriented object detection in aerial images. *IEEE Trans. Geosci. Remote. Sens.* **62**, 5611411, DOI: 10.1109/TGRS.2024.3368630 (2024).
18. Zhang, C. *et al.* Edge-optimized lightweight YOLO for real-time SAR object detection. *Remote. Sens.* **17**, 2168, DOI: 10.3390/rs17132168 (2025).
19. Shan, D., Cai, D., Tong, X., Li, Y. & Liu, D. LDA-YOLO: A YOLO-based rotated object detection method for remote sensing with large kernel attention and deformable alignment. *Appl. Sci.* **16**, 4168, DOI: 10.3390/app16094168 (2026).
20. Kuang, M. *et al.* DDM-YOLO: A lightweight oriented detection model for mature daylily fruits in complex environments. *J. King Saud Univ.-Comput. Inf. Sci.* **38**, 91, DOI: 10.1007/s44443-026-00559-z (2026).
21. Zhu, Y., Hong, J. & Wang, T. Weakly supervised oriented object detection in remote sensing via geometry-aware enhancement network. *Remote. Sens.* **18**, 1253, DOI: 10.3390/rs18081253 (2026).

22. Jamshidi, O., Abbasi, M., Ramazani, A., Salimi Shahraki, A. & Taherkordi, A. Modifier guided resilient CNN inference enables fault-tolerant edge collaboration for IoT. *Sci. Rep.* **15**, 45458, DOI: 10.1038/s41598-025-28454-z (2025).

23. Huang, F., Zheng, L., Wen, H., Dai, M. & Zhang, Z. A novel data augmentation method for few-shot industrial surface defect detection based on segment anything model adapter. *J. Intell. Manuf.* **37**, 2611–2631, DOI: 10.1007/s10845-025-02658-6 (2026).

24. Xia, G. S. *et al.* DOTA: A large-scale dataset for object detection in aerial images. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 3974–3983, DOI: 10.1109/CVPR.2018.00418 (2018).

25. Kong, C. *et al.* Object detection on low-compute edge SoCs: A reproducible benchmark and deployment guidelines. *Sci. Rep.* **16**, 5875, DOI: 10.1038/s41598-026-36862-y (2026).

26. Liu, Z. *et al.* Real-time beef cattle body condition scoring using EdgeBCS-YOLO: A lightweight framework for edge deployment. *Animals* **16**, 1143, DOI: 10.3390/ani16081143 (2026).

27. Zhou, Z. *et al.* Linear Gaussian bounding box representation and ring-shaped rotated convolution for oriented object detection. *Pattern Recognit.* **155**, 110677, DOI: 10.1016/j.patcog.2024.110677 (2024).

28. Mittal, P. A comprehensive survey of deep learning-based lightweight object detection models for edge devices. *Artif. Intell. Rev.* **57**, 242, DOI: 10.1007/s10462-024-10877-1 (2024).

29. He, Y. & Chen, Y. Research on the edge-discrepancy collaborative method for defect detection in casting DR images. *Materials* **19**, 900, DOI: 10.3390/ma19050900 (2026).

30. Cao, J., Bao, W., Shang, H., Yuan, M. & Cheng, Q. GCL-YOLO: A GhostConv-based lightweight YOLO network for UAV small object detection. *Remote. Sens.* **15**, 4932, DOI: 10.3390/rs15204932 (2023).

31. Huang, Q. *et al.* Oriented object detector with Gaussian distribution cost label assignment and task-decoupled head. *IEEE Trans. Geosci. Remote. Sens.* **62**, 5621916, DOI: 10.1109/TGRS.2024.3395440 (2024).

32. Nachet, R., Garrigós, J. & Boudghene Stambouli, T. GhostFaceNet++: Boosting efficiency and accuracy via CSP bottlenecks and channel attention. *J. Real-Time Image Process.* **2**, 201, DOI: 10.1007/s11554-025-01768-x (2025).

33. Zhao, X. & Chen, Y. YOLO-DroneMS: Multi-scale object detection network for unmanned aerial vehicle images. *Drones* **8**, 609, DOI: 10.3390/drones8110609 (2024).

34. Su, H. AdaResNet: Enhancing residual networks with dynamic weight adjustment for improved feature integration, DOI: 10.48550/arXiv.2408.09958 (2024). ArXiv:2408.09958.

35. Zheng, Q. *et al.* Boundary-aware network with two-stage partial decoders for salient object detection in remote sensing images. *IEEE Trans. Geosci. Remote. Sens.* **61**, 5605713, DOI: 10.1109/TGRS.2023.3260825 (2023).

36. Tang, Q. *et al.* ATDIOU: Arctangent differential loss function for bounding box regression. *Sensors* **26**, 1545, DOI: 10.3390/s26051545 (2026).

37. Li, C. Oriented object detection in remote sensing images based on angle quality estimation. *Sci. Rep.* **16**, 1576, DOI: 10.1038/s41598-025-31034-w (2026).

38. Wen, R., Yang, Z., Chen, T., Li, H. & Li, K. Progressive representation recalibration for lightweight super-resolution. *Neurocomputing* **504**, 240–250, DOI: 10.1016/j.neucom.2022.07.050 (2022).

39. Wang, D. & Wu, H. IoU regression with H+L-sampling for accurate detection confidence. *Sensors* **21**, 4433, DOI: 10.3390/s21134433 (2021).

40. Wu, S., Li, X. & Wang, X. IoU-aware single-stage object detector for accurate localization. *Image Vis. Comput.* **97**, 103911, DOI: 10.1016/j.imavis.2020.103911 (2020).

41. He, Y., Li, C., Shi, M. & Li, X. Lightweight remote sensing image dehazing via dual-stage feature calibration and hierarchical context aggregation. *Remote. Sens.* **18**, 564, DOI: 10.3390/rs18040541 (2026).

42. Wang, L. *et al.* A systematic literature review of lightweight YOLO models for object detection. *PeerJ Comput. Sci.* **11**, e3357, DOI: 10.7717/peerj-cs.3357 (2025).

Funding

This work was supported by the 2025 University-level Scientific Research Project of Liuzhou Polytechnic University (Special Project for the "Double High" Professional Group of the College of Mechatronic Engineering), Grant No. 2025JD06.

Data availability

The dataset and improved modules generated and used during the current study are publicly available at https://github.com/TPNS-007/luosifen-obb-detection.

Author contributions statement

G.L. conceived the research, designed the network modules, wrote the manuscript, and supervised the study. S.Z. performed dataset preprocessing, model training, and ablation experiments. Y.L. conducted data analysis, prepared figures and tables, and revised the manuscript. All authors reviewed and approved the final manuscript.

Additional information

Competing interests: The authors declare no competing interests.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AblationExperimentDataSummaryTable.xlsx
- PerformanceComparisonofReplacementPositions.xlsx
- stream.jpg