Enhanced obstacle detection for intelligent agricultural machinery via transfer learning and improved YOLO11

Huan Wan[1,2], Xianlu Guan[1,2], Yuanzhen Ou[1,3], Rui Jiang[1,2,3,4,5,6,7], Zhiyan Zhou[1,2,3,4,5,6,7,8*]

(1. *College of Engineering, South China Agricultural University and Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou 510642, China*;
2. *Guangdong Provincial Key Laboratory of Agricultural Artificial Intelligence (GDKL-AAI), Guangzhou 510642, China*;
3. *Guangdong Engineering Research Center for Agricultural Aviation Application (ERCAAA), Guangzhou 510642, China*;
4. *Key Laboratory of Key Technology on Agricultural Machine and Equipment (South China Agricultural University), Ministry of Education, Guangzhou 510642, China*;
5. *State Key Laboratory of Agricultural Equipment Technology, Guangzhou 510642, China*;
6. *Key Technology Innovation Research Center of Rice Smart Farming of South China Agricultural University - Dongyuan County, Heyuan 517554, Guangdong, China*;
7. *Heyuan Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Heyuan 517000, Guangdong, China*;
8. *The Centre for Pesticide Application and Safety (CPAS), School of Agriculture and Food Sciences, the University of Queensland, Gatton, QLD 4343, Australia*)

Abstract: Reliable obstacle detection is critical for the safe operation of intelligent agricultural machinery in unstructured farmland environments. However, achieving robust detection in agricultural settings remains challenging due to limited annotated data and significant domain shifts between generic visual datasets and agricultural scenes, which constrain the effectiveness of existing object detection models. To address these challenges, this study proposes an enhanced YOLO11 framework with task-aligned transfer learning for detecting obstacles in farmland. The proposed approach consists of three core components: 1) architectural refinements, including the integration of CBAM attention modules and an improved SPPF structure to enhance multi-scale feature representation; 2) task-aligned pretraining on a curated COCO subset comprising 70 552 images containing task-relevant object categories; and 3) a staged fine-tuning strategy that combines backbone freezing with subsequent end-to-end optimization on a farmland obstacle dataset. Experimental results demonstrate that the proposed method achieves a mAP@0.5 of 0.934 on the test set, improving mAP@0.5 by 6.4 percentage points over YOLO11-s. Furthermore, after TensorRT optimization, the model reaches 82.6 FPS on the Jetson AGX Orin platform while maintaining a mAP@0.5 of 0.927, confirming its suitability for real-time deployment. These findings indicate that task-aligned transfer learning, combined with targeted architectural enhancements, effectively mitigates data scarcity and domain shift in agricultural obstacle detection.

Keywords: agricultural machinery, obstacle detection, transfer learning, YOLO11, embedded deployment
DOI: 10.25165/j.ijabe.20261904.10644

Citation: Wan H, Guan X L, Ou Y Z, Jiang R, Zhou Z Y. Enhanced obstacle detection for intelligent agricultural machinery via transfer learning and improved YOLO11. Int J Agric & Biol Eng, 2026; 19(4): 256–267.

1 Introduction

The development of intelligent agricultural machinery requires reliable environmental perception, particularly the detection of obstacles such as humans, vehicles, livestock, and agricultural facilities[1,2]. Unlike structured urban scenarios, agricultural environments are inherently unstructured and visually complex, characterized by crop residues, cluttered backgrounds, varying illumination, and domain-specific obstacles, which pose significant challenges for robust obstacle detection[3,4]. Vision-based perception offers rich semantic cues, high spatial resolution, and cost-effectiveness, making it a widely adopted solution for intelligent agricultural machinery[5,6].

Early vision-based methods relied on handcrafted features, such as HOG[7] and SIFT[8], which exhibit limited robustness in agricultural conditions[9]. Deep learning-based object detection has substantially improved detection performance. The YOLO family[10], owing to its single-stage design and favorable accuracy–speed trade-off, has been applied to agricultural obstacle detection, including lightweight YOLOv3 variants for orchard environments[11], YOLOv4-based farmer detection in farmland[2], and attention-enhanced YOLOv5s models for orchard obstacle detection[12]. More recently, YOLOv8 has also been adopted for field obstacle detection[13]. Building on this line of work, YOLO11 has further advanced the YOLO family: relative to YOLOv8, it replaces the C2f block with a more parameter-efficient C3K2 module and introduces a C2PSA attention block after SPPF to strengthen spatial feature discrimination; relative to YOLOv10, which instead pursues

Received date: 2026-04-15 **Accepted date:** 2026-07-21
Biographies: **Huan Wan**, PhD candidate, research interest: agricultural information collection and data processing, Email: wanhuan@scau.edu.cn; **Xianlu Guan**, PhD candidate, research interest: agricultural artificial intelligence technology, Email: guanxl@stu.scau.edu.cn; **Yuanzhen Ou**, PhD candidate, research interest: remote sensing technology, Email: yzou@stu.scau.edu.cn; **Rui Jiang**, PhD, Associate Professor, research interest: agricultural engineering, Email: ruiojiang@scau.edu.cn.
***Corresponding author: Zhiyan Zhou**, PhD, Professor, research interest: agricultural artificial intelligence technology. College of Engineering, South China Agricultural University, Guangzhou 510642, China. Tel: +86-20-38676975, Email: zyzhou@scau.edu.cn.

efficiency through NMS-free dual-assignment training at the cost of a specialized detection head[14], YOLO11 retains conventional NMS-based inference while achieving a more favorable accuracy-parameter trade-off. Although these approaches demonstrate the effectiveness of YOLO architectures in agricultural scenes, they are typically designed for narrowly defined environments and limited obstacle categories, which constrains their adaptability to diverse and safety-critical farmland conditions. Moreover, deep learning models require large-scale, domain-specific training data[15]. However, general-purpose datasets such as COCO[16] and KITTI[17] exhibit substantial domain gaps with agricultural environments in object appearance, background structure, and illumination. Due to high annotation costs, agricultural obstacle datasets remain limited in scale, causing performance degradation when models pretrained on generic datasets are directly transferred to farmland scenarios.

Transfer learning offers an effective paradigm for alleviating data scarcity by transferring knowledge from large-scale source datasets to task-specific targets[18,19]. While transfer learning has demonstrated effectiveness in agricultural applications such as plant disease recognition[20], fruit detection[21], and livestock monitoring[22], agricultural obstacle detection poses additional challenges including strong domain shifts, complex backgrounds, and stringent safety requirements. To explicitly address such domain shifts, a growing body of work has applied domain adaptation (DA) techniques in agricultural vision, including diffusion-based domain-adapted data synthesis for vineyard shoot detection[23], Fourier-transform- and CycleGAN-based unsupervised DA for cross-sensor and cross-field weed/crop segmentation[24,25], and domain-adversarial training for rice disease recognition and cross-region yield prediction[26,27]. These studies confirm that explicit distribution alignment can mitigate domain shift-induced performance degradation, but they operate at the pixel or feature level and target single-category classification or segmentation tasks; extending such alignment to multi-category obstacle detection, where illumination variation, cluttered backgrounds, and heterogeneous obstacle categories interact simultaneously, remains largely unaddressed. This motivates the task-aligned transfer learning strategy adopted in the present work, which pursues comparable domain-specific adaptation benefits through selective source-data curation and staged fine-tuning rather than explicit pixel- or feature-level alignment, offering a more scalable path for multi-category detection under limited annotation budgets. Overall, most existing approaches, where based on straightforward fine-tuning or the DA techniques discussed above, do not simultaneously address domain relevance, architectural adaptation, or optimization strategies. Specifically, three critical gaps remain: 1) the lack of task-aligned pretraining strategies that selectively exploit relevant knowledge from source domains; 2) insufficient architectural adaptations to capture the distinctive visual characteristics of agricultural obstacles; and 3) limited investigation of staged optimization strategies that balance generic feature preservation with domain-specific adaptation.

To bridge these gaps, this study proposes an integrated framework that combines architectural enhancement with task-aligned transfer learning for robust obstacle detection in agricultural environments. First, the YOLO11 architecture is enhanced by integrating Convolutional Block Attention Modules (CBAM) at multiple detection scales to improve feature discrimination in cluttered backgrounds, and by redesigning the Spatial Pyramid Pooling Fast (SPPF) module to better preserve features of small and slender objects. Second, a task-aligned transfer learning strategy is developed by selectively extracting 70 552 images containing relevant object categories from the COCO dataset, thereby reducing negative transfer from irrelevant classes. Third, a two-stage fine-tuning strategy is employed, consisting of an initial frozen-backbone phase followed by end-to-end optimization. During training, MPDIoU loss is employed to improve localization accuracy for slender and partially occluded obstacles. A farmland obstacle dataset covering seven safety-critical categories is constructed and expanded through data augmentation. Extensive experiments demonstrate that the proposed framework achieves superior detection accuracy while maintaining real-time inference performance on edge computing platforms.

Accordingly, this study aims to develop an effective and efficient obstacle detection framework for autonomous agricultural machinery operating in complex farmland environments. Specifically, the objectives are to: 1) develop an enhanced YOLO11-based obstacle detection framework by integrating CBAM at multiple detection scales and redesigning the SPPF module to improve feature discrimination in cluttered farmland environments and enhance the detection of small and slender obstacles; 2) design a task-aligned transfer learning strategy that combines selective pretraining on task-relevant source-domain data with a two-stage fine-tuning scheme to effectively balance generic feature preservation and domain-specific adaptation; and 3) construct a farmland obstacle dataset covering seven safety-critical categories to serve as a basis for validating the proposed framework under complex agricultural conditions.

2 Materials and methods

2.1 Data collection and preparation

This study constructed an obstacle detection dataset tailored for intelligent agricultural machinery in farmland environments. The images were collected in Xinghua City, Jiangsu Province, China, during November 2024 and March 2025, capturing seasonal and climatic variations that improve model generalization. Data acquisition was performed using a GoPro Hero9 and a ZED 2i camera (Figure 1). The Hero9 provided stable recordings under outdoor conditions, while the ZED 2i, with a 110° field of view and high optical quality, enabled the capture of diverse scenes and small-scale obstacles.

Figure 1 Data acquisition platform

To support task-specific transfer learning, this study defined a taxonomy of obstacles relevant to agricultural navigation, including persons, livestock (cows), vehicles (cars, trucks, motorcycles, and agricultural machinery), and infrastructure (wire poles). Based on this taxonomy, 406 images were manually collected and annotated into seven categories: agricultural machinery, car, cow, motorcycle, person, truck, and wire pole.

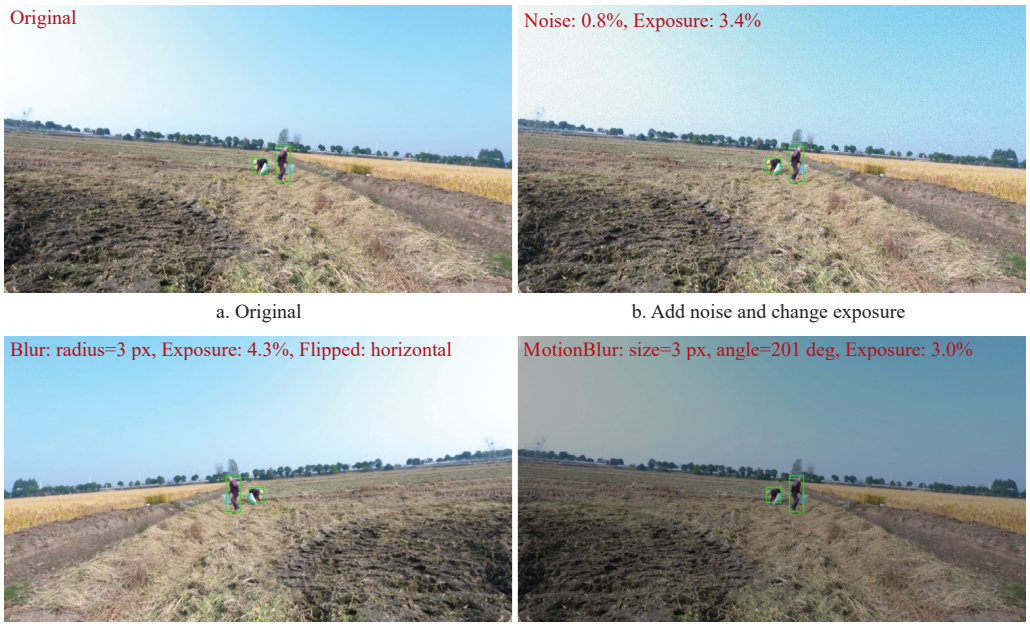
To enhance robustness under field conditions, a hierarchically controlled multi-strategy data augmentation scheme was designed to

simulate agricultural interferences such as illumination variation, motion blur, and particulates (e.g., dust and mist). As listed in Table 1, each sample was augmented using two or three combined strategies, ensuring diversity while preserving critical features.

As shown in Figure 2, the scheme retains critical target features while simulating realistic field interferences, making training samples more representative of actual environments and enhancing model robustness in complex agricultural scenarios. This process expanded the dataset to 1624 images, which were split into training, validation, and test sets at a ratio of 7.0:1.5:1.5. The class distribution is shown in Figure 3a, where agricultural machinery, person, and wire pole are the most frequent categories. In addition, to facilitate transfer learning, a domain-relevant subset of the COCO dataset was constructed. Images containing at least one instance of five categories—person, car, motorcycle, truck, and cow—were selected, and irrelevant annotations were removed. The resulting subset contains 70 552 training and 2992 validation images, effectively complementing the target dataset (Figure 3b) and providing strong initialization for model training.

Table 1 Image augmentation strategies

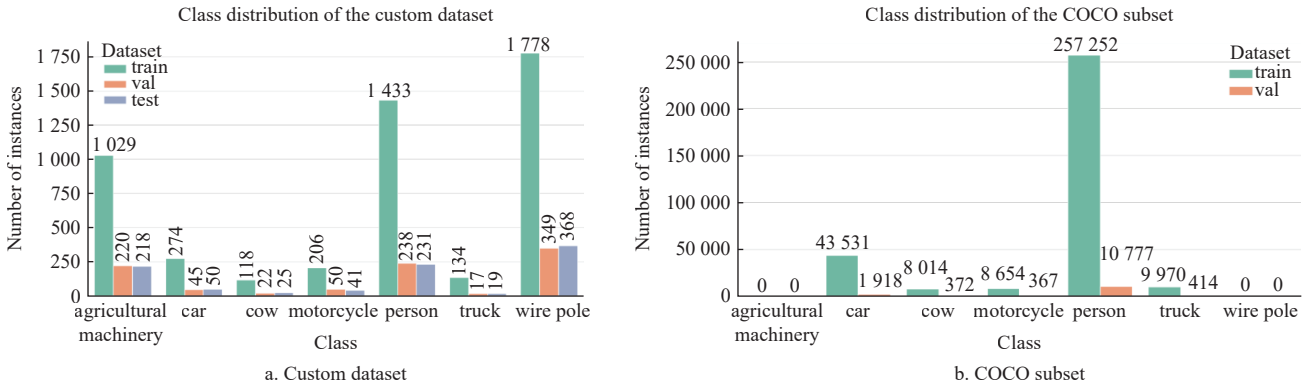
Augmentation type	Parameter range	Scenarios
Exposure adjustment	3%-8%; –8%-(–3%)	Light fluctuations (avoiding over- or under-exposure)
Gaussian blur	1-3 px	Camera defocus
Motion blur	3-5 px (0°-360°)	Agricultural machinery moving, vibration and obstacle movement
Noise injection	0.3%-1.0%	Low-light sensor noise, dust, mist
Horizontal flip	-	Obstacles appearing on left or right sides of agricultural machinery's field of view



a. Original b. Add noise and change exposure

c. Gaussian blur, change exposure, and flip horizontally d. Motion blur and change exposure

Figure 2 Comparison between the original image and its augmented counterparts



a. Custom dataset b. COCO subset

Figure 3 Class distributions of the custom dataset and the COCO subset used for task-aligned transfer learning

2.2 Network architecture enhancement

YOLO11 provides five model variants (n, s, m, l, x) with a unified architecture but different scaling factors in depth, width, and channel size. Among them, YOLO11-s was selected as the baseline model because it delivers substantially better performance than YOLO11-n while remaining more computationally efficient than larger variants (m, l, x), making it suitable for practical deployment on edge devices.

However, agricultural obstacle detection presents several unique challenges that limit the effectiveness of the baseline model. These include complex backgrounds, varying illumination, sensor noise, and significant object scale differences. Additionally, precise localization is essential in safety-critical scenarios, yet the baseline model often suffers from suboptimal bounding box regression under these conditions. Therefore, it is necessary to improve the architecture of the baseline model to enhance feature representation and multi-scale detection in agricultural scenes. The overall architecture of the improved YOLO11 is illustrated in Figure 4.

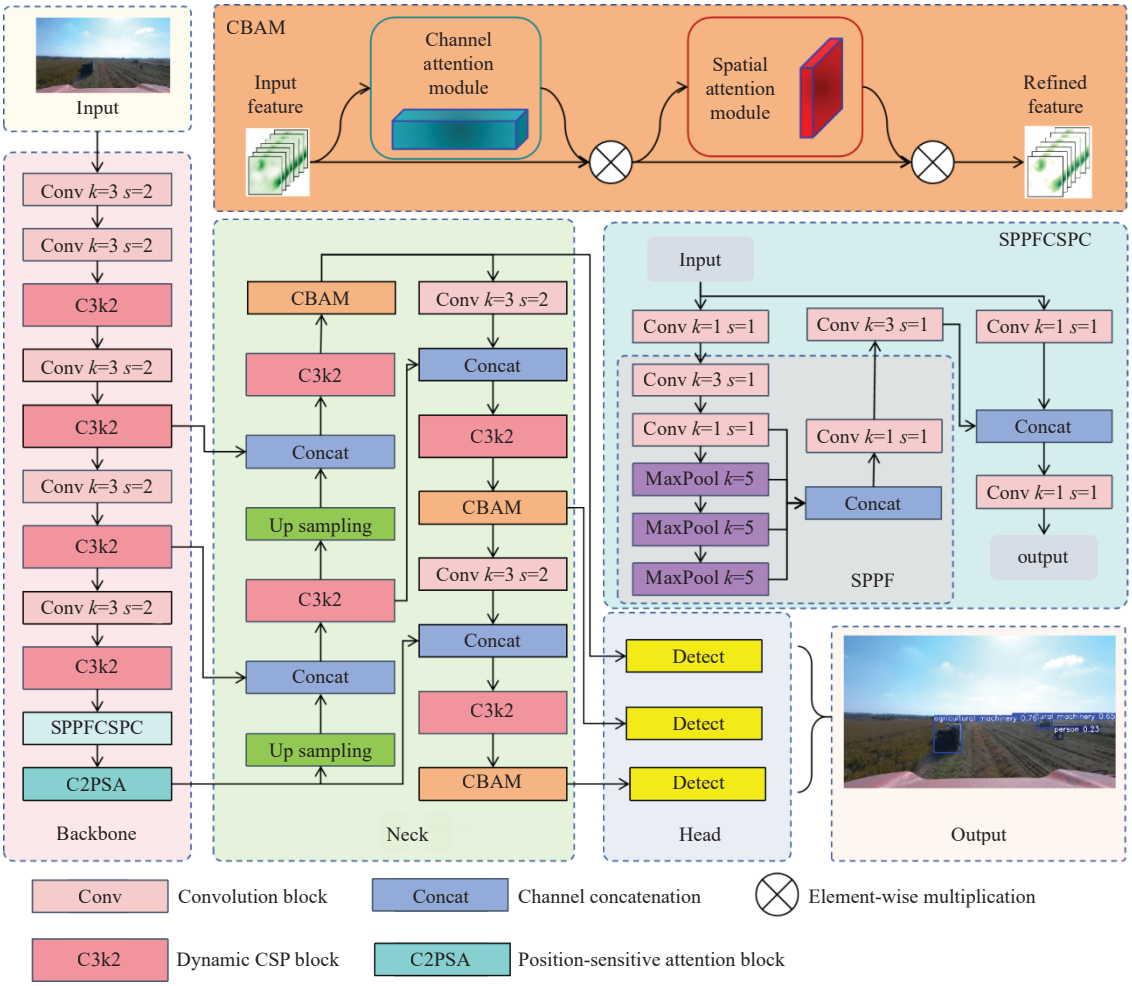


Figure 4 Overall architecture of improved YOLO11

The original YOLO11 employs the SPPF module to enhance receptive fields through multi-scale pooling operations. However, our empirical analysis reveals two critical limitations of SPPF in agricultural obstacle detection scenarios: 1) insufficient feature diversity due to the lack of cross-stage feature fusion, which is essential for distinguishing obstacles with significant intra-class variations (e.g., person in different postures); 2) constrained gradient flow in deeper layers, which hampers the model's ability to learn discriminative features for both small-scale objects (e.g., wire pole, motorcycle) and large-scale entities (e.g., truck, agricultural machinery) simultaneously.

To address these limitations, this study proposes replacing SPPF with the SPPFCSPC module, which introduces two key enhancements: First, by integrating the Cross Stage Partial (CSP) mechanism[28], SPPFCSPC splits feature channels into two branches—one undergoing spatial pyramid pooling and the other bypassing the pooling operations—before merging them through concatenation. This design explicitly preserves multi-granularity feature representations while maintaining computational efficiency. Second, the CSP structure facilitates enhanced gradient propagation by creating shortcut paths, thereby alleviating gradient vanishing issues and reducing redundant gradient information during backpropagation. This architectural refinement is particularly crucial for agricultural scenarios, where the model must simultaneously handle extreme scale variations and dynamic appearance changes.

The general formulation of SPPFCSPC is:

$$Y = \text{Concat}(\phi(F_1), \Psi(\text{SPPF}(F_2))) \quad (1)$$

where, the input feature map F is split into two parts (F_1, F_2). The mapping $\phi(\cdot)$ denotes a lightweight convolutional transformation consisting of a 1×1 convolution followed by a 3×3 convolution and another 1×1 convolution. The mapping $\Psi(\cdot)$ represents the feature refinement process following the SPPF block, in which the multi-scale pooled features are fused by a 1×1 convolution and refined by a 3×3 convolution. Finally, the outputs of the two branches are concatenated to form the final feature representation.

Furthermore, to strengthen the model's discriminative capability, three Convolutional Block Attention Modules (CBAM) were inserted into the neck, each positioned immediately before a detection head. This configuration guides the network to emphasize informative regions while suppressing interference from complex agricultural backgrounds[29].

To improve the discrimination of small and occluded obstacles in farmland scenes, this study inserted CBAM immediately before each of the three detection heads of YOLO11. As illustrated in Figure 4, the CBAM module sequentially refines an input feature map in two steps: the Channel attention module and the Spatial attention module. This dual attention mechanism allows the network to focus on "what" and "where" to attend, thereby improving feature representation.

The channel attention module focuses on refining the inter-channel relationship of features. As shown in Figure 5a, global average pooling (GAP) and global max pooling (GMP) were applied to the input feature map $F \in \mathbb{R}^{C\times H\times W}$ along the spatial dimensions, yielding two spatial context descriptors, which were then forwarded to a shared multi-layer perceptron (MLP) with one hidden layer, followed by element-wise summation and a sigmoid

activation to produce the channel attention map $M_c \in \mathbb{R}^{C\times1\times1}$:

$$M_c(F) = \sigma\left(\mathrm{MLP}\left(\mathrm{GAP}(F)\right) + \mathrm{MLP}\left(\mathrm{GMP}(F)\right)\right) \quad (2)$$

where, σ denotes sigmoid activation.

The refined feature map F' is then obtained by channel-wise multiplication:

$$F' = M_c(F) \otimes F \quad (3)$$

The spatial attention module focuses on refining the spatial locations of the features (Figure 5b). It applies max pooling and average pooling operations along the channel axis, generating two 2D maps, which were concatenated and passed through a standard convolution layer followed by a sigmoid activation to generate the spatial attention map $M_s \in \mathbb{R}^{1\times H\times W}$:

$$M_s(F') = \sigma\left(f^{7\times7}\left(\left[\mathrm{AvgPool}(F'); \mathrm{MaxPool}(F')\right]\right)\right) \quad (4)$$

where, σ is sigmoid activation, $f^{7\times7}$ is a convolution layer with a kernel size of 7 ×7.

The final refined feature map F'' is computed as:

$$F'' = M_s(F') \otimes F' \quad (5)$$

Where ⊗ denotes the element-wise multiplication.

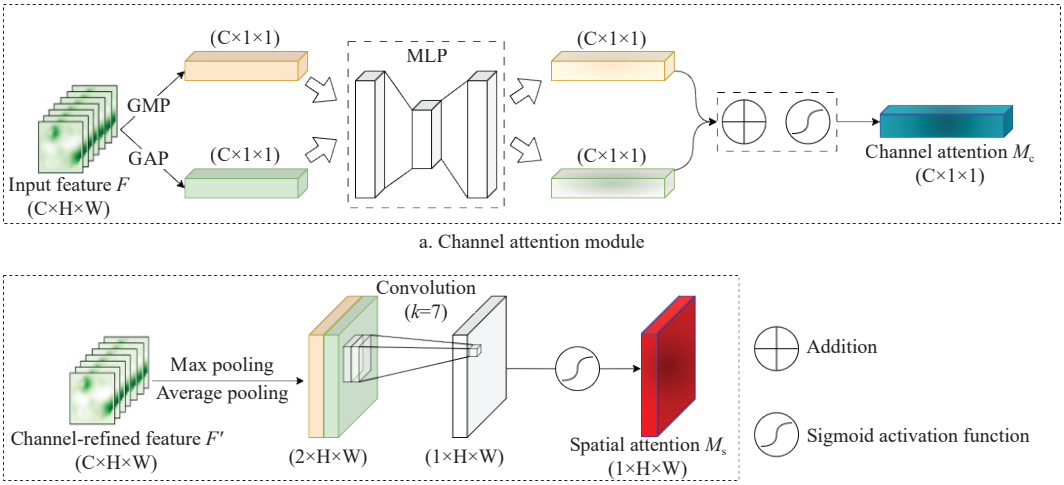


Figure 5 Structures of the channel attention module and spatial attention module in the CBAM

2.3 Loss function of the improved YOLO11

The YOLO11 detection loss comprises bounding box regression loss, classification loss, and distribution focal loss (DFL). In agricultural scenes, obstacles with low contrast and irregular shapes often cause traditional IoU-based losses (e.g., CIoU) to provide insufficient gradient signals when predicted and ground-truth boxes share similar aspect ratios but differ in scale. To address this, this study replaces the default CIoU loss[30] with Minimum Point Distance IoU (MPDIoU) loss[31], while keeping DFL and classification loss unchanged. MPDIoU directly minimizes the distances between corresponding top-left and bottom-right corners of predicted and ground-truth boxes, offering more precise localization guidance. MPDIoU measures the spatial misalignment between predicted and ground-truth bounding boxes based on the distances between their top-left and bottom-right corner points (Figure 6).

Let A and B denote ground-truth and predicted bounding boxes with top-left coordinates $\left(x_1^{gt}, y_1^{gt}\right), \left(x_1^{prd}, y_1^{prd}\right)$ and bottom-right coordinates $\left(x_2^{gt}, y_2^{gt}\right), \left(x_2^{prd}, y_2^{prd}\right)$. The squared Euclidean distances between corresponding corners are:

$$d_1^2 = \left(x_1^{prd} - x_1^{gt}\right)^2 + \left(y_1^{prd} - y_1^{gt}\right)^2 \quad (6)$$

$$d_2^2 = \left(x_2^{prd} - x_2^{gt}\right)^2 + \left(y_2^{prd} - y_2^{gt}\right)^2 \quad (7)$$

MPDIoU is defined as:

$$\mathrm{MPDIoU} = \frac{A \cap B}{A \cup B} - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2} \quad (8)$$

The corresponding loss function is formulated as:

$$\mathcal{L}_{\mathrm{MPDIoU}} = 1 - \mathrm{MPDIoU} = 1 - \mathrm{IoU} + \frac{d_1^2}{d^2} + \frac{d_2^2}{d^2} \quad (9)$$

where, $d^2 = w^2 + h^2$ ensures scale invariance. This formulation provides meaningful gradients even for non-overlapping boxes, improving convergence and localization accuracy for thin or irregularly shaped agricultural obstacles.

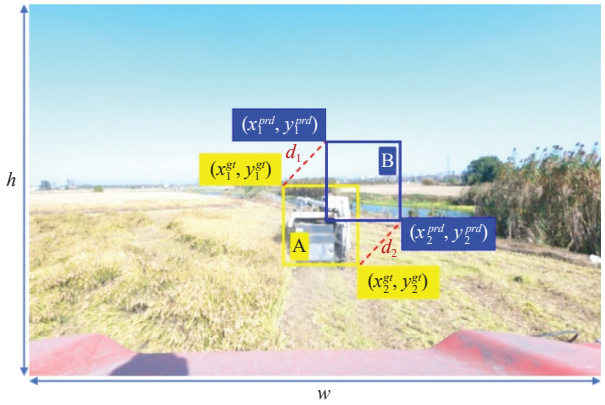


Figure 6 Factors of MPDIoU loss

2.4 Transfer learning strategy

To address the limited size of the manually annotated agricultural dataset, a two-stage transfer learning strategy was employed. The model was initialized with weights pre-trained on a 5-class COCO subset (as presented in Section 2.1).

Pre-training Configuration. The base model was trained for 150 epochs with an input size of 640×640 pixels and a batch size of 32. The initial learning rate was set to 0.01 with a cosine decay schedule to a final factor of 0.0001, and early stopping was triggered if no improvement occurred within 30 epochs. Standard YOLO training protocols were applied unless otherwise specified, including SGD with momentum 0.937 and weight decay 0.0005.

Notably, Mosaic and Mixup augmentations were disabled during this phase to ensure a clean feature alignment with the subsequent custom dataset. The bounding box regression loss was configured with a weight of 7.5, classification loss with 0.5, and DFL with 1.5, utilizing the proposed MPDIoU for localization.

Two-stage Fine-tuning. To prevent catastrophic forgetting of generic features, this study adopted a progressive fine-tuning scheme: 1) Partial Fine-tuning: The backbone (first 10 layers) was frozen for 100 epochs. Training utilized a conservative initial learning rate (lr0=0.001) alongside aggressive data augmentation (Mosaic=1.0, Mixup=0.1) to enhance generalization on the limited agricultural samples. 2) Full Fine-tuning: All layers were unfrozen and trained for an additional 100 epochs with a reduced learning rate (lr0=0.0001). Augmentation intensity was simultaneously reduced (Mosaic=0.5, Mixup=0.0) to refine the detection heads for precise obstacle localization.

Optimization Details. For all training phases, mixed precision training was enabled. The remaining hyperparameters, including optimizer settings, warm-up epochs, and validation intervals, followed the default YOLO11 settings as detailed in the official Ultralytics[32] release documentation.

Mosaic and Mixup augmentations were intentionally disabled during the pre-training stage because the objective of pre-training was to learn stable and transferable visual representations from the selected COCO subset while preserving the natural distribution of object appearance and contextual information. During fine-tuning, however, the target agricultural dataset was considerably smaller than the pre-training dataset. Therefore, Mosaic and Mixup augmentations were enabled to increase sample diversity, improve robustness to scale variation and occlusion, and reduce the risk of overfitting. Following the common training strategy adopted in the YOLO framework, Mosaic augmentation was disabled during the final training epochs to facilitate stable convergence.

2.5 Experimental environment and model evaluation

All models were trained with PyTorch 2.2.2+cu118 and Ultralytics 8.3.76. Model training was conducted on a Lenovo workstation equipped with an Intel i7-11700 CPU, 64 GB RAM, and an NVIDIA RTX A4000 GPU with 16GB VRAM. The trained model was converted to a TensorRT engine on the NVIDIA Jetson AGX Orin to evaluate its real-time inference performance and verify its applicability to embedded deployment in intelligent agricultural machinery.

The models were evaluated on the custom dataset as described in Section 2.1. Standard metrics were employed, including Precision (P), Recall (R), Average Precision (AP), mean Average Precision (mAP), and F1-score. The formulas for these metrics are as follows:

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{10}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{11}$$

where TP, FP, and FN denote true positives, false positives, and false negatives, respectively.

$$\mathrm{AP} = \int_0^1 P(R)\,\mathrm{d}R \tag{12}$$

where, $P(R)$ denotes precision as a function of recall. AP is computed as the area under the Precision–Recall (P-R) curve for a specific class.

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=0}^{N-1}\mathrm{AP}_i \tag{13}$$

where, N is the number of classes, AP_i is the AP of class i.

$$\text{F1-score} = 2\times\frac{P\times R}{P+R} \tag{14}$$

F1-score balances precision and recall, a harmonic mean of precision (P) and recall (R), providing a more comprehensive measure of a model's performance, especially in imbalanced datasets.

3 Results and discussion

3.1 Two-stage transfer learning analysis

The proposed two-stage transfer learning strategy demonstrates effectiveness in adapting the improved YOLO11 model to agricultural obstacle detection tasks. Figure 7 shows the comprehensive training process, revealing distinct learning patterns and convergence behaviors.

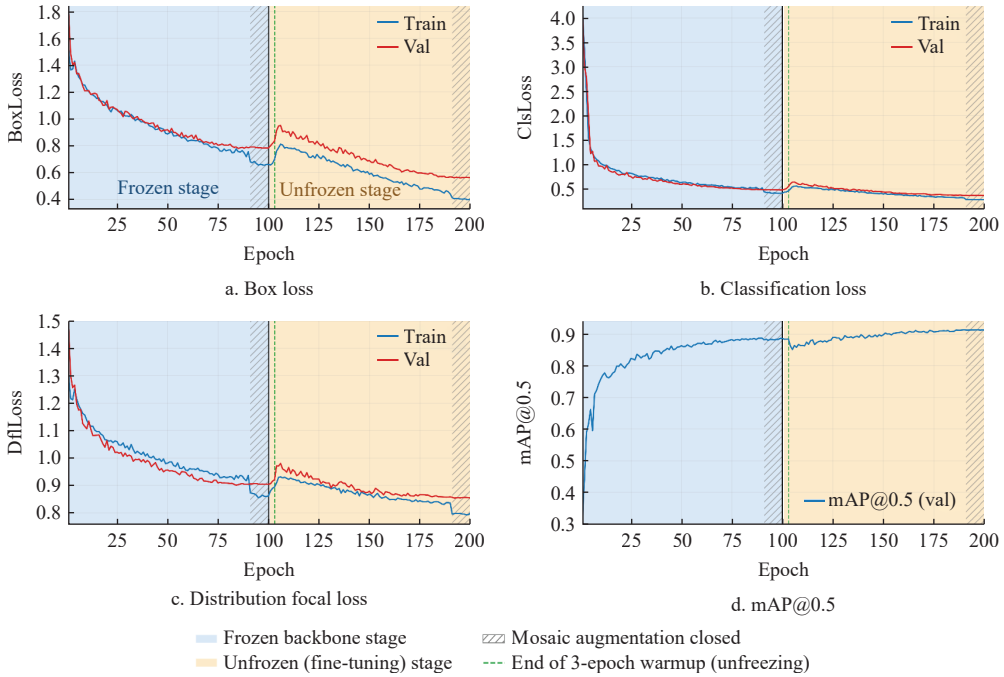


Figure 7 Training dynamics of the proposed model

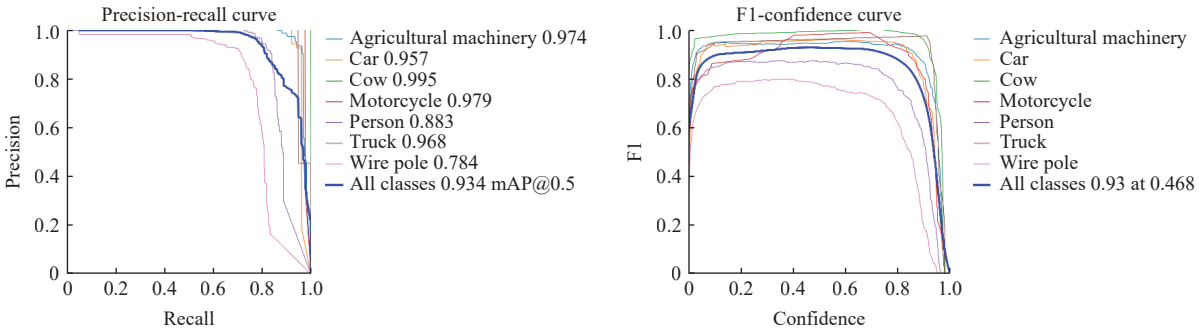
During the initial stage with frozen backbone weights, all loss components exhibit rapid convergence within the first 25 epochs. This rapid initial convergence indicates that the pre-trained features from COCO agricultural objects (five classes) provide a strong foundation for the target domain adaptation. Throughout most of stage 1, validation losses track the training loss closely, suggesting good generalization despite intensive mosaic augmentation (mosaic=1.0). The mAP@0.5 shows steady improvement, reaching approximately 0.87 by epoch 80 and staying stable. A notable change occurs in the final 10 epochs (91-100) when mosaic augmentation was closed. The training losses experience a sudden drop, particularly visible in BoxLoss and DflLoss. However, the validation losses and mAP@0.5 remain relatively stable during this period. This behavior is expected when close mosaic augmentation, the model can better fit the training data, but validation performance plateaus as the frozen backbone limits further adaptation capacity.

The second stage, initiated with backbone unfreezing and a three-epoch warmup, demonstrates more nuanced learning dynamics. The warmup period shows controlled gradient updates, preventing catastrophic forgetting of the learned representations. All loss components continue their downward trend. The reduced mosaic augmentation (0.5) during this stage provides a balance between regularization and convergence stability, as evidenced by the smoother loss curves compared to stage 1. The mAP@0.5 shows continued improvement, ultimately reaching and maintaining values above 0.91. The final 10 epochs (191-200) with closed mosaic augmentation show further refinement, with all metrics converging to their optimal values.

3.2 Quantitative analysis

The quantitative performance of our proposed model, the improved YOLO11 model with two-stage transfer learning, was evaluated on the test dataset using both Precision-Recall (*P*-*R*) and F1-Confidence curves. As shown in Figure 8a, the model achieves high precision across most recall levels, with the mAP@0.5 reaching 0.934. Among individual classes, cow and motorcycle achieved the highest precision with mAP@0.5 of 0.995 and 0.979, respectively, while wire pole had the lowest mAP@0.5 at 0.784, most likely due to its individual shape and difficulty to detect in the complex background in agricultural scenes.

The F1-confidence curves in Figure 8b illustrate the trade-off between precision and recall across different confidence thresholds. The model maintains high F1-score across a wide range of confidence thresholds. The peak F1-score for all classes is 0.93 at a confidence of 0.468, suggesting that the model strikes a good balance between precision and recall. Different object classes exhibit varying sensitivity to confidence thresholds. Classes such as cow, car, agricultural machinery, and motorcycle maintain high F1-scores across a broader range of confidence values, indicating more consistent detection performance. Wire pole remains the most challenging category.



a. Precision-recall curves for individual classes and overall performance (mAP@0.5 = 0.934)

b. F1-score vs. confidence threshold curves showing optimal performance at confidence = 0.468

Figure 8 Quantitative evaluation results of the improved YOLO11-s model with two-stage transfer learning

These quantitative results validate the effectiveness of the proposed methodology in accurately detecting diverse obstacles for intelligent agricultural machinery in complex rural environments. The improved YOLO11-s model with two-stage transfer learning achieves robust performance for large and medium-sized targets, while small or slender obstacles remain more challenging. High overall precision values (>0.95 for most classes) indicate a successful reduction of false positives, and balanced F1-scores demonstrate that the model maintains good recall while achieving high precision. Overall, these results confirm that the proposed approach effectively supports reliable obstacle perception for autonomous agricultural operations, facilitating practical deployment in real-world scenarios.

3.3 Ablation study

Ablation studies evaluated the contribution of each proposed module, loss function, and category-relevant two-stage transfer learning. Starting from a YOLO11-s baseline, this study incrementally integrated CBAM, SPPFCSPC, and MPDIoU loss. Table 2 reports precision, recall, mAP@0.5, mAP@0.5:0.95, parameter size, and FLOPs.

Table 2 Ablation study results on the custom test dataset

Model	Pretrain	BoxLoss	Precision	Recall	mAP@0.5	mAP@0.5:0.95	Params (M)	FLOPs (G)
Baseline	✘	Default	0.914	0.823	0.870	0.701	9.42	21.3
+CBAM	✘	Default	0.914	0.864	0.897	0.722	9.76	21.6
+SPPFCSPC	✘	Default	0.953	0.826	0.892	0.731	15.84	26.5
+CBAM + SPPFCSPC	✘	Default	0.951	0.844	0.898	0.751	16.19	26.7
+CBAM + SPPFCSPC	✘	MPDIoU	0.930	0.858	0.900	0.755	16.19	26.7
+CBAM + SPPFCSPC	✔	MPDIoU	0.964	0.895	0.934	0.828	16.19	26.7

The baseline achieved 0.914 precision, 0.823 recall, 0.870 mAP@0.5, and 0.701 mAP@0.5:0.95 (9.4M parameters, 21.3 GFLOPs). Adding CBAM raised recall to 0.864 and mAP@0.5 and mAP@0.5:0.95 to 0.897 and 0.722, respectively, with negligible complexity increase (9.8 M, 21.6 GFLOPs), confirming enhanced feature discrimination and background suppression. Replacing SPPF with SPPFCSPC boosted precision to 0.953 and mAP@0.5:0.95 to 0.731, albeit at higher computational cost (15.8 M, 26.5 GFLOPs), indicating that improved multi-scale aggregation benefits

variable-scale target detection in farmland scenes.

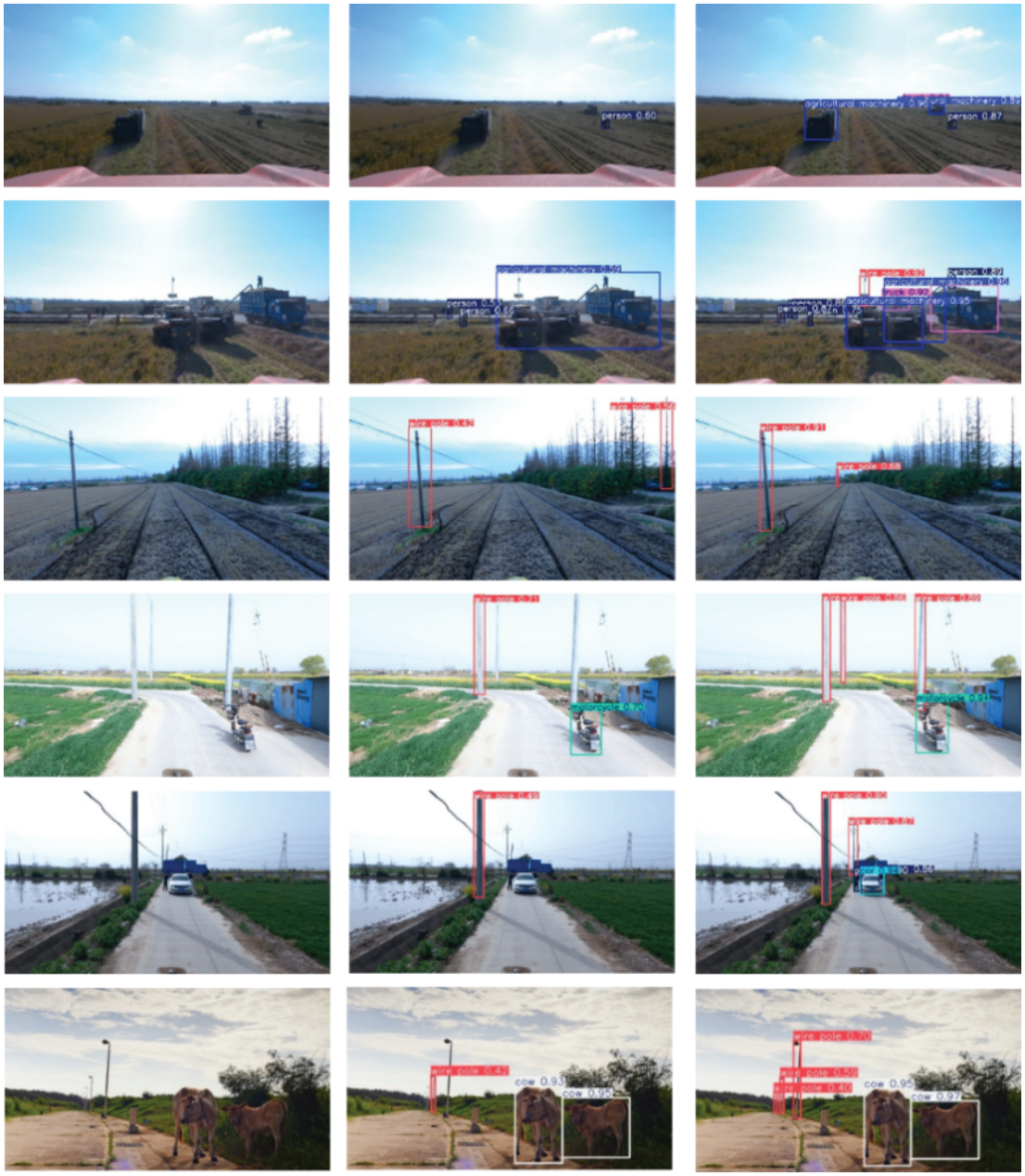
Combining CBAM and SPPFCSPC yielded 0.898 mAP@0.5 and 0.751 mAP@0.5:0.95 (16.2 M, 26.7 GFLOPs). Substituting MPDIoU loss further improved these metrics to 0.900 and 0.755, respectively. Although the improvement is relatively limited compared with other components, MPDIoU consistently improves localization accuracy under stricter IoU thresholds, indicating its effectiveness in refining bounding box regression. However, this improvement remains constrained when the model is trained solely on the small-scale agricultural dataset, since the quality of bounding box optimization is strongly dependent on the discriminative power of the learned feature representations. Under such conditions, optimizing the regression loss alone provides limited gains due to insufficient semantic feature richness.

After introducing the proposed task-aligned transfer learning strategy, the model achieved the best overall performance, with 0.964 precision, 0.895 recall, 0.934 mAP@0.5, and 0.828 mAP@0.5:0.95. Compared with the MPDIoU-only configuration, mAP@0.5 and mAP@0.5:0.95 increased by 3.4 and 7.3 percentage points, respectively. These results demonstrate a complementary relationship between transfer learning and MPDIoU, where pretraining enhances feature representation while MPDIoU improves bounding box regression, leading to substantially better performance than either component alone.

The ablation results clearly demonstrate that each proposed component contributes positively to model performance. More importantly, their joint integration with transfer learning yields the most significant gains in both mAP@0.5 and mAP@0.5:0.95, reflecting a complementary effect between feature representation learning and localization optimization.

3.4 Qualitative detection results

To further illustrate the effectiveness of the proposed model beyond quantitative metrics, this study performed qualitative comparisons with the baseline YOLO11-s on six representative scenarios: three from farmland and three from rural roads. As shown in Figure 9, column (a) presents the original images, column (b) shows the detection results of the baseline model, and column (c) displays the results of our proposed model.



a. Original images b. Detection results of the baseline YOLO11-s c. Detection results of the proposed model

Figure 9 Qualitative detection results

The baseline model occasionally suffers from missed detections in cluttered backgrounds and partial occlusions. In contrast, the proposed model consistently produced more accurate bounding boxes, successfully detecting both near and far field targets, and exhibited stronger robustness under conditions of occlusion, small objects, and cluttered backgrounds. After integrating CBAM, the model is better able to highlight relevant object regions, thus reducing false negatives. SPPFCSPC enhances detection stability in scenarios with large-scale variations, such as distant vehicles and small agricultural machinery. The combination of CBAM and SPPFCSPC further improves object completeness, particularly for elongated targets like wire poles. With MPDIoU, bounding boxes align more closely with object boundaries, improving localization accuracy. The pretrained variant exhibits the most robust detection capability, effectively handling complex farmland scenes with multiple overlapping obstacles. These qualitative results confirm that the proposed improvements not only enhance overall detection accuracy but also reduce false negatives and false positives, thereby providing more reliable perception for intelligent agricultural machinery in real-world environments.

3.5 Feature representation analysis

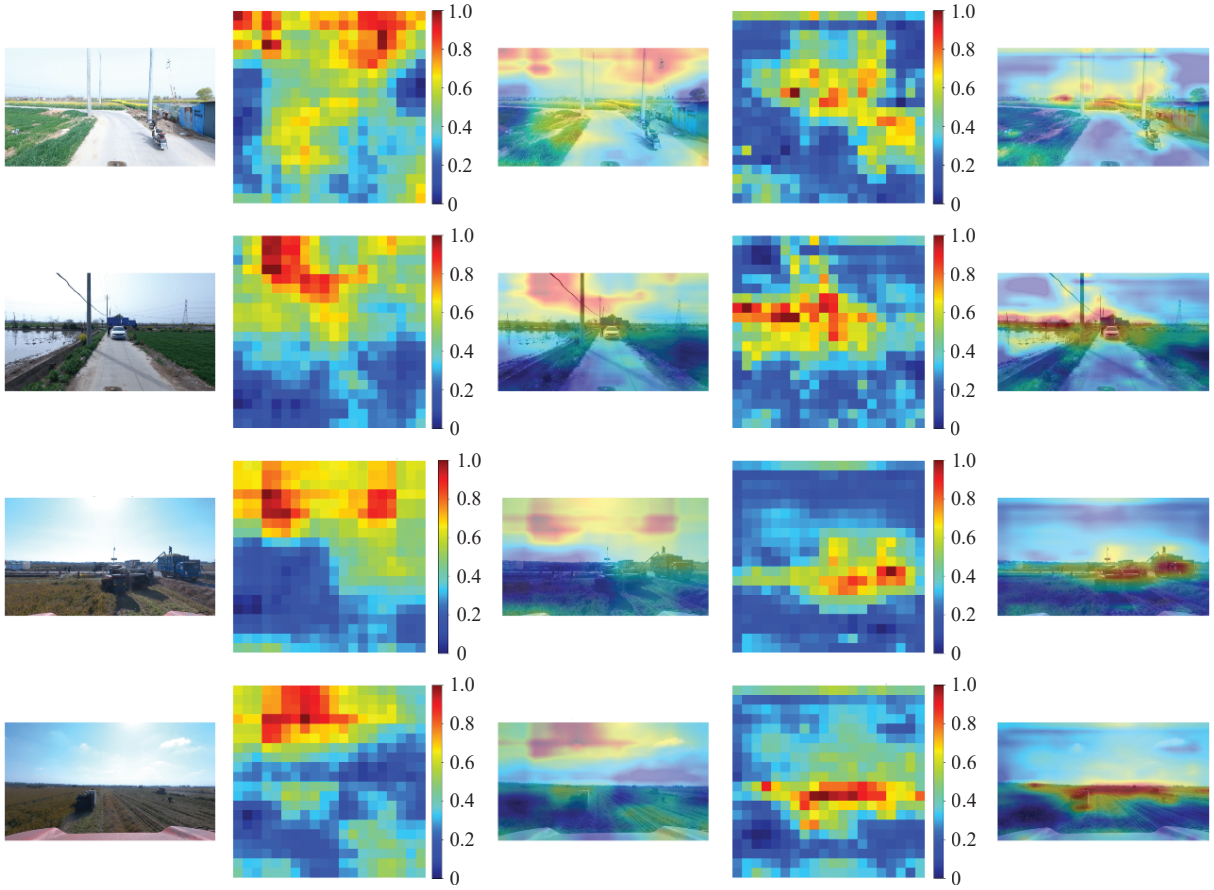
3.5.1 Feature preservation via SPPFCSPC

In YOLO11, backbone features are projected to a fixed low resolution (P5, 1/32 scale) following successive downsampling. Modules operating at this resolution, including spatial pyramid pooling and the subsequent C2PSA, rely predominantly on high-level semantics, while fine-grained structural cues, particularly for small or slender obstacles, are prone to degradation.

Figure 10 compares SPPF and SPPFCSPC activation patterns in farmland scenes. SPPF aggregates multi-scale context via repeated max-pooling, which enhances global receptive fields but often propagates strong background responses from homogeneous regions (e.g., sky, soil, crop textures), yielding spatially diffuse activations poorly aligned with obstacles (Figure 10b). In contrast, SPPFCSPC produces structured, target-aligned responses (Figure 10d): obstacle-related regions exhibit higher consistency while background activations are suppressed. This distinction is critical for small-scale and slender targets (wire poles, pedestrians, motorcycles), whose limited spatial extent and reliance on structural continuity render them vulnerable to dilution under standard pooling; SPPFCSPC preserves their spatial signatures more reliably.

This improvement stems from the Cross Stage Partial (CSP) design. By splitting feature propagation into parallel branches, CSP avoids uncontrolled accumulation of pooling-induced responses, enabling one stream to retain structural information without excessive contextual mixing. Subsequent fusion and 1×1 convolution adaptively rebalance pooled context and preserved structure, yielding more discriminative low-resolution representations.

For larger obstacles (machinery, trucks), SPPFCSPC generates activation regions that better correspond to object extents, providing clearer spatial delineation before C2PSA modeling. Overall, Figure 10 demonstrates that replacing SPPF with SPPFCSPC enhances feature robustness during low-resolution semantic aggregation, producing cleaner, more informative backbone features for subsequent attention refinement and multi-scale detection.



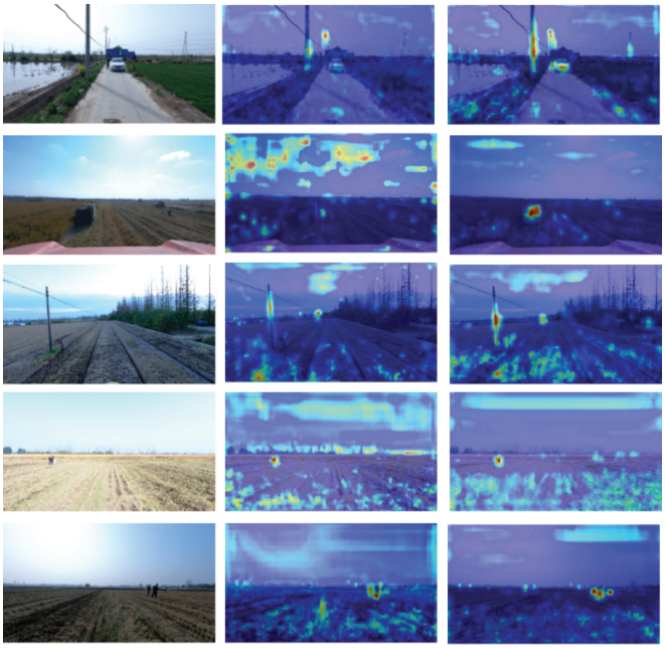
a. Original image b. SPPF feature activation (layer 9) c. Feature overlay d. SPPFCSPC feature activation (layer 9) e. Feature overlay

Note: SPPFCSPC demonstrates improved background suppression and more structured activations aligned with obstacle regions.

Figure 10 Feature activation comparison between SPPF and SPPFCSPC

3.5.2 Attention-guided feature refinement via CBAM

In the baseline YOLOv11, C3K2 modules at layers 16, 19, and 22 produce feature maps for the P3, P4, and P5 detection branches, respectively. In the proposed model, a Convolutional Block Attention Module (CBAM) is inserted immediately after each C3K2 module, yielding attention-refined features at layers 17, 21, and 25 before the detection heads. Figure 11 presents Grad-CAM++[33] visualizations comparing baseline C3K2 outputs with CBAM-enhanced features across multiple farmland scenes.



a. Original image b. Baseline (layers 16/19/22) c. CBAM-enhanced model (layers 17/21/25)

Figure 11 Grad-CAM++ visualization comparison between baseline C3K2 features and CBAM-enhanced features

Across all examples, CBAM consistently suppresses diffuse low-relevance activations while concentrating responses on obstacle regions. Baseline C3K2 features often exhibit scattered activations over background areas (e.g., roads, crops, waterlogged surfaces), indicating limited target–background discrimination. After CBAM integration, background-dominated responses are markedly reduced, and attention shifts toward semantically relevant regions.

For large objects (e.g., agricultural machinery, Row 2), CBAM condenses fragmented baseline activations into compact, high-response regions centered on the object. For small or distant targets (Rows 3-5), CBAM amplifies weak baseline responses into distinct activation peaks, improving perceptual separation from cluttered farmland backgrounds, critical in open-field environments where repetitive background textures can interfere with detection.

This behavior aligns with CBAM's sequential channel and spatial attention mechanisms, which reweight features based on global context and spatial saliency. By refining C3K2 outputs before prediction, CBAM steers the detection heads toward task-relevant features while suppressing irrelevant activations. The improved alignment between attention maps and obstacle locations in Figure 11 qualitatively supports the observed gains in detection accuracy, particularly for small and visually ambiguous obstacles.

3.6 Real-time test on Jetson

To evaluate real-time performance, all inference experiments were conducted on a Jetson AGX Orin platform. The detailed hardware and software configurations are summarized in Table 3. Real-time inference speed and detection accuracy (mAP) were evaluated on the customized test dataset described in Section 2.1.

Table 3 Hardware and software environment of the Jetson AGX Orin

Component	Specification
Platform	NVIDIA Jetson AGX Orin
JetPack Version	5.1.2
CUDA Version	11.4
TensorRT Version	8.5.2.2
PyTorch Version	2.0.0+nv23.5
CPU	12-core ARM Cortex-A78AE
GPU	2048-core NVIDIA Ampere architecture
VRAM	64 GB

As listed in Table 4, the PyTorch FP32 (32-bit float-point precision) model reached 44.3 FPS (22.6 ms total latency, mAP@0.5=0.934). After TensorRT optimization, the FP32 engine achieved 71.4 FPS (14.0 ms, mAP@0.5=0.928), and FP16 further increased speed to 82.6 FPS (12.1 ms, mAP@0.5=0.927) with a reduced model size of 36.7 MB. Preprocessing and postprocessing account for about 40% of total time, highlighting their impact on real-time efficiency. These results demonstrate that the model can achieve a good balance between accuracy and computational efficiency, suitable for on-board perception in intelligent agricultural machinery.

Table 4 Real-time performance of the proposed model on Jetson AGX Orin

Mode format	Precision mode	Size/MB	Latency/ms			FPS	mAP@0.5
			Preprocess	Inference	Postprocess		
Pytorch	FP32	32.8	0.3	13.7	8.6	44.3	0.934
TensorRT	FP32	68.0	1.0	9.7	3.3	71.4	0.928
TensorRT	FP16	36.7	1.1	7.3	3.7	82.6	0.927

3.7 Comparison with existing work

To provide a rigorous and fair comparison, we retrained and evaluated YOLOv5-s, YOLOv8-s, YOLOv10-s, YOLO11-n, YOLO11-s, YOLO11-m, and RT-DETR (ResNet-50 backbone) on our self-built farmland obstacle dataset, using the same training/validation/test split, input resolution (640×640), training schedule, and hardware platform (RTX A4000) to ensure a fair comparison. Table 5 reports Params, FLOPs, FPS, mAP@0.5, mAP@0.5:0.95, and Recall for all models, measured under the same testing environment.

Table 5 Comparison with mainstream detectors on the same dataset

Model	Params (M)	FLOPs (G)	FPS	mAP@0.5	mAP@0.5:0.95	Recall
YOLOv5-s	9.15	24.22	123.9	0.869	0.693	0.823
YOLOv8-s	11.17	28.82	133.9	0.873	0.710	0.841
YOLOv10-s	8.13	25.10	82.0	0.864	0.688	0.795
YOLO11-n	2.62	6.62	100.1	0.842	0.637	0.780
YOLO11-s	9.42	21.32	100.0	0.870	0.701	0.823
YOLO11-m	20.11	68.52	78.7	0.918	0.792	0.867
RT-DETR (ResNet-50)	31.01	108.34	32.8	0.939	0.799	0.912
Ours	**16.19**	**26.72**	**85.3**	**0.934**	**0.828**	**0.895**

As listed in Table 5, the proposed method achieves the second-highest mAP@0.5 (0.934) and the highest mAP@0.5:0.95 (0.828) among all compared models, trailing RT-DETR by only 0.5

percentage points in mAP@0.5 and by 1.7 percentage points in Recall, while surpassing it in mAP@0.5:0.95. This suggests that although RT-DETR detects a marginally higher proportion of ground-truth objects, our method produces more precisely localized bounding boxes overall. Notably, this level of accuracy is achieved while reducing the parameter count by 47.8%, the FLOPs by 75.3%, and increasing the inference speed by 160.1% relative to RT-DETR, indicating a substantially better accuracy-efficiency trade-off. Compared with the larger YOLO11-m model, our method achieves higher accuracy across all three metrics while using fewer parameters, fewer FLOPs, and running at a higher FPS, demonstrating a favorable balance between detection accuracy and computational efficiency suitable for deployment on edge platforms.

3.8 Limitations and future work

The proposed task-aligned transfer learning framework incorporating the CBAM attention mechanism, SPPFCSPC architecture, and MPDIoU loss function markedly improves agricultural object detection. However, several limitations remain. First, although the dataset integrates both custom-collected data and selected samples from public datasets, it still does not fully cover the full range of agricultural objects. Second, reliance on static images hinders performance in dynamic, real-world scenarios involving moving objects and time-varying conditions, critical for autonomous systems requiring real-time obstacle avoidance. Third, architectural enhancements increase computational complexity, impeding deployment on resource-constrained edge devices despite acceptable performance on platforms such as Jetson AGX Orin.

Future work will focus on several complementary conditions. Incorporating temporal information, such as optical flow[34,35] or video-based training[36], can improve the detection of moving and partially occluded objects in dynamic agricultural environments. Integrating LiDAR[37] and vision data can enhance robustness under challenging conditions. Further efforts will also explore tighter integration with agricultural machinery control systems to support real-time obstacle avoidance and improve operational safety and efficiency.

4 Conclusions

This study proposes a task-aligned transfer learning framework for agricultural obstacle detection based on an improved YOLO11 architecture, systematically addressing three key challenges: domain misalignment in transfer learning, insufficient architectural adaptation to complex agricultural environments, and suboptimal optimization strategies for balancing generic and domain-specific features. By combining domain-relevant pretraining on a curated COCO subset with targeted architectural enhancements, including CBAM and a redesigned SPPF module, and a staged fine-tuning strategy with MPDIoU loss, the proposed approach effectively improves detection accuracy and robustness under farmland conditions. Experimental results demonstrate that the model achieves strong detection performance on a seven-category agricultural test set while maintaining real-time inference capability on edge devices such as Jetson AGX Orin. Although the current dataset is relatively specific and does not cover all possible farmland scenarios, the proposed framework exhibits robust generalization within the target agricultural domain. Future work will focus on expanding and diversifying the dataset, integrating temporal information to better handle dynamic and partially occluded obstacles, and exploring multi-modal sensing strategies to further enhance the applicability of the system in real-world intelligent agricultural operations.

Acknowledgements

This work was supported by the National Key R&D Program of China (2022YFD2001501), in part by Laboratory of Lingnan Modern Agriculture Project (NT2025006), Science and Technology Plan of Guangdong Province of China (2023B10564002), the National Natural Science Foundation of China (32301707), the Natural Science Foundation of Guangdong Province (2024A1515012608), and Guangzhou Key Research and Development Program (2024B03J1356). The authors would like to thank the reviewers and editors for their constructive comments.

[References]

[1] Campos Y, Sossa H, Pajares G. Spatio-temporal analysis for obstacle detection in agricultural videos. Applied Soft Computing, 2016; 45(C): 86–97.

[2] Yu Y, Liu Y F, Wang J C, Noguchi N, He Y. Obstacle avoidance method based on double DQN for agricultural robots. Computers and Electronics in Agriculture, 2023; 204: 107546.

[3] Christiansen P, Nielsen L N, Steen K A, Jørgensen R N, Karstoft H. DeepAnomaly: Combining background subtraction and deep learning for detecting obstacles and anomalies in an agricultural field. Sensors, 2016; 16(11): 1904.

[4] Skoczeń M, Ochman M, Spyra K, Nikodem M, Krata D, Panek M, et al. Obstacle detection system for agricultural mobile robot application using RGB-D cameras. Sensors, 2021; 21(16): 5292.

[5] Ghazal S, Munir A, Qureshi W S. Computer vision in smart agriculture and precision farming: Techniques and applications. Artificial Intelligence in Agriculture, 2024; 13: 64–83.

[6] Wang T H, Chen B, Wang N, Ji Y H, Li H, Zhang M. Zero-shot obstacle detection using panoramic vision in farmland. Journal of Field Robotics, 2024; 41(7): 2169–2183.

[7] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 2005; pp.886–893. doi: 10.1109/CVPR.2005.177.

[8] Lowe D G. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 2004; 60(2): 91–110.

[9] Wang T H, Chen B, Zhang Z Q, Li H, Zhang M. Applications of machine vision in agricultural robot navigation: A review. Computers and Electronics in Agriculture, 2022; 198: 107085.

[10] Sharma A, Kumar V, Longchamps L. Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species. Smart Agricultural Technology, 2024; 9: 100648.

[11] Li Y, Li M, Qi J T, Zhou D Y, Zou Z F, Liu K. Detection of typical obstacles in orchards based on deep convolutional neural network. Computers and Electronics in Agriculture, 2021; 181: 105932.

[12] Zhang J C, Tian M M, Yang Z R, Li J H, Zhao L L. An improved target detection method based on YOLOv5 in natural orchard environments. Computers and Electronics in Agriculture, 2024; 219: 108780.

[13] Zhou X, Chen W, Wei X. Improved field obstacle detection algorithm based on YOLOv8. Agriculture, 2024; 14(12): 2263.

[14] Wang A, Chen H, Liu L H, Chen K, Lin Z J, Han J G, et al. YOLOv10: Real-time end-to-end object detection. Advances in Neural Information Processing Systems, 2024; pp.107984–108011.

[15] Kamilaris A, Prenafeta-Boldú F X. Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 2018; 147: 70–90.

[16] Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft COCO: Common objects in context. Computer Vision – ECCV, Springer, 2014; pp.740–755. doi: 10.1007/978-3-319-10602-1_48.

[17] Geiger A, Lenz P, Stiller C, Urtasun R. Vision meets robotics: The KITTI dataset. The International Journal of Robotics Research, 2013; 32(11): 1231–1237.

[18] Wang M, Deng W H. Deep visual domain adaptation: A survey. Neurocomputing, 2018; 312: 135–153.

[19] Zhuang F Z, Qi Z Y, Duan K Y, Xi D B, Zhu Y C, Zhu H S, et al. A comprehensive survey on transfer learning. Proceedings of the IEEE, 2021;

109(1): 43–76.

[20] Karki S, Basak J K, Tamrakar N, Deb N C, Paudel B, Kook J H, et al. Strawberry disease detection using transfer learning of deep convolutional neural networks. Scientia Horticulturae, 2024; 332: 113241.

[21] Cao B, Zhang B., Zheng W, Zhou J, Lin Y, Chen Y. Real-time, highly accurate robotic grasp detection utilizing transfer learning for robots manipulating fragile fruits with widely variable sizes and shapes. Computers and Electronics in Agriculture, 2022; 200: 107254.

[22] Ruchay A, Kolpakov V, Guo H, Pezzuolo A. On-barn cattle facial recognition using deep transfer learning and data augmentation. Computers and Electronics in Agriculture, 2024; 225: 109306.

[23] Hirahara K, Nakane C, Ebisawa H, Kuroda T, Iwaki Y, Utsumi T, et al. D4: Text-guided diffusion model-based domain adaptive data augmentation for vineyard shoot detection. Computers and Electronics in Agriculture, 2025; 230: 109849.

[24] Magistri F, Weyler J, Gogoll D, Lottes P, Behley J, Petrinic N, et al. From one field to another—Unsupervised domain adaptation for semantic segmentation in agricultural robotics. Computers and Electronics in Agriculture, 2023; 212: 108114.

[25] Bertoglio R, Mazzucchelli A, Catalano N, Matteucci M. A comparative study of Fourier transform and CycleGAN as domain adaptation techniques for weed segmentation. Smart Agricultural Technology, 2023; 4: 100188.

[26] Ma Y C, Zhang Z, Yang H L, Yang Z W. An adaptive adversarial domain adaptation approach for corn yield prediction. Computers and Electronics in Agriculture, 2021; 187: 106314.

[27] Chen L, Zou J X, Yuan Y, He H Y. Improved domain adaptive rice disease image recognition based on a novel attention mechanism. Computers and Electronics in Agriculture, 2023; 208: 107806.

[28] Wang C-Y, Mark Liao H-Y, Wu Y-H, Chen P-Y, Hsieh J-W, Yeh I-H. CSPNet: A new backbone that can enhance learning capability of CNN. Seattle: IEEE, 2020; pp.1571–1580. doi: 10.1109/CVPRW50498.2020.00203.

[29] Woo S, Park J, Lee J-Y, Kweon I S. CBAM: Convolutional block attention module. Computer Vision – ECCV 2018, Springer, 2018; pp.3–19.

[30] Zheng Z H, Wang P, Ren D W, Liu W, Ye R G, Hu Q H. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. IEEE Transactions on Cybernetics, 2022; 52(8): 8574–8586.

[31] Ma S, Xu Y. MPDIoU: A loss for efficient and accurate bounding box regression. arXiv: 2307.07662.

[32] Jocher G, Qiu J, Chaurasia A. Ultralytics YOLO. 2023. https://github.com/ultralytics

[33] Chattopadhay A, Sarkar A, Howlader P, Balasubramanian V N. Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe: IEEE, 2018; pp.839–847. doi: 10.1109/WACV.2018.00097.

[34] Baker S, Matthews I. Lucas-Kanade 20 years on: A unifying framework. International Journal of Computer Vision, 2004; 56: 221–255.

[35] Xu H Z, Li S C, Ji Y H, Cao R Y, Zhang M. Dynamic obstacle detection based on panoramic vision in the moving state of agricultural machineries. Computers and Electronics in Agriculture, 2021; 184: 106104.

[36] Cheng K Y, Zhu X S, Zhan Y Z, Pei Y S. Video deblurring and flow-guided feature aggregation for obstacle detection in agricultural videos. International Journal of Multimedia Information Retrieval, 2022; 11(4): 577–588.

[37] Rivera G, Porras R, Florencia R, Sánchez-Solís J P. LiDAR applications in precision agriculture for cultivating crops: A review of recent advances. Computers and Electronics in Agriculture, 2023; 207: 107737.