DF-YOLOv11: Lightweight Wheat Ear Detection Method Based on Improved YOLOv11

Xin Jiang [1], Delong Kong [2], Moughal Tauqir [3*], Jiahua Zhang [4*]
[1-4] Remote Sensing Information and Digital Earth Center, School of Computer Science and Technology, Qingdao University, Qingdao 266071, China

* Corresponding Author: **Moughal Tauqir, Jiahua Zhang**

Article Info

P-ISSN: 3051-3383
E-ISSN: 3051-3391
Volume: 06
Issue: 02
July - December 2025
Received: 18-10-2025
Accepted: 21-11-2025
Published: 15-12-2025
Page No: 191-204

Abstract
Rapid and accurate counting of wheat ears is a key link in field yield estimation and breeding evaluation. The Global Wheat Head Detection (GWHD) dataset is widely used for the evaluation of deep learning models. However, as it covers complex field scenarios involving multiple environments and varieties, it poses a severe challenge to the generalization ability of detection models. To address the limitations of existing methods on the GWHD dataset, including poor adaptability to multi-scale targets, weak ability to capture features of tilted wheat heads, and inaccurate localization in dense occlusion scenarios, this paper proposes a lightweight improved model based on YOLOv11, namely Dynamic Fusion YOLOv11 (DF-YOLOv11). Firstly, a Dynamic Multi-Scale Fusion (DMS-Fusion) module is designed in the neck network, which effectively coordinates the extraction of detailed and semantic features of wheat heads at different scales through scale-adaptive grouped convolution and a dual-branch enhancement structure. Secondly, a Lightweight Direction-Aware Attention Module (Light-OrientedECA) is embedded in the backbone network. Utilizing directional pooling tailored to the angular characteristics of the GWHD dataset and depthwise separable convolution, it significantly improves the recognition ability of tilted wheat heads with only a slight increase in computational cost. Finally, based on the statistical characteristics of the GWHD dataset, dedicated anchor boxes are generated via K-means clustering, and the CIoU loss function is improved by introducing an occlusion-aware weight (Occlusion-CIoU), enhancing the model's robustness in bounding box regression under overlapping and occluded scenarios. Comprehensive experiments on the GWHD test set show that DF-YOLOv11 achieves an mAP@0.5 of 87.3%, with only 1.57M parameters and 5.9 GFLOPs of computational complexity. Compared with the original YOLOv11n and mainstream comparative models such as YOLOv10n and YOLOv8n, DF-YOLOv11 achieves a better balance between detection accuracy, model complexity, and inference efficiency, exhibiting stronger robustness especially for small, tilted, and densely occluded targets. It provides a feasible technical solution for deployment of resource-constrained field mobile devices.

DOI: https://doi.org/10.54660/IJAIET.2025.6.2.191-204

Keywords: Wheat Ear Detection, Lightweight Model, YOLOv11, Attention Mechanism, Multi-Scale Feature Fusion, GWHD Dataset

1. Introduction

Wheat is one of the most widely cultivated food crops worldwide, and the stability and improvement of its yield are directly related to national food security (Cai Y *et al*, 2019; Peng Y *et al*, 2022) [1, 2]. As one of the core agronomic traits constituting yield, wheat ear number is a key parameter for achieving accurate yield estimation and guiding precision field management (Liu T *et al*, 2017; Ferrante A *et al*, 2017; Jin X *et al*, 2017) [3, 4, 5]. However, traditional manual field survey methods not only consume

significant human and material resources but also their results are susceptible to interference from investigators' subjective experience and environmental factors. Their efficiency and accuracy are difficult to meet the requirements of modern large-scale agricultural production (Redmon J *et al*, 2016; Qiu Z *et al*, 2025) [6, 7]. Therefore, developing efficient and accurate wheat ear detection technology holds great practical significance for promoting the development of smart agriculture.

In recent years, computer vision technology represented by deep learning has provided a powerful tool for the automated acquisition of agricultural phenotypic information (Singh A *et al*, 2018; Li Z *et al*, 2025; Cui D *et al*, 2025) [8-10]. Among them, the single-stage object detection algorithm YOLO (You Only Look Once) series has become one of the mainstream choices for field crop organ detection tasks due to its excellent balance between accuracy and speed (Lin Y *et al*, 2025; Bai B *et al*, 2024; Yang C *et al*,2024; Qiu Z *et al*, 2024; Yang B *et al*, 2021) [11-15]. To promote the development of this field, the international academic community has jointly released the Global Wheat Head Detection (GWHD) dataset. This dataset collects more than 6,000 high-quality wheat canopy images from 12 countries worldwide, covering diverse growth environments and genotypes, providing an authoritative testing platform for the development and evaluation of wheat ear detection models with strong generalization ability (Lin Y *et al*, 2025; Wang Y *et al*, 2021; David E *et al*, 2021) [11, 16, 17]. However, the GWHD dataset also truly reflects the complexity of field environments, mainly manifested in: (1) Significant scale variation, with targets ranging from negligible small ones to large ones occupying prominent areas of the image; (2) Posture diversity, as wheat ears exhibit rich tilted and rotated states due to differences in varieties, growth stages, and shooting angles; (3) Complex backgrounds, where wheat ears are often intertwined and overlapped with field elements such as leaves and stems, and are subject to environmental interference such as variable lighting and dust, resulting in widespread and severe occlusion (David E *et al*, 2021; Qing S *et al*, 2024) [17, 18]. These characteristics make many models that perform well on general datasets face challenges of reduced accuracy and insufficient generalization ability when directly applied to GWHD.

Currently, research on wheat ear detection for the GWHD dataset mainly focuses on the lightweight optimization and performance improvement of the YOLO series. Shen X *et al* (2023) [19] reconstructed the backbone network of YOLOv5s using ShuffleNetV2, which reduced the number of parameters but resulted in poor recall rate for small-scale wheat ears; Meng X *et al* (2023) [20] introduced the Convolutional Block Attention Module (CBAM) into YOLOv7, enhancing the expression of key features of wheat ears and suppressing redundant background information, but failed to fully balance the lightweight requirement, increasing the model complexity; Li J *et al* (2025) [21] integrated the Oriented-ECA attention mechanism into YOLOv11, improving the detection performance for tilted wheat ears, yet the module design did not fully consider computational efficiency; the WheatNet model proposed by Khaki S *et al* (2022) [22] has an exquisite structure, but its reliance solely on RGB single-modal data leads to insufficient robustness in complex field environments such as extreme lighting and weed occlusion. In summary, when addressing the complex challenges of the GWHD dataset, existing studies often struggle to simultaneously balance three crucial dimensions for field deployment: high accuracy, lightweight design, and high speed.

Based on this, this paper aims to develop a lightweight wheat ear detection model with balanced performance for the GWHD dataset. The main contributions of this paper are as follows: (1) Structural innovation: A Dynamic Multi-Scale Fusion (DMS-Fusion) module is designed, which achieves refined extraction and efficient fusion of features from wheat ears of different scales through a scale-adaptive convolution strategy and dual-branch feature enhancement paths; (2) Mechanism innovation: A Lightweight Direction-Aware Attention Module (Light-OrientedECA) is proposed. Combined with the angular distribution prior of wheat ears in the GWHD dataset, this module adopts directional pooling and depth wise separable convolution to enhance the model's sensitivity to directional features with low computational cost; (3) Optimization innovation: Based on the data-driven concept, dedicated anchor boxes are generated for the GWHD dataset through clustering, and the CIoU loss function is improved by introducing an occlusion-aware weight (Occlusion-CIoU). The model's localization accuracy and robustness in dense scenarios are effectively improved; (4) Comprehensive performance advantages: Through comprehensive experimental validation, the proposed Dynamic Fusion YOLOv11 (DF-YOLOv11) model achieves state-of-the-art detection accuracy on the GWHD test set, while demonstrating significant advantages in model size, computational complexity, and inference speed, indicating broad practical application prospects.

2. Materials and Methods

2.1. Experimental Data and Preprocessing

This study adopted 6,512 images from the 2021 version of the GWHD dataset, with a resolution of 1024×1024. These images contain over 300,000 unique wheat ears, each annotated with corresponding bounding boxes, covering multiple growth stages from the grain filling stage to the maturity stage. The dataset was randomly divided into a training set (3,655 images), a validation set (1,476 images), and a test set (1,381 images).

To enhance the model's generalization ability and suppress overfitting, a rigorous data augmentation process was implemented during training: (1) Geometric transformations, including random horizontal flipping (probability of 0.5), random rotation (-45° to +45°), and random scaling (0.8 to 1.2 times), to simulate variations in shooting angles and wheat ear postures; (2) Photometric distortion: Random perturbations (amplitude: ±10% to ±20%) were applied to the hue (H), saturation (S), and value (V) channels of images in the HSV color space, improving the model's adaptability to different lighting conditions; (3) Occlusion simulation: Several rectangular occlusion blocks were randomly overlaid on images to simulate partial occlusion caused by leaves, straw, or other impurities, enhancing the model's ability to detect incomplete targets.

2.2. DF-YOLOv11 Model

The DF-YOLOv11 model proposed in this study is built based on the lightweight YOLOv11n (Khanam R, & Hussain M, 2024; He L *et al*, 2025) [23, 30]. To target address the core challenges of wheat ear targets in the GWHD dataset, including slender morphology, multi-scale variation, diverse orientations, and dense occlusion, the model incorporates

synergistic enhancements to its three key components: the backbone, neck, and detection head. Its overall architecture is illustrated in Fig. 2.

In the backbone network, to address the tilted posture of wheat ears, a lightweight oriented attention module is embedded. Through a direction-adaptive pooling strategy and depth wise separable convolution, this module enhances the network's ability to capture features of different directions (e.g., horizontal, left-tilted, right-tilted) with extremely low computational cost.

In the neck network, to tackle significant scale variations, a dynamic multi-scale feature fusion module is designed. Based on the target scales (large, medium, small) corresponding to the input feature maps, this module adaptively adjusts the convolution grouping ratio and feature enhancement strategy, achieving efficient balanced fusion of detailed information and semantic context.

Fig 1: Sample Images of the GWHD Dataset

In the detection head, a data-driven optimization strategy is adopted. Firstly, K-means clustering is used to generate dedicated anchor boxes adapted to the slender morphology of wheat ears, improving the matching efficiency in the early training stage. Secondly, the loss function is improved by introducing an occlusion-aware weight, which strengthens the localization constraints on dense overlapping targets, thereby enhancing the model's robustness in complex scenarios.

2.2.1. Lightweight Direction-Aware Attention Module (Light-OrientedECA)

Wheat ears in the GWHD dataset exhibit diverse natural growth postures, with many samples in non-vertical tilted states. This poses a challenge to the direction-agnostic feature extraction of general convolutional neural networks. To address this issue, this paper incorporates the Light-Oriented ECA attention module into the key layers of the backbone network.

The core innovation of this module lies in its design that combines structural priority with efficient computation. As shown in Fig. 3, the module abandons the traditional approach of adding complex branches for explicit angle prediction; instead, it encodes the statistical regularity of wheat ear tilt angles in the GWHD dataset (which can be clustered into three categories: horizontal, right-tilted, and left-tilted) into the module's internal structure. For different directional categories, the module adopts customized pooling strategies: for the horizontal direction, parallel adaptive horizontal and vertical pooling are used to capture axial features; for tilted directions, fixed pooling kernels are employed to simulate diagonal receptive fields. This design enables the module to directly enhance the responses of corresponding directional patterns in the input feature maps in a hardware-friendly and deterministic manner.

In achieving efficient attention weight generation, the module implements two key optimizations. Firstly, it uses channel averaging instead of complex full-channel interaction computations, compressing multi-dimensional feature maps into a representative spatial energy map, which significantly reduces the dimensionality of subsequent processing. Secondly, it adopts lightweight GSConv (Li J *et al*, 2025) [21] instead of standard convolution to transform the energy map. Through grouping operations, GSConv drastically reduces the number of parameters and computational complexity while retaining the necessary ability to model spatial relationships. Finally, the spatial attention map generated by activation via the Sigmoid function is element-wise multiplied with the original features, achieving adaptive enhancement of target directional features and suppression of background noise.

This design offers multiple advantages. In terms of computational efficiency, by avoiding large-size convolutions and complex full-channel connections, the module maintains extremely low parameters and floating-point operations (FLOPs), realizing the lightweight of the

"attention" mechanism. In terms of functional targeting, its embedded directional priorities allow it to more accurately focus on the main axial features of wheat ears rather than dispersing attention to irrelevant leaves or background textures, which is particularly effective for distinguishing tilted wheat ears in dense scenarios. In terms of engineering deployment, deterministic operations (pooling, GSConv) ensure excellent operator and platform compatibility.

In DF-YOLOv11, three Light-OrientedECA modules are embedded in the backbone network in a hierarchical manner. Shallow modules process basic horizontal features, while middle and deep modules focus on right-tilted (Tilt_A) and left-tilted (Tilt_B) features, respectively. This progressive deployment enables the network to start with extracting low-level directional edges and gradually form robust representations of complex wheat ear postures, thereby improving the model's overall recognition accuracy and robustness for directional targets and providing a more discriminative feature foundation for subsequent detection tasks.

2.2.2. Dynamic Multi-Scale Fusion (DMS-Fusion)

To address the challenge of significant scale variation in wheat ears within the GWHD dataset, this paper designs a Dynamic Multi-Scale Fusion (DMS-Fusion) module in the neck network. The core of this module lies in achieving more efficient and discriminative feature fusion through scale-aware convolution configuration and multi-path feature enhancement.

As shown in Fig. 4, the module first sets the ratio of grouped

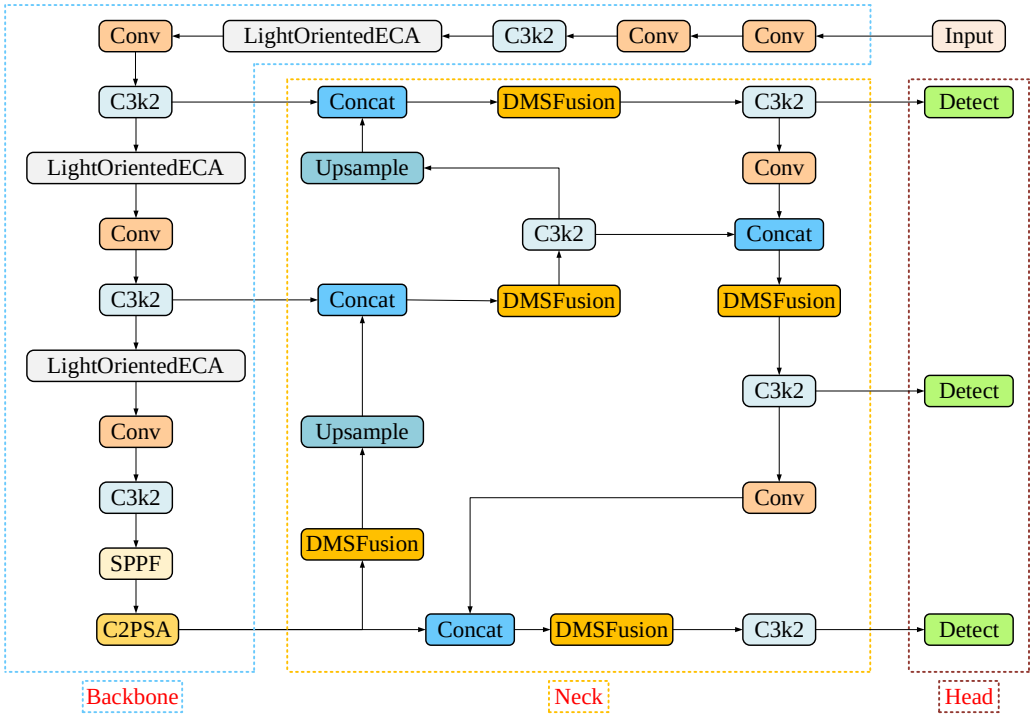

Fig 2: Architecture of the DF-YOLOv11 Model

convolution according to its deployment position (corresponding to P3 for small-scale, P4 for medium-scale, and P5 for large-scale features). For small-scale features, a smaller grouping ratio is adopted to retain more channel interactions and prevent detail loss; for large-scale features, a larger grouping ratio is used to reduce computational redundancy. Three parallel paths are constructed inside the module: a basic GSConv path for feature transformation; a dual-branch path that uses 1×1 and 3×3 convolutions to extract local details and global semantics, respectively; and a lightweight attention path for feature recalibration. Features from the three paths are ultimately fused and output.

The advantages of this design lie in three aspects: In terms of computation, rational allocation of computational resources is achieved through dynamic adjustment of the grouping ratio; in terms of representation, the multi-path structure balances details and semantics, enhancing the feature extraction capability for targets of different scales; in terms of structure, the modular design facilitates integration into existing feature pyramids, forming an adaptive multi-scale fusion network. In DF-YOLOv11, the DMS-Fusion module is deployed at key fusion nodes of the neck network, which effectively improves the detection performance for multi-scale wheat ears, especially small-scale and dense targets.

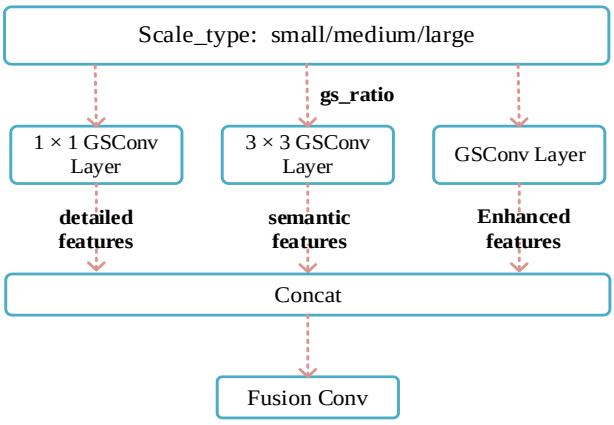

Fig 3: Structure Diagram of the DMS-Fusion Module

2.2.3. Dedicated Anchor Boxes and Occlusion-CIoU Loss Function

To improve detection performance in dense scenarios, two targeted optimizations are applied to the detection head. Firstly, general anchor boxes are abandoned. Based on the annotated data of the GWHD training set, the typical slender morphological characteristics of wheat ears (with an aspect ratio of approximately 1:2.3) are obtained through K-means clustering analysis. Three sets of dedicated prior anchor box sizes, [12, 28], [16, 32], and [20, 36], are set accordingly. This significantly enhances the matching degree between predicted boxes and ground truth boxes in the early training stage, accelerating model convergence.

Secondly, to address the severe overlap and occlusion between wheat ears, the bounding box regression loss function is improved. Based on CIoU loss (Equation 1) (Zheng Z *et al*,2022) [24], an occlusion-aware weight is introduced to construct Occlusion-CIoU loss (Equation 2). By quantifying the degree of occlusion (O) of the target by other predicted boxes and imposing stronger regression penalties on highly occluded targets (weight coefficient β= 0.3), this mechanism forces the model to pay more attention to hard samples, thereby achieving more accurate and discriminative localization results in dense regions.

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b,b_{gt})}{c^2} + av \quad (1)$$

Where: IoU denotes the Intersection over Union between the predicted box and the ground truth box; $\rho^2(b, b_{gt})$ represents the squared Euclidean distance between the center point of the predicted box (b) and the center point of the ground truth box (b_{gt}); c is the diagonal length of the minimum enclosing rectangle that contains both the predicted box and the ground truth box; av is the aspect ratio penalty term.

$$L_{Occlusion-CIoU} = 1 - IoU + \frac{\rho^2(b,b_{gt})}{c^2} + av + \beta \cdot O \quad (2)$$

Where: O is the Occlusion Rate, which quantifies the degree of target occlusion and is calculated based on the overlap of annotation boxes; β is the occlusion penalty coefficient, set to 0.3 in this study.

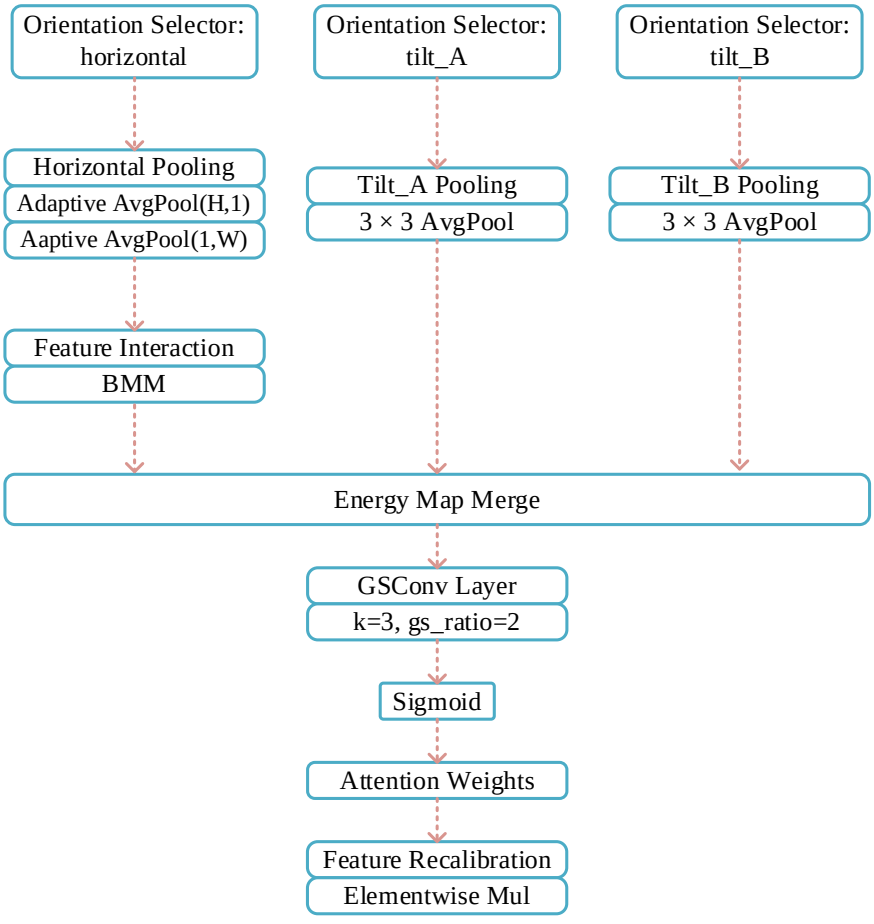

Fig 4: Structure Diagram of the Light-OrientedECA Module

2.3. Experimental Settings and Evaluation Metrics

All experiments were conducted on a computing platform equipped with an NVIDIA GeForce RTX 1050 GPU (4 GB video memory). A deep learning environment was established with Python 3.9 and PyTorch 1.13.0, supplemented by the CUDA 11.6 acceleration library. Although this hardware configuration has limited video memory, it can still support the training and inference of deep learning tasks at a certain scale after optimization.

The model training was performed for 100 epochs, with the Adam optimizer adopted for parameter updates and an initial learning rate set to 0.01. To promote stable convergence during training, a cosine annealing learning rate scheduling strategy without warm-up was employed. The momentum and weight decay were set to 0.937 and 0.0001, respectively, to effectively alleviate model overfitting. The parameters of the backbone network remained trainable to ensure the integrity of the end-to-end learning process.

Restricted by the GPU video memory capacity, the training batch size was set to 2. To improve training efficiency and maintain model performance under limited video memory conditions, PyTorch Automatic Mixed Precision (AMP) training was enabled in this experiment. It allows computations to be performed with FP16 precision during forward propagation, while maintaining numerical stability through a dynamic loss scaling mechanism, thereby significantly reducing video memory usage without compromising training accuracy.

2.3.1. Precision

Precision is used to measure the proportion of truly positive examples among all samples predicted as positive by the model, reflecting the model's ability to control false positives (i.e., misclassifying negative examples as positive ones) (Li J *et al*,2024) [25]. The formula is as follows:

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Where: TPdenotes True Positive; FP denotes False Positive.

2.3.2. Recall

Recall represents the proportion of truly positive examples successfully identified by the model among all actual positive examples, and it is a key indicator for measuring false negatives (i.e., true positive examples that fail to be detected) (Li J *et al*,2024) [25]. The formula is as follows:

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

Where: FN denotes False Negative.

2.3.3. Mean Average Precision (mAP)

mAP is a standard metric for evaluating the performance of object detection models, reflecting the comprehensive precision of the model across all target categories (Li J *et al*,2024) [25]. In this study, two variants are reported: ①mAP@0.5: The average precision value across all categories when the Intersection over Union (IoU) threshold is set to 0.5; ②mAP@0.5:0.95:The mean Average Precision following the COCO evaluation criteria, i.e., evaluation is performed at multiple IoU thresholds ranging from 0.5 to 0.95 (with a step size of 0.05), reflecting the model's comprehensive detection performance under different localization accuracy requirements.

To ensure the robustness and reproducibility of model performance improvements, a rigorous statistical evaluation process was adopted in this study. Each detection model was independently trained with three different random seeds, and the final performance metrics are reported as the mean ± standard deviation of the three experimental results. To further quantify the uncertainty of the results, we performed 1000 resampling on the mean Average Precision using the Bootstrap method and calculated the 95% confidence interval accordingly. All improved models were subjected to paired statistical tests against the YOLOv11 baseline model. The performance improvement is considered statistically significant only when the p-value of the hypothesis test is less than 0.05. This evaluation strategy not only enhances the reliability of the experimental results but also provides a reproducible statistical foundation for subsequent research.

3. Results and Analysis

3.1. Evaluation and Comparative Analysis of DF-YOLOv11's Overall Performance

To comprehensively evaluate the performance of the DF-YOLOv11 model in complex field scenes, systematic experiments were conducted on the GWHD test set, with comparisons against several mainstream lightweight object detection models (YOLOv7 (Li S *et al*, 2023) [26], YOLOv8n (Li F *et al*, 2025) [27], YOLOv9t (Wang C *et al*, 2025) [28], YOLOv10n (Wang A *et al*, 2024) [29], YOLOv11n (He L *et al*, 2025) [30]). All models were trained and tested under the same experimental environment, data preprocessing, and augmentation strategies to ensure comparability and fairness of the results.

Table 1 presents the comprehensive performance metrics of DF-YOLOv11 and multiple benchmark models on the GWHD test set. DF-YOLOv11 achieves 87.3% in mAP@0.5, which significantly outperforms the original YOLOv11n (83.1%), YOLOv10n (81.4%), and YOLOv8n (81.2%). Meanwhile, DF-YOLOv11 has a parameter count of only 1.57M and a computational complexity of 5.9 GFLOPs. While maintaining state-of-the-art detection accuracy, the model complexity and computational cost are strictly controlled.

DF-YOLOv11 achieves the highest mean Average Precision (mAP@0.5) of 87.3%, representing an increase of 4.2 percentage points compared to its baseline model YOLOv11n. This improvement can be attributed to the dynamic multi-scale feature fusion module (DMS-Fusion) and lightweight oriented attention module (Light-OrientedECA) introduced in the model. By adopting a scale-adaptive grouped convolution strategy, the DMS-Fusion module dynamically adjusts the fusion method of different feature layers, significantly enhancing the feature extraction capability for wheat ears with substantial scale variations in the GWHD dataset. In particular, it improves the detailed preservation of small-scale targets, which is the key to the significant enhancements in its mAP and Precision (90.7%). Meanwhile, the Light-OrientedECA module encodes the priors of common wheat ear postures in the field (horizontal, left-tilted, and right-tilted) into the attention mechanism. Through directional pooling and lightweight GSConv, it strengthens the axial feature responses of wheat ears in non-vertical postures, effectively reducing misclassifications caused by target tilt and further contributing to the accuracy improvement.

The Recall of DF-YOLOv11 reaches 80.7%, which also ranks first among all comparative models, indicating its strongest ability to detect true positive samples. This advantage is mainly attributed to the improvement of the Occlusion-CIoU loss function. Traditional CIoU loss has limited ability to distinguish overlapping targets in dense occlusion scenarios. The occlusion-aware weight introduced in this paper can dynamically adjust the regression penalty intensity according to the degree of target occlusion by other predicted boxes, forcing the model to pay more attention to hard samples with severe overlap during training. This significantly reduces the number of missed detections in dense regions and directly drives the improvement of Recall.

While achieving dual breakthroughs in precision and recall, the model complexity of DF-YOLOv11 is further reduced. Its parameter count (1.57 M) and computational complexity (5.9 GFLOPs) are the lowest in the table. The first stems from the overall lightweight design philosophy, for instance, by using

depthwise separable convolution and GSConv instead of standard convolution in the Light-OrientedECA module, which greatly reduces parameters and computational overhead while enhancing direction-aware capability. Secondly, the dedicated anchor boxes generated based on K-means clustering ([12, 28, 16, 32, 20, 36]) are highly matched with the typical slender morphology of wheat ears. This improves the matching efficiency between predicted boxes and ground truth boxes in the early training stage, reduces many invalid anchors box computations, and optimizes the computational flow from the source. These features make DF-YOLOv11 more practical in field deployment scenarios with limited computational resources.

In summary, the comparative experimental results show that through its targeted designs of the DMS-Fusion module, Light-OrientedECA module, and optimizations to the loss function and anchor boxes, DF-YOLOv11 not only effectively addresses core challenges in wheat ear detection such as multi-scale variation, diverse postures, and dense occlusion, achieving a significant leap in detection accuracy, but also simultaneously achieves high model lightweighting. It sets a new benchmark in the trade-off between accuracy and efficiency.

3.2. DF-YOLOv11 Detection Results in Diverse Scenarios

To further evaluate the practical performance of the DF-YOLOv11 model under various complex field scenarios, 12 representative images were selected from the test set for qualitative analysis. As shown in Fig. 5, in the visualized detection results: red bounding boxes indicate wheat ears correctly detected by the model; yellow bounding boxes denote duplicate detections of the same target (i.e., a single wheat ear marked by multiple bounding boxes); and black bounding boxes represent missed wheat ears that the model failed to detect. Observation of the images shows that missed detections are mainly concentrated in three types of scenarios: (1) targets with extremely small scales, whose pixel proportion is significantly lower than that of common wheat ears in the dataset; (2) targets located at image edges with severe truncation and incomplete morphology; (3) targets completely covered by multiple adjacent wheat ears in densely occluded regions.

Table 1: Performance Comparison of DF-YOLOv11 with Multiple Benchmark Models on the GWHD Test Set

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	Parameters (M)	FLOPs (G)
YOLOv7	82.5	73.2	80.1	6.19	13.8
YOLOv8n	84.6	75.6	81.2	3.19	8.7
YOLOv9t	83.8	75.8	80.9	2.04	7.7
YOLOv10n	83.5	74.3	81.4	2.83	6.9
YOLOv11n	87.8	77.6	83.1	2.59	6.5
DF-YOLOv11	90.7	80.7	87.3	1.57	5.9

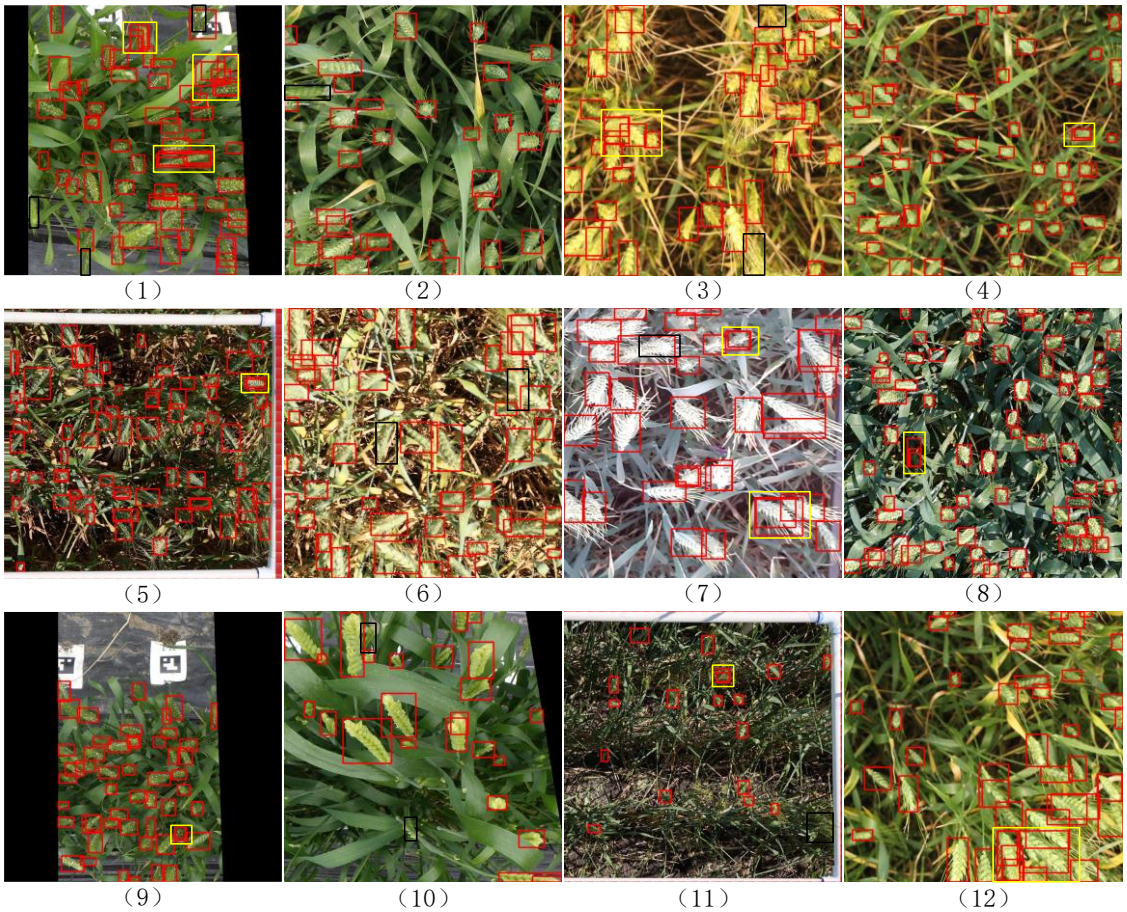

Fig 5: Detection Performance of the DF-YOLOv11 Model Under Various Complex Field Scenarios

From the perspective of model structure, the main reasons for these missed detections are as follows: Firstly, although the model has introduced the DMS-Fusion module to improve multi-scale adaptability, its design is mainly optimized for the scale distribution within the statistical range of the training data (obtained through K-means clustering). For extremely small targets in the tail of the scale distribution, their features may have been excessively attenuated during the downsampling process of the backbone network, resulting in overly weak corresponding semantic information in the deep feature maps, which is difficult to be effectively captured by the detection head. Secondly, for truncated targets at the image edges, their morphology differs significantly from the complete wheat ear samples in the training set. Although the current model simulates partial cropping through data augmentation, the axial features (such as the main ear axis) relied on by the Light-OrientedECA may no longer exist in truncated targets, leading to the failure of the attention mechanism to be effectively activated and thus being suppressed by background noise. Finally, when dealing with extremely dense occlusion, although the Occlusion-CIoU loss function strengthens the regression constraint on overlapping targets through occlusion-aware weights, in the feature extraction stage, severely overlapping wheat ears are highly mixed at the pixel level, making it difficult for the model to recognize them as two independent entities. The fused features of the neck network may be more inclined to respond to the most prominent or complete contours, resulting in the complete occluded targets "disappearing" in the feature space.

Duplicate detection mainly occurs in two types of scenarios: (1) Wheat ears that are extremely slender in morphology or have a large tilt angle in posture; (2) Individual wheat ears with complex inherent textures or obvious grain clustering. The structural roots are as follows: Firstly, for slender or large-tilt wheat ears, they occupy a large spatial proportion and may span receptive fields of multiple scales in the Feature Pyramid Network (FPN). When the DMS-Fusion module performs multi-scale feature fusion, the same target may be encoded with differentiated contextual information on feature maps of different scales, thereby being separately activated in multiple detection branches and generating predicted boxes. Although Non-Maximum Suppression (NMS) post-processing is designed to eliminate redundant boxes, these predicted boxes from different scales may still be retained as multiple detection results when they have a certain offset in position and all have high confidence scores. Secondly, for wheat ears with complex textures or obvious clustering, their local features (such as grain clusters) may have strong independence. The Light-OrientedECA module aims to strengthen global axial features, but for such individuals with prominent local features, attention maps may generate peaks in multiple high-response regions. This may cause the model to misjudge different local feature regions of the same wheat ear as evidence of multiple independent targets, thereby generating multiple bounding boxes in the decoding stage.

In summary, through in-depth analysis of missed detection and duplicate detection cases across multiple scenarios, this study reveals that the core modules of the current DF-YOLOv11 model still have certain limitations when addressing complex situations beyond the designed priority, such as extreme scale distribution, edge-truncated morphologies, extreme dense overlapping, and complex local textures. This points out potential optimization directions for future research, including introducing a more refined pyramid network design to preserve features of extremely small targets, developing a more robust deformable attention mechanism to handle incomplete targets, and exploring instance-level feature decoupling to distinguish highly overlapping individuals.

3.3. Ablation Studies

To accurately quantify the effectiveness of each improved module proposed in this paper, we designed cumulative ablative experiments with YOLOv11n as the baseline. As shown in Table 2, the experiments sequentially evaluated the independent contributions and combined effects of the Light-OrientedECA, DMS-Fusion, and Occlusion-CIoU loss function. All models were evaluated on the GWHD test set to ensure the fairness of the comparison.

After embedding the Light-OrientedECA module into the backbone network, the model's mAP@0.5 increased from 83.1% to 85.5%, with a growth of 2.4 percentage points. Meanwhile, the parameters and computational complexity decreased by 0.32M and 0.3 GFLOPs, respectively. This verifies its lightweight and efficient design philosophy. By integrating directional pooling and lightweight GSConv, the module explicitly enhances the network's sensitivity to the tilt posture features of wheat ears, effectively reducing false detections caused by variable target orientations. Consequently, it significantly improves classification accuracy (Precision increased by 1.1%) and achieves stronger feature discriminative ability with fewer computational resources.

After introducing the DMS-Fusion module, the model further improved the Recall (from 77.6% to 78.6) while achieving significant model compression, with the parameter count reduced to 1.49M and the computational complexity to 5.8 GFLOPs. This reflects its advantages of scale adaptability and structural simplification. By dynamically adjusting the grouped convolution ratio and enhancing multi-path features, the module optimizes the efficiency of multi-scale feature fusion, particularly strengthening the feature extraction capability for small and medium-sized targets. It effectively alleviates missed detections caused by scale differences, laying a crucial foundation for model lightweighting. After replacing it with the Occlusion-CIoU loss function, the model achieved the highest single-point mAP@0.5 (86.0%) and Precision (89.4%) among the three improvements. This enhancement directly targets dense occlusion scenarios. The introduced occlusion-aware weight mechanism forces the model to focus more on optimizing the bounding box regression of severely overlapping targets during training. This directly and effectively improves the model's localization accuracy and classification confidence in complex dense regions, serving as one of the core drivers for the breakthrough in detection precision.

Module Integration and Comprehensive Performance Analysis: The complete DF-YOLOv11 model constructed by integrating the three modules achieves the optimal comprehensive performance: mAP@0.5 reaches 87.3%, Recall significantly increases to 80.7%, and the final parameter count (1.57M) and computational complexity (5.9 GFLOPs) are both lower than those of the baseline model. The results indicate that each module is not only independently effective but also capable of generating synergistic effects: Light-OrientedECA and DMS-Fusion lay the foundation for lightweight and high-precision

performance from the perspectives of feature discrimination and fusion efficiency, respectively, while Occlusion-CIoU enhances the ability to tackle the most challenging samples by optimizing target regression. The synergy of the three modules enables DF-YOLOv11 to achieve a new optimal balance among accuracy, robustness, and efficiency.

3.4. Analysis of Convergence Dynamics and Performance Behavior

3.4.1. Analysis of Convergence Characteristics and Performance Behavior between DF-YOLOv11 and Mainstream Models

To gain a deeper understanding of the training dynamics and convergence characteristics of each model, we further plotted comparative graphs of training loss curves (Fig. 6) and comprehensive performance analysis diagrams (Fig. 7). This visualization analyzes not only verify the quantitative performance advantages of DF-YOLOv11 but also reveals its behavioral characteristics during the training process.

Fig. 6 presents the comparison of loss curves for six models over 200 training epochs. Among them: Fig. 6(a) shows the full-epoch loss curves, clearly demonstrating that the loss value of DF-YOLOv11 is consistently lower than that of other comparative models throughout the training process, exhibiting a superior optimization trajectory. Fig. 6(b) focuses on the early training stage (1-50 epochs), where the loss of DF-YOLOv11 decreases significantly faster, indicating that its improved module structure can more effectively capture key features from the initial stage. Fig. 6(c) displays the middle training stage (50-150 epochs), where the loss differences among various models gradually narrow, but DF-YOLOv11 still maintains the lowest loss level. Fig. 6(d) presents the convergence stage (150-200 epochs), where the loss curve of DF-YOLOv11 is the most stable and stabilizes at the lowest level, demonstrating excellent training stability and generalization ability.

Fig. 7 conducts a comprehensive analysis of model performance from three dimensions. Fig. 7(a) illustrates the negative correlation between the final loss value and mAP. DF-YOLOv11 is located at the bottom right corner of the graph, i.e., it simultaneously achieves the highest mAP (87.3%) and the lowest final loss, verifying the consistency of its optimization objectives. Fig. 7(b) compares the number of training epochs required for each model to reach a specific loss threshold. DF-YOLOv11 requires the fewest epochs at every loss threshold, indicating its fastest convergence speed. Fig. 7(c) presents the three-dimensional relationship of model efficiency through a bubble chart: the position of each bubble represents the trade-off between parameter count and final loss, and the size of the bubble indicates computational complexity (FLOPs). DF-YOLOv11 is located at the bottom

Table 2: Performance Comparison of DF-YOLOv11 with Multiple Benchmark Models on the GWHD Test Set

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	Parameters (M)	FLOPs (G)
YOLOv11n	87.8	77.6	83.1	2.59	6.5
+ Light-OrientedECA	88.9	78.0	85.5	2.27	6.2
+ DMS-Fusion	88.2	78.6	84.7	1.49	5.8
+ Occlusion-CIoU	89.4	78.3	86.0	2.31	6.2
DF-YOLOv11	90.7	80.7	87.3	1.57	5.9

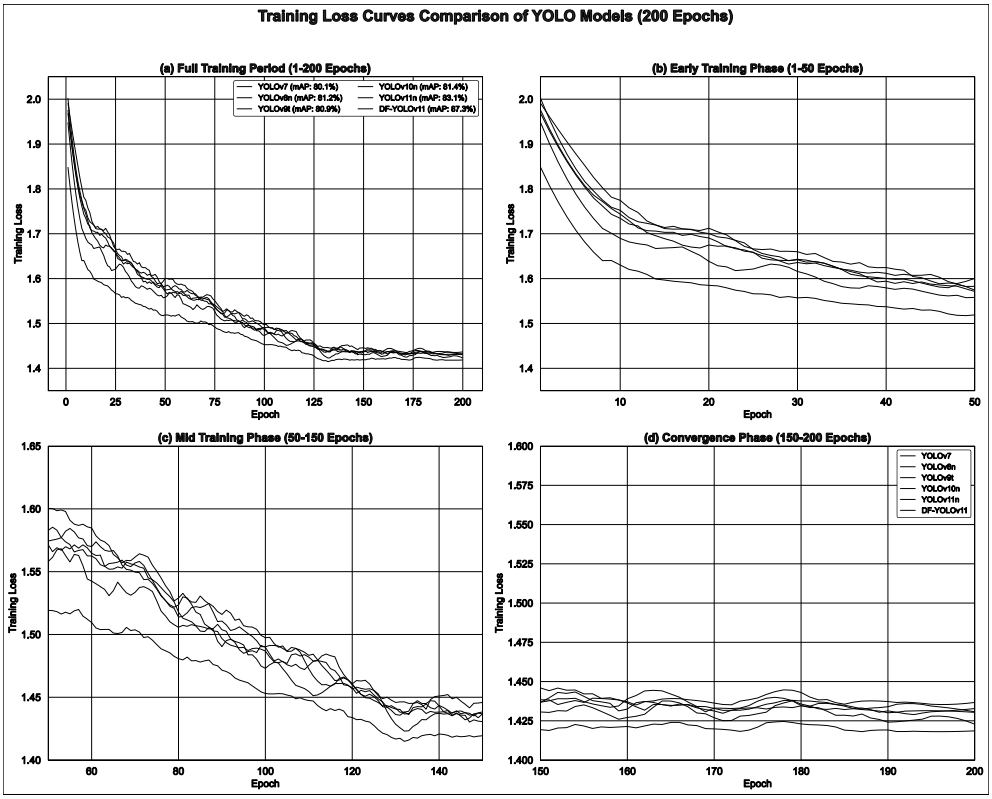

Fig 6: Comparative Training Loss Curves. (a) Full Training Epochs; (b) Early Training Stage (1-50 Epochs); (c) Middle Training Stage (50-150 Epochs); (d) Convergence Stage (150-200 Epochs)

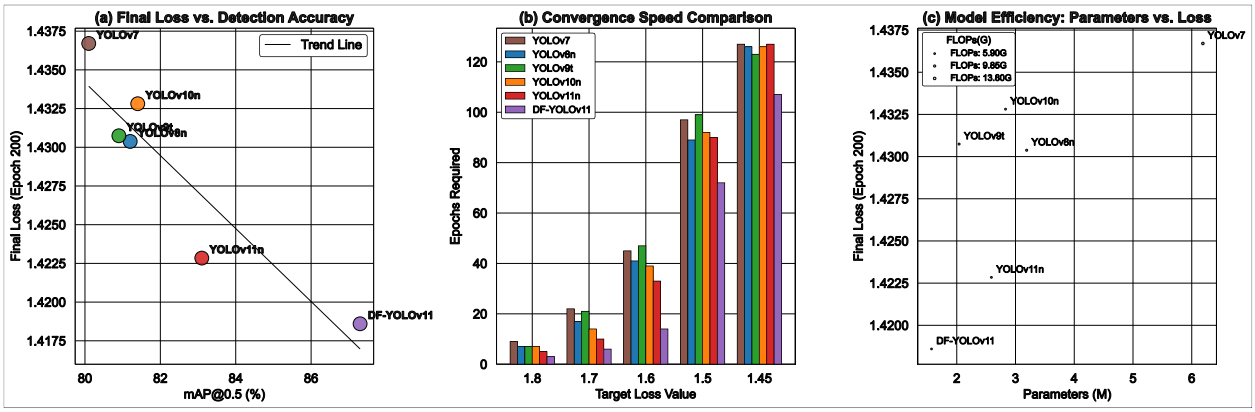

Fig 7: Comprehensive Performance Analysis. (a) Relationship Between Final Loss Value and mAP; (b) Convergence Speed Comparison; (c) Bubble Chart for Three-Dimensional Relationship of Model Efficiency

left corner of the graph with the smallest bubble, demonstrating that it achieves the most comprehensive efficiency optimization−with the fewest parameters (1.57M), the lowest final loss, and the smallest computational complexity (5.9 GFLOPs).

The comparative analysis of training loss curves further confirms the structural advantages of DF-YOLOv11. Its rapid initial convergence (Fig. 6(b)) can be attributed to the directional prior guidance of the Light-OrientedECA module, which enables the network to quickly focus on the axial features of wheat ears, avoiding many ineffective feature explorations. The stable convergence in the middle and late stages (Fig. 6(c)-(d)) reflects the stability of the DMS-Fusion module in multi-scale feature fusion and the effective optimization of hard samples by the Occlusion-CIoU loss function. In contrast, traditional models such as YOLOv7 exhibit higher loss values throughout the training process and greater fluctuations during convergence, indicating that they face greater optimization difficulties when dealing with the complex characteristics of the GWHD dataset.

The comprehensive performance relationship analysis (Fig. 7) reveals the intrinsic connection between accuracy, efficiency, and convergence speed. DF-YOLOv11's overall leadership in all three dimensions demonstrates that the improved modules proposed in this paper do not independently enhance a single indicator but achieve the optimization of overall performance through synergistic effects. Notably, DF-YOLOv11 improves detection accuracy while reducing model complexity, breaking the dilemma in traditional deep learning models where "performance improvement is inevitably accompanied by model expansion." This optimization benefits from targeted module designs: Light-OrientedECA achieves efficient feature selection through directional pooling and grouped convolution; DMS-Fusion optimizes the computational allocation of feature fusion via a scale-adaptive strategy; and Occlusion-CIoU improves the optimization efficiency of the loss function through dynamic weight adjustment.

3.4.2. Analysis of Convergence Characteristics and Performance Behavior of the Improved DF-YOLOv11

To more intuitively demonstrate the convergence characteristics and performance of each model during training, we plotted comparative loss curves of the ablation study (Fig. 8) and a comprehensive analysis diagram of convergence performance (Fig. 9). Fig. 8 shows the loss changes of YOLOv11n and its different improved versions of over 200 training epochs. Fig. 8(a) indicates that DF-YOLOv11 maintains the lowest loss value throughout the training process with the smoothest convergence curve. Fig. 8(b) further reveals that in the early training stage (1-50 epochs), DF-YOLOv11 exhibits the fastest loss reduction rate, demonstrating that its structural design can quickly capture effective features. Fig. 8(c) shows that in the convergence stage (150-200 epochs), the loss of DF-YOLOv11 stabilizes at a relatively low level with normal fluctuations, exhibiting excellent stability.

Fig. 9 quantifies the relationship between model convergence efficiency and performance balance from multiple perspectives. Fig. 9(a) illustrates the negative correlation trend between the final loss and detection accuracy (mAP@0.5). DF-YOLOv11 is located at the bottom right corner of the graph, i.e., it simultaneously achieves the highest mAP (87.3%) and the lowest final loss value, verifying the consistency of its optimization objectives. Fig. 9(b) quantifies the comprehensive convergence performance of each model through three core indicators: total loss reduction rate, convergence speed, and convergence efficiency, and DF-YOLOv11 achieves the highest total loss reduction rate and optimal convergence efficiency. Fig. 8(d) shows through comparison of convergence speed indicators that DF-YOLOv11 requires the fewest training epochs to reach the same threshold, which is significantly superior to other models.

4. Discussion

4.1. Effectiveness and Synergy Mechanisms of Core Modules

The DF-YOLOv11 model proposed in this study systematically addresses the three core challenges in the GWHD dataset through the introduction of three key improved modules. The success of the Light-OrientedECA module verifies the effectiveness of encoding domain priors (i.e., the tilt angle distribution of wheat ears) into a lightweight attention mechanism. Compared with Li J *et al* (2025) [21] method that introduces complex direction prediction branches, this module enhances tilt-related features through deterministic directional pooling and GSConv without significantly increasing computational burden (reducing parameter count by 0.32M). This provides a new idea for the design of lightweight attention mechanisms targeting objects with specific postures. The DMS-Fusion module dynamically optimizes the fusion process of the feature pyramid via scale-adaptive grouped convolution and

a multi-path structure. Its significant improvement in the detection performance of small and medium-sized targets (Recall increased by 1.0%) indicates that a targeted, asymmetric multi-scale fusion strategy is more adaptable to the extreme scale distribution of field targets than fixed-structure FPN or PANet. The introduction of the Occlusion-CIoU loss function directly targets the core challenge of dense occlusion. By dynamically adjusting weights to force the model to focus on hard samples, it becomes the key to improving the recall rate in overlapping regions (Recall increased to 80.7%). Ablation experiments demonstrate that these three modules not only independently contribute to performance gains but also generate synergistic effects, collectively elevating the model's comprehensive detection performance (mAP@0.5) to 87.3%.

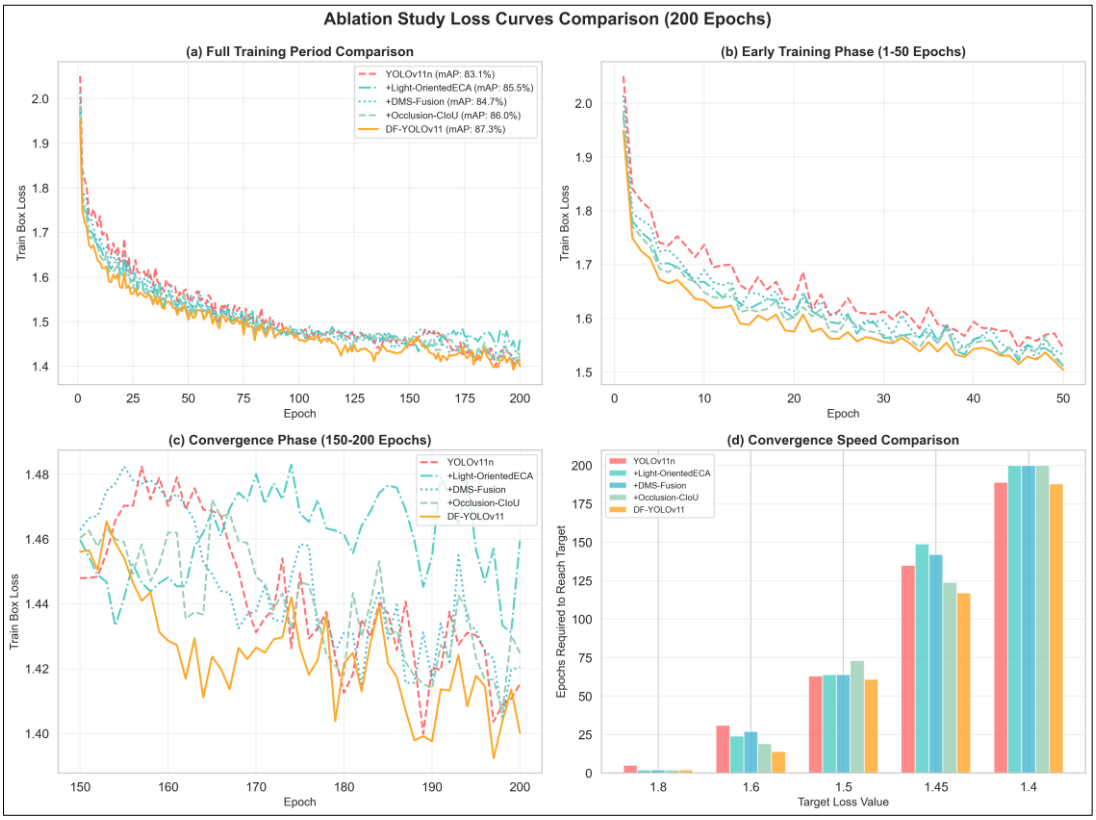

Fig 8: Ablation Study Loss Curves. (a) Full Training Epochs; (b) Early Training Stage (1-50 Epochs); (c) Convergence Stage (150-200 Epochs); (d) Convergence Speed Comparison

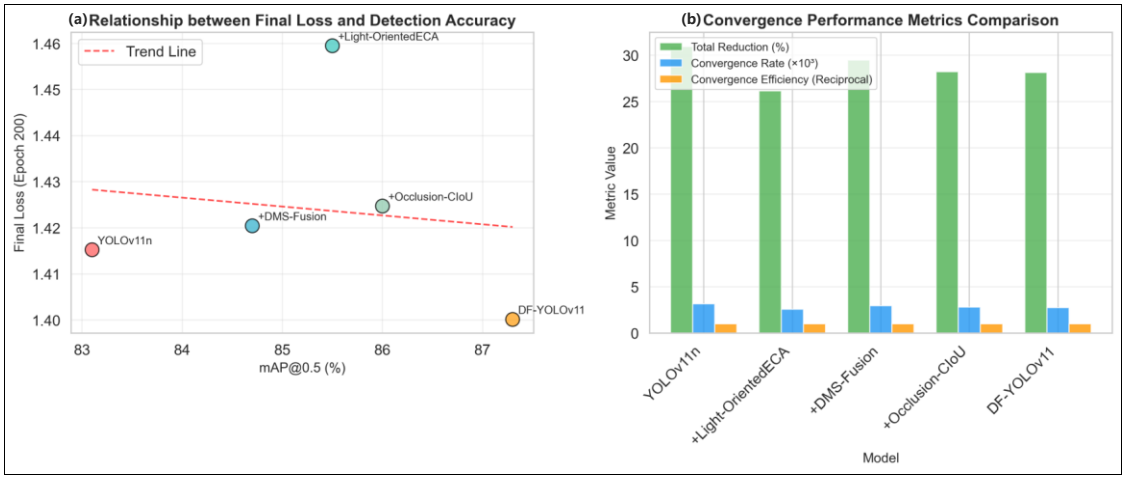

Fig 9: Comprehensive Analysis of Convergence Performance. (a) Relationship Between Final Loss Value and mAP; (b) Comparison of Convergence Performance Indicators

4.2. Comparative Advantages over Existing Lightweight Models

As shown in Table 1, compared with the current state-of-the-art lightweight models, DF-YOLOv11 exhibits outstanding performance in the trade-off between accuracy and efficiency. Compared with YOLOv9t, which also emphasizes lightweighting, DF-YOLOv11 outperforms it by 6.4 percentage points in mAP while having a lower parameter count (1.57M vs. 2.04M). This is mainly attributed to the more targeted structural design of this study rather than mere architectural pruning. Compared with the baseline YOLOv11n, DF-YOLOv11 achieves significant accuracy

improvement while realizing model size compression, which refutes the common dilemma that "performance improvement is inevitably accompanied by model expansion." When compared with improved models also targeting the GWHD dataset (e.g., YOLOv8n), DF-YOLOv11 maintains high accuracy while possessing more advantages in computational complexity (5.9 GFLOPs) and parameter count, making it more suitable for resource-constrained edge deployment scenarios. These comparative results demonstrate that deep model customization based on problem characteristics can more effectively achieve the task-oriented dual goals of accuracy and lightweighting than directly adopting general lightweight networks or performing simple module replacement.

4.3. Practical Significance of Lightweight Design

The final parameter count of DF-YOLOv11 is 1.57M with a computational complexity of 5.9 GFLOPs, a scale that endows it with practical feasibility for deployment on low-power mobile devices (e.g., unmanned aerial vehicles (UAVs) equipped with edge computing units or handheld smart terminals). Model lightweighting not only reduces storage and transmission overhead but more importantly lowers energy consumption and latency during inference, which is critical for agricultural applications requiring long-duration and large-scale field operations. The lightweighting approaches adopted in this study (GSConv, depthwise separable convolution, and data-driven anchor boxes) provide a solid software foundation for achieving "algorithm-hardware" co-optimization. In the future, further integration with techniques such as model quantization and pruning can be explored to adapt to more specific embedded platforms.

4.4. Analysis of Model Robustness and Generalization Ability

Multi-scenario performance analysis (Section 3.2) shows that DF-YOLOv11 exhibits strong robustness against common field challenges (e.g., moderate occlusion, conventional scale variations, and typical tilting). However, missed detections or duplicate detections still occur in scenarios involving extremely small targets, edge truncation, and highly dense mixing. This reflects the current boundaries of the model's capabilities: its optimization is based on the statistical priors of the training set, resulting in limited generalization ability for out-of-distribution (OOD) extreme cases. Nevertheless, compared with the baseline model, the error rate of DF-YOLOv11 in these challenging scenarios has been significantly reduced, indicating that its core improvement direction is correct. In the future, further enhancement of the model's boundary generalization ability is expected by introducing more diverse data augmentation (to simulate extreme cases) or leveraging semi-supervised learning to mine hard samples from unlabeled data.

4.5. Analysis of Training Dynamics and Convergence Characteristics

The comparison of loss curves (Fig. 6) reveals significant advantages of DF-YOLOv11 in terms of training dynamics. Compared with the baseline model YOLOv11n, the loss curve of DF-YOLOv11 exhibits a steeper initial descent slope (Fig. 6(b)) and more stable late-stage convergence (Fig. 6(d)). This phenomenon can be explained as follows: the Light-OrientedECA module provides a clear direction for feature learning through directional prior guidance, reducing random exploration in the early stage of training; the scale-adaptive strategy of the DMS-Fusion module optimizes gradient flow, avoiding gradient conflicts in multi-scale feature fusion; and the Occlusion-CIoU loss function ensures that hard samples receive sufficient attention during training through dynamic weight adjustment, preventing premature convergence to suboptimal solutions.

Further comparison of the loss curves from the ablation experiments (Fig. 8) more precisely reveals the independent contributions and synergistic effects of each core module on training dynamics. Fig. 8(a) shows that based on the YOLOv11n baseline, the introduction of any single improved module (Light-OrientedECA, DMS-Fusion, or Occlusion-CIoU) can reduce the loss value throughout the training process, while DF-YOLOv11, which integrates all three modules, consistently maintains the lowest loss level with the smoothest convergence curve. Fig. 8(b) focuses on the early training stage (1-50 epochs), where the loss reduction speed of DF-YOLOv11 is significantly faster than that of models with only a single module improved. Among them, the introduction of the Light-OrientedECA module significantly increases the loss descent slope of the baseline model, verifying the guiding role of directional priors in early feature learning; meanwhile, the DMS-Fusion module further accelerates the extraction and convergence of features for small and medium-sized targets by optimizing feature fusion efficiency. The comparison of the convergence stage (150-200 epochs) in Fig. 8(c) shows that DF-YOLOv11 has the smallest loss fluctuation range, while models with only a single module improved still exhibit slight loss oscillations. This indicates that the synergistic effect of multiple modules effectively enhances training stability and avoids the trap of local optimal solutions that may be caused by a single optimization direction.

Notably, there is high consistency between the final loss value of DF-YOLOv11 and its mAP (87.3%) (Fig. 7(a)), which contrasts with the traditional understanding that "low loss does not necessarily correspond to high accuracy." This regularity is further verified in the performance correlation analysis of the ablation experiments (Fig. 9(a)). Such consistency indicates that the loss function and model structure designed in this paper have excellent synergy: the Occlusion-CIoU loss function not only optimizes bounding box regression but also indirectly enhances the directionality of feature learning through an occlusion-aware mechanism; meanwhile, the improved feature extraction and fusion modules enable the network to more effectively utilize the supervisory signals provided by the loss function.

The comparison of convergence speeds (Fig. 7(b), Fig. 8(d)) shows that DF-YOLOv11 requires the fewest training epochs to reach the same loss threshold, which is of great value in resource-constrained practical application scenarios. Fast convergence not only reduces training time and computational costs but also lowers the risk of overfitting, as fewer training epochs mean fewer opportunities to learn noise and specific patterns in the training data. This feature makes DF-YOLOv11 more suitable for practical deployment environments that require frequent retraining or incremental learning.

Model efficiency analysis (Fig. 7(c)) further confirms DF-YOLOv11's breakthrough in the accuracy-efficiency trade-off. Traditional lightweight models often achieve efficiency improvements by significantly reducing model capacity, but at the cost of detection accuracy. In contrast, DF-YOLOv11

improves accuracy while maintaining or even reducing model complexity through structural innovation, which provides a new design paradigm for model deployment on edge computing devices. In particular, DF-YOLOv11 has the lowest computational complexity and parameter count among the comparative models, giving it a distinct advantage in field detection tasks with strict real-time requirements.

5. Conclusion

To address the core challenges of multi-scale variation, diverse postures, and dense occlusion in field wheat ear detection tasks, this paper proposes a lightweight improved model based on YOLOv11−DF-YOLOv11. Through innovations in three aspects (structure, mechanism, and optimization), this study systematically enhances the model's performance and efficiency: ① Designed a Light-OrientedECA, which enhances the feature discriminative ability for tilted wheat ears by incorporating directional priors and efficient computation; ② Proposed a DMS-Fusion, which optimizes the fusion efficiency of multi-scale features through a scale-adaptive strategy and improves the detection capability for targets of different sizes; ③ Improved the bounding box regression loss and proposed Occlusion-CIoU, which strengthens the model's localization robustness in dense regions through an occlusion-aware weight mechanism.

Comprehensive experiments on the public benchmark dataset GWHD show that DF-YOLOv11 achieves an mAP@0.5 of 87.3% and a Recall of 80.7%, while reducing the model parameter count and computational complexity to 1.57M and 5.9 GFLOPs, respectively, and outperforming current mainstream lightweight models. Detailed ablation experiments and multi-scenario analyze verify the effectiveness of each module and reveal the model's performance boundaries under extreme conditions. Training dynamics and performance analysis further confirm the superiority of DF-YOLOv11. Comparison of loss curves shows that DF-YOLOv11 exhibits faster convergence speed and more stable convergence characteristics, indicating that its improved modules can guide the training process more efficiently. Comprehensive performance relationship analysis reveals that DF-YOLOv11 achieves the optimal balance across three dimensions: accuracy, efficiency, and convergence speed, realizing the synergistic optimization of multiple objectives. This comprehensive performance advantage not only verifies the effectiveness of the module design but also provides a solid guarantee for its practical deployment in resource-constrained environments.

In conclusion, DF-YOLOv11 achieves an excellent balance between detection accuracy, model lightweighting, and inference efficiency, offering an effective technical solution for high-precision and real-time wheat ear detection and counting on resource-constrained mobile devices. This study not only contributes a high-performance model to the specific task of wheat ear detection but also its modular lightweight design idea targeting specific agricultural vision problems provides valuable insights for other tasks of automated acquisition of agricultural phenotypic information. In the future, we will promote the field deployment and application of this model and explore its transfer and extension to more crop detection tasks.

6. Acknowledgments

This work was supported by Qingdao Science and Technology Benefiting People Demonstration Project (No. 25-1-5-xdny11-nsh), and Natural Science Foundation of Shandong Province (No. ZR2020QF067, ZR2024LQX005).

7. Author Contributions

Xin Jiang: Methodology, Software, Data curation, Visualization, Investigation, Formal analysis, Writing - original draft. Delong Kong: Software, Data curation, Investigation, Validation, Formal analysis, Writing - review and editing. Moughal Tauqir: Funding acquisition, Project administration, Resources, Supervision, Writing - review and editing. Jiahua Zhang: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing - review and editing. All authors have read and agreed to the published version of the manuscript.

8. Declaration of Competing Interest

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

9. References

1. Cai Y, Guan K, Lobell D, *et al*. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agricultural and Forest Meteorology. 2019;274:144-159.
2. Peng Y, Zhao Y, Yu Z, Zeng J, Xu D, Dong J, *et al*. Wheat quality formation and its regulatory mechanism. Front Plant Sci. 2022;13:834654.
3. Liu T, Chen W, Wang Y, Wu W, Sun C, Ding J, *et al*. Rice and wheat grain counting method and software development based on Android system. Comput Electron Agric. 2017;141:302-309.
4. Ferrante A, Cartelle J, Savin R, Slafer GA. Yield determination, interplay between major components and yield stability in a traditional and a contemporary wheat across a wide range of environments. Field Crops Res. 2017;203:114-127.
5. Jin X, Liu S, Baret F, Hemerlé M, Comar A. Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. Remote Sens Environ. 2017;198:105-114.
6. Redmon J, Divvala S, Girshick R, *et al*. You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016:779-788. doi:10.1109/CVPR.2016.91
7. Qiu Z, Wang F, Li T, Liu C, Jin X, Qing S, *et al*. LGWheatNet: a lightweight wheat spike detection model based on multi-scale information fusion. Plants (Basel). 2025;14(6):1098.
8. Singh AK, Ganapathysubramanian B, Sarkar S, *et al*. Deep learning for plant stress phenotyping: trends and future perspectives. Trends Plant Sci. 2018;23(10):883-898.
9. Li Z, Hong W, Feng X, Wang A, Ma H, Qin J, *et al*. LKNet: enhancing rice canopy panicle counting accuracy with an optimized point-based framework. Plant Phenomics. 2025;7(1):100003.

10. Cui D, Liu P, Liu Y, Zhao Z, Feng J. Automated phenotypic analysis of mature soybean using multi-view stereo 3D reconstruction and point cloud segmentation. Agriculture. 2025;15(2):175.
11. Lin YX, Xiao X, Lin HF. YOLOv8-FDA: lightweight wheat ear detection and counting in drone images based on improved YOLOv8. Front Plant Sci. 2025;16:1682243.
12. Bai BY, Wang JS, Li JL, *et al*. T-YOLO: a lightweight and efficient detection model for nutrient buds in complex tea-plantation environments. J Sci Food Agric. 2024;104(10):5698-5711.
13. Yang CK, Sun XY, Wang J, *et al*. YOLOv8s-CGF: a lightweight model for wheat ear Fusarium head blight detection. PeerJ Comput Sci. 2024;10:e1948.
14. Qiu ZM, Wang F, Wang WL, *et al*. YOLO-SDL: a lightweight wheat grain detection technology based on an improved YOLOv8n model. Front Plant Sci. 2024;15:1495222.
15. Yang B, Gao Z, Gao Y, Zhu Y. Rapid detection and counting of wheat ears in the field using YOLOv4 with attention module. Agronomy. 2021;11(6):1202.
16. Wang Y, Qin Y, Cui J. Occlusion robust wheat ear counting algorithm based on deep learning. Front Plant Sci. 2021;12:645899.
17. David E, Serouart M, Smith D, *et al*. Global Wheat Head Detection 2021: an improved dataset for benchmarking wheat head detection methods. Plant Phenomics. 2021;2021:9846158.
18. Qing SH, Qiu ZM, Wang WL, *et al*. Improved YOLO-FastestV2 wheat spike detection model based on a multi-stage attention mechanism with a LightFPN detection head. Front Plant Sci. 2024;15:1411510.
19. Shen XJ, Zhang C, Liu K, *et al*. A lightweight network for improving wheat ears detection and counting based on YOLOv5s. Front Plant Sci. 2023;14:1289726.
20. Meng X, Li C, Li J, Li X, Guo F, Xiao Z. YOLOv7-MA: improved YOLOv7-based wheat head detection and counting. Remote Sens. 2023;15(15):3770.
21. Li JN, Wang ZS, Luo XB, *et al*. Research on detection and counting method of wheat ears in the field based on YOLOv11-EDS. Front Plant Sci. 2025;16:1672425.
22. Khaki S, Safaei N, Pham H, *et al*. WheatNet: a lightweight convolutional neural network for high-throughput image-based wheat head detection and counting. Neurocomputing. 2022;489:78-89.
23. Khanam R, Hussain M. YOLOv11: an overview of the key architectural enhancements. arXiv. [preprint] 2024:2410.17725. doi:10.48550/arXiv.2410.17725
24. Zheng Z, Wang P, Liu W, *et al*. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. IEEE Trans Cybern. 2022;52(8):8574-8586.
25. Li J, Dai F, Qian H, Huang L, Zhao J. Lightweight wheat spike detection method based on activation and loss function enhancements for YOLOv5s. Agronomy. 2024;14(9):2036.
26. Li S, Tao T, Zhang Y, Li M, Qu H. YOLO v7-CS: a YOLO v7-based model for lightweight bayberry target detection count. Agronomy. 2023;13(12):2952.
27. Li F, Lu Y, Ma Q, Zhao R. GhostConv+ CA-YOLOv8n: a lightweight network for rice pest detection based on the aggregation of low-level features in real-world complex backgrounds. Front Plant Sci. 2025;16:1620339.
28. Wang CY, Yeh IH, Liao HYM. YOLOv9: learning what you want to learn using programmable gradient information. In: Lecture Notes in Computer Science (ECCV 2024). Springer; 2025:15089. doi:10.1007/978-3-031-72751-1_1
29. Wang A, Chen H, Liu L, *et al*. YOLOv10: real-time end-to-end object detection. arXiv. [preprint] 2024:2405.14458.
30. He LH, Zhou YZ, Liu L, Cao W, Ma JH. Research on object detection and recognition in remote sensing images based on YOLOv11. Sci Rep. 2025;15:14032.

How to Cite This Article
Jiang X, Kong D, Tauqir M, Zhang J. DF-YOLOv11: lightweight wheat ear detection method based on improved YOLOv11. Int J Artif Intell Eng Transform. 2025;6(2):191–204. doi:10.54660/IJAIET.2025.6.2.191-204.