ECBAM-YOLOv8: A DEEP LEARNING MODEL GUIDED BY EFFICIENTTEACHER FOR PRECISE WHEAT GRAIN DETECTION

/

ECBAM -YOLOv8：基于高效教师引导的深度学习模型实现小麦籽粒精准检测

Xiao CUI, Huiqin LI, Jiangchen ZAN, Jianhua CUI, Pengzhi HOU, Qian ZHAO, Jisheng LIU, Xiaoying ZHANG*)
Faculty of Software Technologies, Shanxi Agricultural University, Taigu 030801, China;
Tel: +86-158-0344-9361; E-mail: xiaoyingzhang@sxau.edu.cn
DOI: https://doi.org/10.35633/inmateh-77-114

Keywords: *YOLOv8; grain recognition; small object detection; EfficientTeacher learning; wheat grain*

ABSTRACT

Real-time, high-precision detection of wheat grains is crucial for food security and intelligent management, yet fully supervised methods require extensive annotations and struggle with occlusion and overlap. This paper proposes a lightweight YOLOv8-CoT model based on EfficientTeacher. FasterNet is integrated with CoTAttention to optimize the FC-C2f unit, enhancing channel–spatial feature representation, while a CBAM module is inserted at the end of the neck to improve recognition of occluded and overlapping grains. A pseudo-label self-training strategy is adopted using 80% unlabeled data and 20% labeled samples. The proposed method achieves 91.7% accuracy in field scenarios, improves efficiency by 6.6%, and reduces annotation cost to one-fifth.

摘要

小麦籽粒实时高精度检测对粮食安全与智能管理至关重要，但全监督方法依赖大量标注且难应对遮挡重叠。本文提出基于 EfficientTeacher 的轻量化 YOLOv8-CoT：融合 FasterNet 与 CoTAttention 优化 FC-C2f，提升通道-空间特征表征；在颈部引入 CBAM 增强遮挡与重叠识别；采用 80% 未标注与 20% 标注数据进行伪标签自训练。在田间场景实现 91.7% 精度，效率提升 6.6%，标注成本降至 1/5。

INTRODUCTION

Wheat, as a widely cultivated staple crop worldwide, has long ranked among the top three food crops globally in terms of cultivation area and total production. Approximately one-third of the world's population relies on wheat as a primary food source, providing food security for about 30% of the global population (*Adam, 2023*). Increasing wheat yield per unit area remains a central objective of modern breeding programs *(Bastos et al., 2020)*. Automatic detection and counting of wheat grains can rapidly and accurately obtain grain number and spatial distribution at the spike, plant, and population scales, thereby enabling refined characterization of spike grain number, grain morphology, and fertility. This provides key parameters for estimating yield at the plant, unit-area, and regional scales, and significantly improves the efficiency of germplasm screening, yield component analysis, and field trial evaluation (*Wu et al., 2020*). With the continuous development of image processing, machine learning, and deep learning, such artificial intelligence-based methods have been increasingly applied in wheat yield prediction and multi-scale phenotypic analysis (*Zaji et al., 2022*).

In recent years, the YOLO series of models has been increasingly applied to agricultural object detection. *Li et al. (2024)*, introduced RGB-D depth information into an improved YOLOv7 framework to achieve three-dimensional detection and localization of fruits such as strawberries, significantly improving detection accuracy and pose estimation performance in harvesting scenarios and fully demonstrating the advantages of integrating depth information into object detection networks. However, this approach relies on relatively expensive depth sensors and fixed mechanical structures, making it difficult to scale to large-area wheat fields. Focusing on greenhouse tomatoes and other fruit targets, *Yang et al. (2023)* performed lightweight modifications to the feature extraction and attention modules of YOLOv8, substantially reducing the number of parameters and computational cost while maintaining high accuracy, thus validating the effectiveness of enhancing C2f structures through feature enhancement and attention mechanisms. Nevertheless, such studies mainly target fruits with "rounded, single, and well-defined" shapes, and their network designs and prior assumptions are not directly applicable to targets such as wheat spikes and grains, which are "slender, extremely small, and densely overlapped."

Aiming at UAV-based field wheat ear detection and counting, *Li et al. (2025)* and *Lin et al. (2025)* introduced P2 micro-scale detection layers, SPDConv, DySample, and efficient attention modules into the YOLOv8 framework under fully supervised settings, significantly improving the detection accuracy and model compactness for spike-level small targets and demonstrating the strong potential of YOLOv8 in this domain. However, these works primarily focus on spike-level targets in long-range aerial imagery, pay insufficient attention to the segmentation and separation of grain-level tiny targets in close-range static images, and still heavily depend on large-scale manual annotation.

In contrast, *Zhang et al. (2025)* incorporated the EfficientTeacher semi-supervised framework into an improved YOLOv8-based wheat ear detector. By introducing SPDConv and PSA modules to fully exploit unlabeled images, they further improved spike-level detection performance and initially verified the effectiveness of semi-supervised strategies in agricultural object detection. However, their attention and feature modules are still designed around spike-level targets, and the backbone and neck retain the original C2f structure, which limits their ability to represent grain-level tiny objects under severe occlusion and strong adhesion.

With the advancement of deep learning in agricultural research, some researchers have begun to integrate YOLOv5 with semi-supervised "EfficientTeacher" learning frameworks to train robust detectors under conditions of limited labeled data. *Xu et al. (2023)* proposed the EfficientTeacher semi-supervised object detection framework, which designs modules such as Dense Detector, Pseudo Label Assigner, and Epoch Adaptor to significantly improve the semi-supervised training efficiency and accuracy of YOLOv5-based single-stage detectors on general datasets such as VOC and COCO, thereby demonstrating the feasibility and generality of the EfficientTeacher concept in one-stage architectures. Building on this, *Zhou et al. (2023)* developed SSDA-YOLO by combining YOLOv5 with domain adaptation and a Mean-Teacher-based knowledge distillation framework. Through scene style transfer to generate cross-domain pseudo-images and the introduction of consistency loss, they improved cross-domain detection performance from source to target domains, indicating that semi-supervised EfficientTeacher mechanisms are also applicable to complex scenarios with significant domain shifts. Furthermore, *Lyu et al. (2022)* coupled a teacher–student model with a strip attention module and proposed a semi-supervised SPM-YOLOv5 for bagged citrus detection. By embedding the strip attention module into the YOLOv5 backbone to highlight strip-shaped citrus and branch features, and using a small number of labeled samples together with large amounts of unlabeled images to generate pseudo-labels, they significantly improved detection accuracy and recall while effectively reducing annotation costs in agricultural scenarios. Overall, these methods have achieved remarkable results in leveraging depth information, feature enhancement and attention mechanisms, multi-scale small object detection, and semi-supervised EfficientTeacher learning. However, most of them focus on fruit or spike-level targets, and still lack structural designs specifically tailored for wheat grains, which are smaller, denser, and more prone to adhesion.

In summary, existing studies either rely on costly depth sensors, or adopt feature module designs that better match fruit or spike-level targets, or remain at the level of generic semi-supervised detection based on the YOLOv5 architecture. There is still a clear gap in dense semi-supervised detection methods specifically targeting wheat grains as agricultural tiny objects. To address this limitation, this study proposes ECBAM-YOLOv8 based on the YOLOv8 and EfficientTeacher frameworks. The method, for the first time, embeds an FC-C2f module tailored to slender textures together with CBAM simultaneously into the backbone and neck, and uses wheat grain shape priors to jointly calibrate channel–spatial weights. Combined with a semi-supervised training strategy that fully exploits unlabeled samples, the proposed approach substantially reduces dependence on manual annotations while effectively improving the accuracy and robustness of wheat grain detection and counting in densely adhered scenes.

MATERIALS AND METHODS

Data collection

The experiment was conducted from September 2023 to June 2024 at the Yangjiazhuang Experimental Site (112.5°E, 37.4°N) of Shanxi Agricultural University. The area is located at an elevation of approximately 800 m and features a typical warm-temperate continental climate, with an average annual temperature of 11°C, a frost-free period of 176 days, and an annual precipitation of 498.85 mm, offering abundant sunlight resources. The soil is calcareous cinnamon soil developed from loess parent material, with moderate fertility: total nitrogen 1.09 g kg^{-1}, total phosphorus 1.32 g kg^{-1}, and total potassium 22.13 g kg^{-1}.The schematic diagram of the test site is shown in Fig.1.

A completely randomized block design was adopted for the experiment, using the winter wheat cultivar "Nongda 212" as the test material. This variety exhibits plump kernels and typical spike morphology, showing no significant morphological differences from common winter wheat, thereby meeting the requirements for kernel-spike feature extraction. Sowing was carried out on September 27, 2023, with a one-time application of base fertilizers: nitrogen (urea) 120 kg ha^{-2}, phosphorus 120 kg ha^{-2}, and potassium (K_2O) 120 kg ha^{-2}, at an organic-to-chemical fertilizer mass ratio of 1:3. No additional fertilizers were applied during the entire growth period. Winter irrigation was completed on November 24.

Fig. 1 - Schematic diagram of the test site

Data Processing

The fully matured grain samples were collected in July 2024. At this stage, the grains exhibited bright coloration and a hard texture, facilitating the extraction of texture and morphological features. To enhance data diversity, over 20 types of dark background fabrics with varying grayscale gradients were used for photography. An iPhone 15 Pro Max was employed as the imaging device, with the lens positioned 25–30 cm vertically above the samples and fixed at a 90° overhead angle to ensure clear delineation of grain edges.

To address the issue of reduced model accuracy caused by grain overlap, manual arrangement was applied during photography to ensure grains were densely packed and stacked, with overlapping samples accounting for >90% of the dataset. A total of 2,200 raw images were acquired. After rigorous screening and augmentation strategies—including random rotation, mirroring, and brightness adjustment—the dataset was expanded to 4,500 high-quality images for subsequent model training and validation.

Improved YOLOv8 object detection algorithm

Principles of the YOLOv8 algorithm

YOLO, which stands for "You Only Look Once," is an end-to-end single-stage object detection model *(Redmon et al., 2016)*. The model primarily consists of three core components: the Backbone, the Neck, and the Head. The Backbone extracts image features through a series of stacked convolutional neural network layers and passes these features to the Neck for further processing. The Neck composed of a Feature Pyramid Network (FPN) and a Path Aggregation Network (PAN), fuses and enhances multi-scale features to improve detection capability for objects of varying sizes *(Lin et al., 2017)*.

The Head then performs predictions on the three generated feature maps at different scales to produce the final detection results. As the latest iteration in the series, YOLOv8 is categorized into five variants—n, s, m, l, and x—based on model width and depth, each with distinct parameter configurations *(Jocher et al., 2023)*. Among them, the YOLOv8n model has the fewest parameters and the fastest detection speed, making it highly suitable for deployment on embedded devices.

Improved YOLOv8 object detection algorithm

To address the characteristics of wheat grain datasets - significant scale variations along with extensive occlusion and overlapping phenomena - this study optimizes the YOLOv8n algorithm. First, an improved FC-C2f module is introduced after convolutional layers to enhance small target feature extraction capability. Subsequently, a CBAM attention module is incorporated at the end of YOLOv8's neck network. This mechanism processes features to generate both channel-wise and spatial attention maps, which are then multiplied with input features to achieve adaptive feature refinement. The CBAM module significantly strengthens feature representation for occluded targets, improves precision in key feature extraction, while suppressing interference from irrelevant features, thereby substantially enhancing grain detection accuracy. Fig. 2 illustrates the architecture of the modified YOLOv8n algorithm after these improvements.

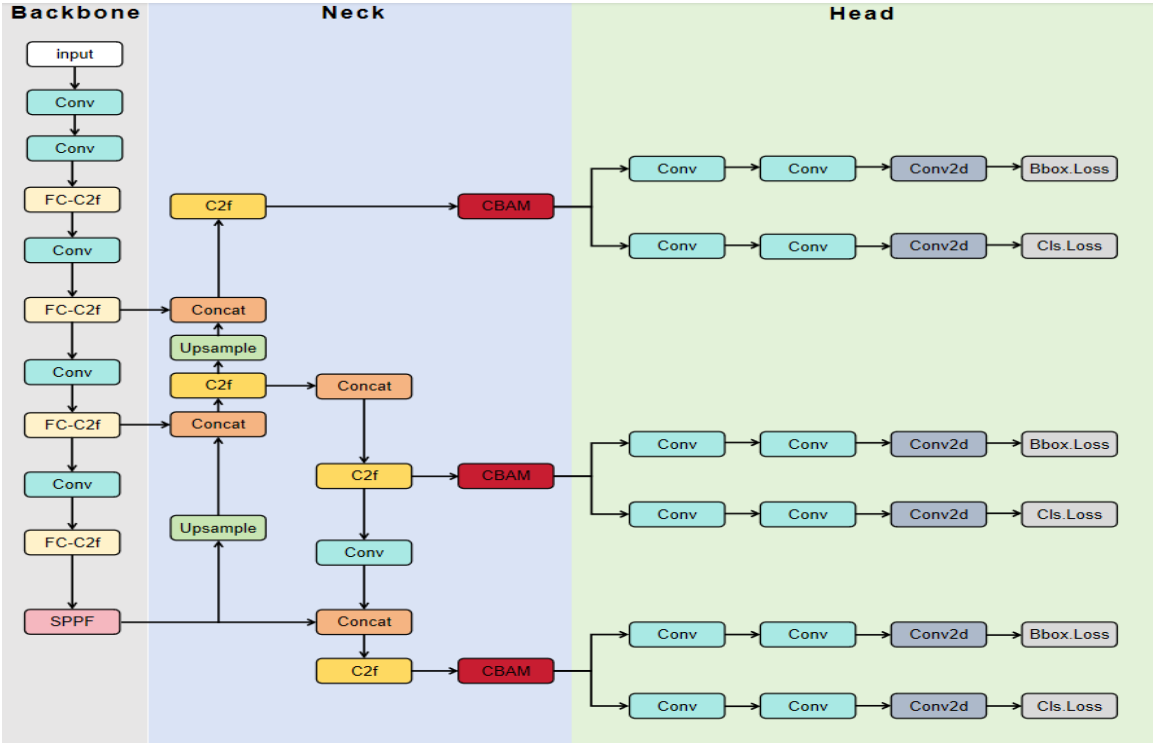


Fig. 2 - Network structure of improved YOLOv8 algorithm

FC-C2f module

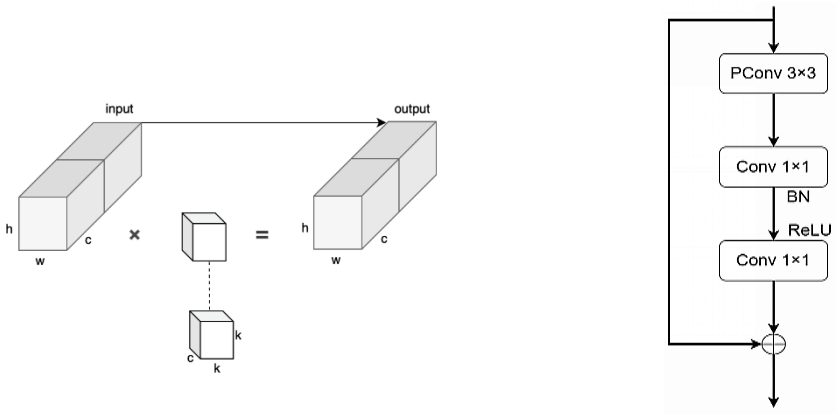
To simultaneously reduce computational load and parameter quantity while preserving channel information in object detection tasks, this study integrates the C2f module, FasterNet Block module, and CoTAttention mechanism to construct a lightweight FC-C2f module. Through the synergistic effects of multi-scale feature fusion and attention mechanisms, this module effectively enhances the model's detection capability for small targets while maintaining high computational efficiency.

C2f Module

In the YOLOv8 architecture, the C2f module improves object detection network performance through multi-level cross-layer connections and optimized design. It incorporates the Bottleneck concept, splitting feature maps for processing to reduce computational load and parameter quantity while enhancing the nonlinear representation capability of features. This enables efficient fusion of multi-scale features. Additionally, the C2f module introduces a split operation to improve the network's performance in capturing multi-scale and semantic information. This design not only strengthens feature extraction but also optimizes gradient flow, ensuring both model performance and efficiency improvements while reducing computational requirements.

FasterNet Block Module

FasterNet, as an innovative rapid network architecture, is built upon the Partial Convolution (PConv) module, achieving higher throughput with low latency. The PConv module significantly reduces redundant computations and memory access costs by performing convolution operations only on partial input channels, thereby improving hardware utilization efficiency for computational resources. The PConv structure and the FasterNet Block structure are shown in Fig. 3.



(a) PCovn （b）FasterNet Block

Fig. 3 - PConv Structure and FasterNet Block Structure

CoTAttention Mechanism

The CoTAttention mechanism successfully captures global dependencies among channels while avoiding dimensionality reduction operations. Through iterative optimization of the chain-of-thought process, it reduces redundant computations and enhances the model's sensitivity to fine-grained features (e.g., textures, edges). In high-resolution vision tasks (such as semantic segmentation and object detection), CoTAttention achieves a balance between efficiency and performance. The network structure of the CoTAttention module is illustrated in Fig. 4.

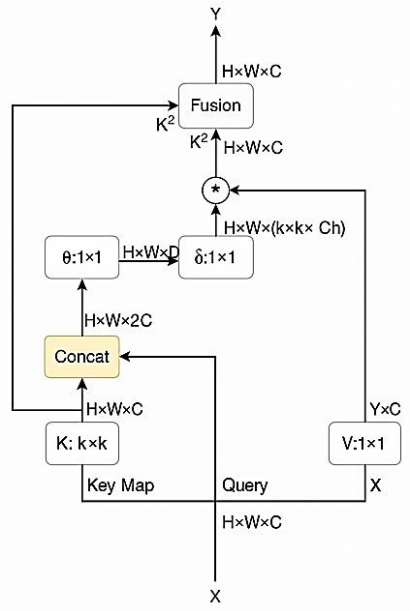


Fig. 4 - CoTAttention Attention Network Structure

FC-C2f Module

The FasterNet Block module consists of one 3×3 PConv and two 1×1 Convs. The CoTAttention mechanism was embedded after the first 3×3 convolution operation to form an FC-Bottleneck structure. By replacing the original Bottleneck structure in the C2f module with the proposed FC-Bottleneck structure, an innovative FC-C2f module was constructed. This module was then used to substitute the original C2f module in YOLOv8's backbone network while maintaining consistent input/output channel dimensions across all network layers. This improvement achieves both model lightweighting and enhanced image detection performance while accelerating processing efficiency for small target detection tasks. Fig. 5 details the network architecture of the FC-C2f module.

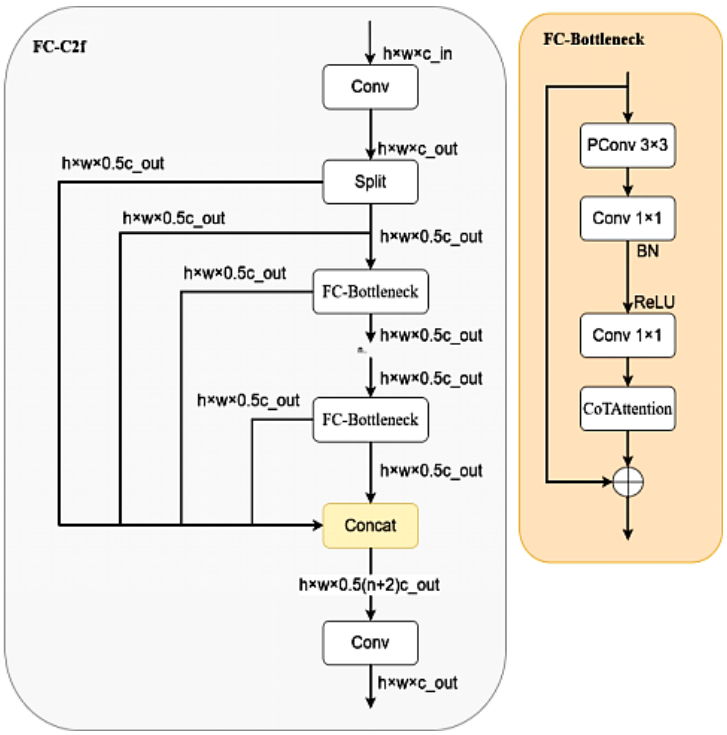


Fig. 5 - FC-C2f network structure

In this study, four SPDConv layers were added following the Conv layers of the YOLOv8 backbone network to enhance feature extraction. This approach takes the feature maps generated from the preceding convolutional operations as input and utilizes the SPD layers to transform spatial dimensions into depth dimensions, thereby increasing feature map depth without information loss. Subsequent consecutive Conv layers then perform convolutional processing, enabling feature extraction without reducing feature map dimensions while effectively preserving image details. Compared with traditional convolution operations, this insertion method demonstrates superior efficiency in retaining intra-channel information, thereby significantly enhancing feature extraction performance for small targets.

CBAM Attention Module

Woo et al. designed and developed an attention component for feed-forward convolutional neural networks, i.e., the Convolutional Block Attention Module (CBAM), which consists of a channel attention part and a spatial attention part, and its specific structure is shown in detail in Fig. 6.

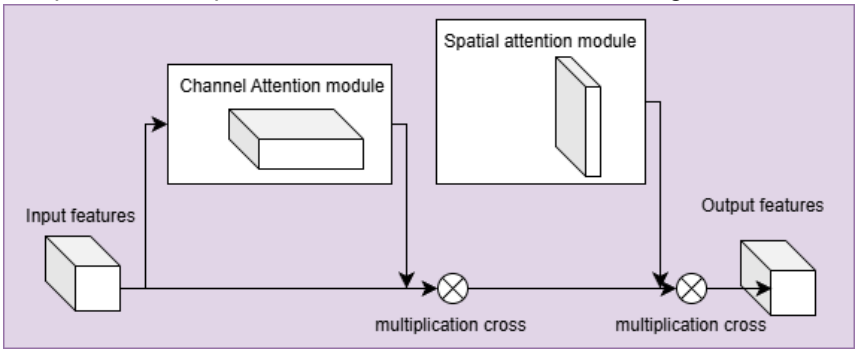


Fig. 6 - Structure of CBAM

This mechanism significantly enhances the feature representation of occluded targets, improves the accuracy of key feature extraction, while suppressing interference from irrelevant features, thereby substantially boosting grain detection accuracy. Experimental validation demonstrates that this structure performs particularly well in identifying overlapping grains in complex scenarios, with markedly improved recognition precision. Moreover, the additional computational overhead remains controllable, ensuring an optimal balance between inference speed and accuracy. This approach provides a more reliable technical solution for high-density crop grain detection.

Optimization of EfficientTeacher object detection algorithm

For wheat grains, dataset collection is relatively straightforward. However, comprehensive annotation incurs high labor costs due to the small size and dense distribution of targets in the scene. To address this issue, this paper proposes a EfficientTeacher object detection algorithm based on an improved EfficientTeacher. The architecture of the EfficientTeacher model is shown in Fig. 7.

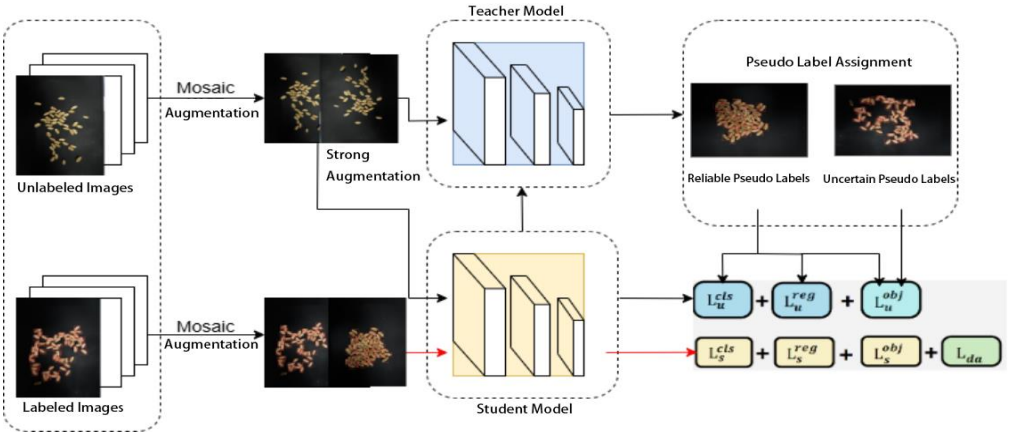


Fig. 7 - EfficientTeacher training framework

EfficientTeacher learning framework

First, labeled samples undergo Mosaic augmentation and are trained using supervised learning, while unlabeled samples undergo both Mosaic augmentation and Strong augmentation for training with pseudo-labels *(Xu et al., 2021)*. After that, the pseudo-label is assigned as reliable and uncertain by high and low screening thresholds τ_1 and τ_2, and the screening formula is shown in equation (1).

$$X_i = \begin{cases} \text{Re}\, liable, P_i > \tau_1 \\ Uncertain, \tau_2 < P_i < \tau_1 \end{cases} \quad (1)$$

where P_i represents the confidence score of the pseudo-labeling. Subsequently, the pseudo-labeled and labeled truth values were used as a guide for training and parameter updating of the student model, respectively, and finally the teacher model was updated by exponential moving average (EMA).

Although EfficientTeacher demonstrates significant improvements in detection accuracy compared to SSOD and one-stage anchor-based detectors, it still faces notable challenges in small object detection accuracy. These challenges primarily stem from the limited pixel occupancy of small objects in images, which complicates feature extraction and consequently impacts detection precision. Despite its outstanding performance in other aspects, EfficientTeacher requires further research and refinement to enhance its capabilities in small object detection scenarios.

Improved EfficientTeacher object detection algorithm

EfficientTeacher provides implementation code for both YOLOv5 and YOLOv8. However, only the EfficientTeacher version is available for YOLOv5. This paper extends the EfficientTeacher functionality to YOLOv8 through code architecture modifications. Since YOLOv5 employs an anchor-based design whereas YOLOv8 adopts an anchor-free approach, it was necessary to adapt YOLOv8 into an anchor-based variant to enable integration with the EfficientTeacher framework.

Given the current model's high compatibility with YOLOv5 but the absence of EfficientTeacher implementation code for YOLOv8, this study designed a comparative experiment to evaluate the performance of the EfficientTeacher model after incorporating YOLOv8n. The primary objective was to verify whether the adapted code could operate stably while enhancing model precision. The experimental results are detailed in Table 1.

Table 1

EfficientTeacher experimental results of different models

Model	mAP50	mAP50:95
YOLOv5s	0.833	0.473
YOLOv5l	0.825	0.466
YOLOv8n	0.881	0.493

The findings demonstrate that when using YOLOv8n as the base model, the algorithm achieved its peak mAP50 score of 88.1% - representing a 4.8% improvement over YOLOv5s and a 5.6% increase compared to YOLOv5l. These results confirm the significant effectiveness of our model replacement approach.

Experimental Environment and Training Design

The experiments were conducted on a Windows 11 operating system, with the development environment configured using Python 3.8, CUDA 11.1.0, and PyTorch 1.11.0. The training process was executed on an NVIDIA GeForce RTX 4060 GPU. For the training, the YOLOv8n model was used as the base model, with the input image size set to 640×640 pixels, a batch size of 32, and a weight decay coefficient of 0.0005. In the fully supervised learning mode, the initial learning rate was set to 0.01, the minimum learning rate to 0.002, and the learning rate adjustment strategy employed a cosine annealing schedule, with a total of 40 training epochs. In the EfficientTeacher learning mode, the learning rate was fixed at 0.01, and the training ran for a total of 300 epochs.

Experimental Dataset and Evaluation Metrics

To enhance the model's generalization capability, the dataset for this study was collected from mature wheat grain samples at the Yangjiazhuang Experimental Base of Shanxi Agricultural University in July 2024. During this period, the grains were firm, plump, and had distinct coloration, facilitating effective feature extraction during model training. To ensure image diversity, the photography was primarily conducted against dark backgrounds. After careful selection, over twenty different gradient shades of backgrounds were used. For the image capture, an iPhone 15 Pro Max was employed, resulting in a total of 2,200 raw images. After rigorous screening and augmentation using image enhancement techniques, the dataset was expanded to 4,500 images. All images were saved in JPG format. Using LabelImg software, the wheat spikes in the images were annotated with rectangular bounding boxes and assigned only the "wheat" class label. The annotations were then saved as "txt" files in YOLO format. Fig. 8 displays sample images from this dataset. The comparative analysis of six sample images clearly demonstrates the differential effects of grain arrangement, shooting height, and background color on imaging quality during winter wheat grain acquisition: (a) shows sparsely and evenly distributed grains, (b) exhibits blurred grain details due to excessive shooting height, (c) presents densely overlapping grain distribution, (d) displays aggregated clusters formed by adhering grains, (e) employs a dark background to enhance grain contour features, and (f) uses a light background for contrast. The study confirms that background color exerts decisive influence on the accuracy, robustness, and processing efficiency of image recognition. By systematically introducing complex scene elements such as high overlap and dark backgrounds, this experiment effectively enhances the generalization capability of the deep learning model *(Wang et al., 2025)*.

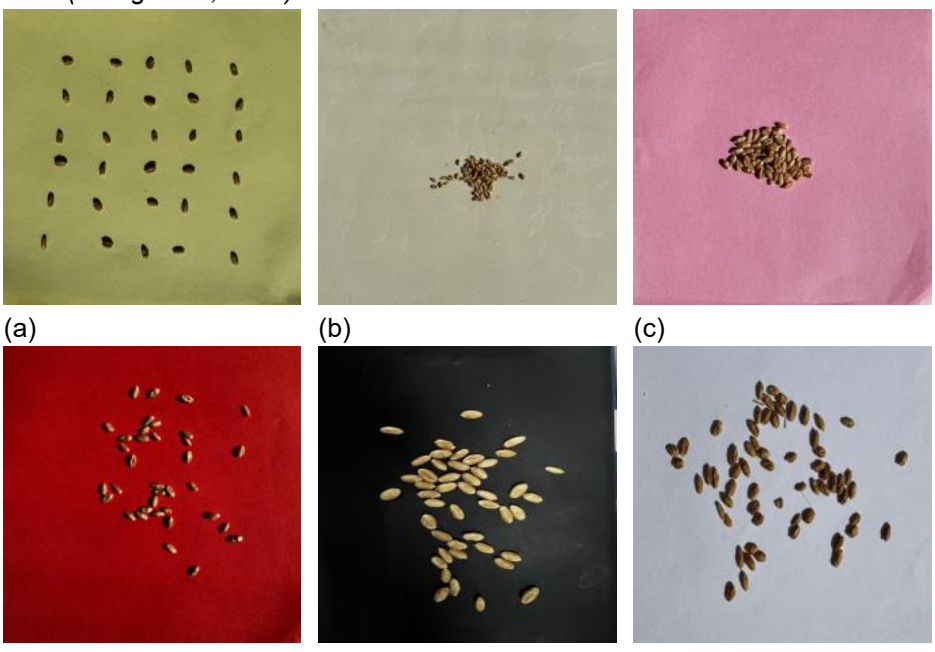


Fig. 8 - Example images of the datasets

The dataset used in this study comprises a total of 4,500 images, with 3,500 images designated as the training set and the remaining 1,000 images serving as the validation set. In the fully supervised learning experiments, 20% of the data—equivalent to 700 images—were randomly selected for training. For the EfficientTeacher learning experiments, these 700 images were used as labeled data, while the remaining 2,800 images were treated as unlabeled, with 20% of them annotated for experimental purposes.

In this study, a series of evaluation metrics recognized in the field of target detection, including Precision (P), Recall (R), and mean Average Precision (mAP) at an IoU threshold of 0.5 were used to determine the mAP50 by integrating the Precision-Recall (P-R) curve. The calculation in equation (2) - (3).

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \times 100\% \quad (2)$$

$$AP = \int_0^1 P(R)dR \quad (3)$$

The mAP50:95 was calculated in steps of 0.05 as the IoU threshold was varied from 0.5 to 0.95. The precision is the proportion of all detections judged to be positive samples in a given category that are actually positive, which is more stringent than the mAP50. It is calculated by averaging mAPs over IoU thresholds ranging from 0.50 to 0.95 at intervals of 0.05, as in equation (4) - (5).

$$P = \frac{TP}{TP + FP} \times 100\% \quad (4)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (5)$$

EfficientTeacher Learning Experiments

This study conducted comparative experiments on the effectiveness of fully supervised and EfficientTeacher training methods under different labeled sample ratios. The experimental results are presented in Fig. 9. In the figure, the blue area on the left represents the mAP50 values obtained using fully supervised training with all labeled samples, while the orange area on the right displays the mAP50 values achieved by further applying EfficientTeacher training methods on top of the fully supervised training. Given that the total number of training samples was 3,500, a 20% labeling ratio corresponds to 700 labeled samples and 2,800 unlabeled samples, with other ratios adjusted accordingly.

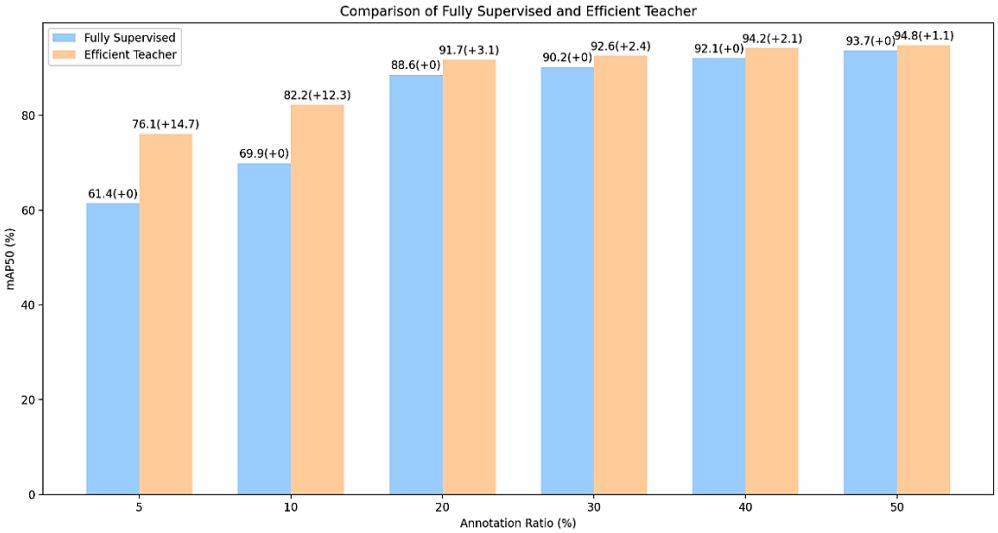


Fig. 9 - mAP50 metric of different annotation sample ratios

As can be observed from Fig. 9, when the labeled sample ratio is 5%, the fully supervised training yields an mAP50 value of 61.4%. Through EfficientTeacher training, this value increases by 14.7%, reaching 76.1%, surpassing the fully supervised training result of 69.9% achieved with a 10% labeled sample ratio. At other labeling ratios, EfficientTeacher training improves accuracy by 1.5%, 0.5%, and 0.5%, respectively, compared to fully supervised training. It can be seen that EfficientTeacher learning can reach a higher mAP50% when the annotated image ratio is low, especially at the annotation ratio of 20%, the performance is close to that of fully supervised learning, but the amount of annotated data used is greatly reduced, so it is the most cost-effective.

With the increase of the proportion of annotated images, the performance improvement of EfficientTeacher learning gradually flattened. Moreover, when the labeled sample ratio is low and the number of unlabeled samples is high, the performance improvement from EfficientTeacher training is particularly significant. This has a positive impact on wheat grain detection. The manually annotated images were randomly divided into 10 groups, with each group containing 4 images for model evaluation.

As shown in Table 2, the enhanced model demonstrated stable accuracy of approximately 90% in real-world scenarios, achieving a peak detection precision of 91.72%.

Table 2

Accuracy Analysis of ECBAM-YOLOv8 Model Detection versus Manual Counting

NO.	Manual count	Model count	Number of missed detection	Omission factor	Precision ratio
1	359	321	38	10.63%	89.37%
2	421	379	42	10.06%	89.94%
3	433	391	42	9.68%	90.32%
4	217	196	21	9.59%	90.41%
5	189	172	17	9.15%	90.85%
6	285	259	26	8.98%	91.02%
7	295	269	26	8.83%	91.17%
8	311	285	26	8.28%	91.72%
9	463	421	42	9.13%	90.87%
10	354	324	30	8.43%	91.57%

RESULTS

Ablation Study and Algorithm Comparison Experiments

Ablation Study

In order to verify the effectiveness of the improvement strategy proposed in this paper in enhancing the model accuracy, this study conducts ablation experiments with the YOLOv8n algorithm as a benchmark. The experimental results are detailed in Table 3. where A represents the FC-C2f module replacing C2f, B represents the introduced CBAM attention module, and C represents the EfficientTeacher target detection method mentioned in this paper.

The objective of this research was to develop an efficient wheat head detection algorithm by modifying the original YOLOv8 architecture. The improvements include: adjusting YOLOv8's network structure through the FC-C2f module, incorporating the CBAM attention mechanism, and embedding these enhancements into a EfficientTeacher learning framework. To elucidate the impact of each modification, ablation experiments were performed under identical training environments and hyperparameter settings. In this section, the YOLOv8n model serves as the baseline, with the aforementioned improvements implemented incrementally for comparative analysis.

According to the data presented in Table 3, the detection performance shows significant improvement with the progressive integration of enhancement modules into YOLOv8n. However, the proposed algorithm does not achieve optimal performance in terms of recall rate. This phenomenon occurs because while improving the feature extraction capability for small targets, the algorithm also optimizes computational complexity and memory usage. Nevertheless, given that both the precision and mean average precision reach their peak values, and considering the inherent trade-off between precision (P) and recall (R), the slight decrease in recall rate remains acceptable.

Table 3

Ablation Study Results of ECBAM-YOLOv8

Model	A	B	C	P	R	mAP50	mAP50:95
1				0.864	0.785	0.857	0.426
2	√			0.901	0.845	0.909	0.491
3		√		0.890	0.825	0.893	0.461
4			√	0.825	0.716	0.797	0.309
5	√	√		0.895	0.835	0.901	0.475
6	√		√	0.883	0.824	0.888	0.454
7		√	√	0.890	0.832	0.895	0.465
8	√	√	√	0.909	0.855	0.917	0.502

Algorithm Comparison Experiment

To further validate the performance advantages of the proposed algorithm on the wheat head detection dataset, comparative experiments were conducted against several commonly used object detection algorithms as well as the EfficientTeacher baseline. For a fair comparison, the default input resolutions of all methods were kept unchanged. The experimental results are detailed in Table 4.

In this experiment, EfficientTeacher employed the YOLOv5s detector, while V8 represents the EfficientTeacher training results using YOLOv8n as the baseline model. ECBAM-YOLOv8 denotes our proposed EfficientTeacher training results based on the enhanced YOLOv8n baseline model. All fully supervised object detection algorithms listed in the table were trained for 300 epochs without using pretrained weights.

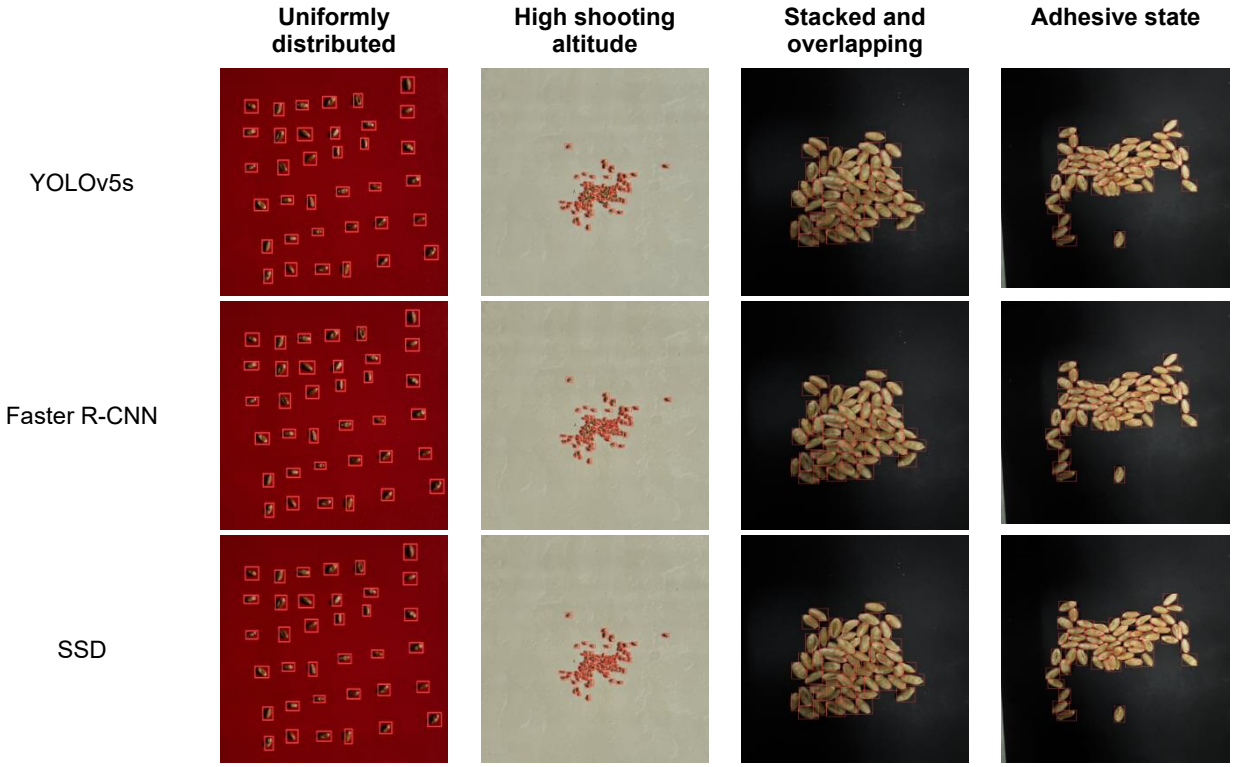
As shown in Table 4, the improved YOLOv8n model achieved a mean average precision (mAP50) of 91.7% under EfficientTeacher training conditions, representing a 1.5% performance improvement over the baseline EfficientTeacher model (EfficientTeacher) and a 1.0% improvement over the latest YOLOv11n (mAP50=0.907). Notably, YOLOv11n, as the most advanced lightweight detector in the YOLO series, has optimized network structure and loss function for small-object detection, yet our ECBAM-YOLOv8 still outperforms it in both mAP50 and mAP50:95. Compared to all other fully supervised models in the table (including YOLOv5s, Faster R-CNN, YOLOv10, and YOLOv11n), our approach demonstrated superior overall accuracy, leading to the conclusion that the enhanced model developed in this study is particularly well-suited for wheat grain detection applications.

Table 4

Comparison of Winter Wheat Grain Recognition Models

Model	Size	P	R	mAP50	mAP50:95
YOLOv5s	1024	0.831	0.747	0.816	0.309
Faster R-CNN	1024	0.855	0.781	0.851	0.405
SSD	1024	0.883	0.814	0.884	0.451
YOLOv8n	1024	0.868	0.804	0.874	0.445
YOLOX	1024	0.891	0.825	0.893	0.461
YOLOv10	1024	0.901	0.845	0.909	0.491
EfficientTeacher	1024	0.896	0.828	0.897	0.465
V8	1024	0.895	0.835	0.902	0.477
YOLOv11n	1024	0.903	0.849	0.907	0.489
ECBAM-YOLOv8	1024	0.909	0.855	0.917	0.502

As evident from Fig. 10, all models demonstrate high accuracy in identifying uniformly distributed grains, while exhibiting distinct advantages and disadvantages when processing images of grains under other conditions.



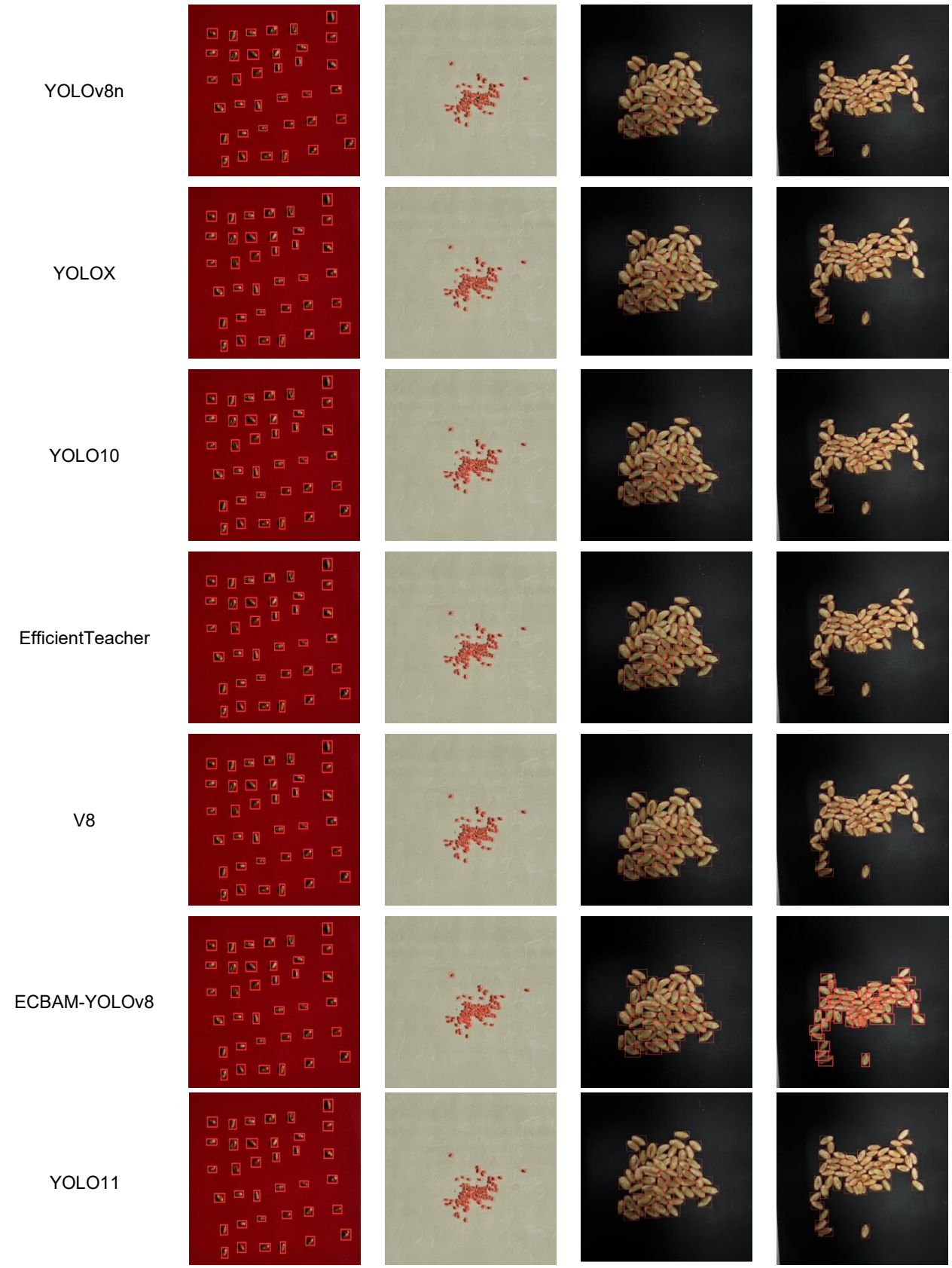


Fig.10 - Comparison of grain recognition results among different models

Specifically: for grains captured at high shooting altitudes, YOLOv5s, Faster R-CNN, and YOLOv8n show significantly higher false detection and missed detection rates compared to other models, which primarily accounts for their final detection accuracy remaining below 88%; YOLOv11n performs better than these models in this scenario but still lags behind ECBAM-YOLOv8 due to insufficient adaptation to the slender and

dense characteristics of wheat grains. In detecting stacked and adhesive grain states, ECBAM-YOLOv8 achieves notably higher accuracy than all comparative models, including YOLOv11n. Particularly for adhesive grains, ECBAM-YOLOv8 maintains substantially lower false detection and missed detection rates, benefiting from the tailored FC-C2f module and dual-embedded CBAM attention mechanism that enhance feature representation for complex grain distributions.

In conclusion, the proposed ECBAM-YOLOv8 model not only outperforms traditional baseline models but also surpasses the latest YOLOv11n detector in key metrics and complex scenario adaptability. It demonstrates robust detection accuracy in practical tests and shows superior application value for grain recognition in complex agricultural environments.

CONCLUSIONS

This study addresses the demand for high-throughput phenotyping of winter wheat, achieving comprehensive optimization of the entire process from raw wheat grain image acquisition to final field grain detection and counting. It focuses on two core issues: the missed detection of small objects in densely occluded scenes and the high annotation costs associated with fully supervised methods. The experimental system comprehensively covers 4,500 grain images from the Yangzhuang Experimental Station of Shanxi Agricultural University, integrating the YOLOv8n baseline model, FC-C2f lightweight units, CBAM attention modules, and the EfficientTeacher framework. To ensure the comprehensiveness of the evaluation, the latest YOLOv11n is also included as a comparative baseline.

Under the EfficientTeacher framework with a 20% annotation ratio, the ECBAM-YOLOv8 grain detection model achieves a precision of 91.7% and a recall of 85.5%, outperforming traditional baseline models such as YOLOv5s and YOLOX as well as the latest YOLOv11n. It effectively mitigates the high false positive and false negative rates caused by grain overlap and occlusion. Through EfficientTeacher training, this study maintains the advantage of detection accuracy while reducing manual annotation costs to one-fifth of that of traditional fully supervised schemes. This lays a technical foundation for the low-cost and large-scale deployment of winter wheat high-throughput phenotyping platforms, and provides a new "lightweight-EfficientTeacher" paradigm for real-time wheat grain detection.

ACKNOWLEDGEMENT

This research is supported by the Project Plan of Shanxi Agricultural University (CXGC2025057).

REFERENCES

[1] Adam S. (2023). Wheat crops: Sustaining global food security [J]. *Advances in Crop Science and Technology*, 11(7): 600.

[2] Bastos, L. M., Carciochi, W., Lollato, R. P., Jaenisch, B. R., Rezende, C. R., Schwalbert, R., Vara Prasad, P. V., Zhang, G., Fritz, A. K., Foster, C., Wright, Y., Young, S., Bradley, P., & Ciampitti, I. A. (2020). Winter wheat yield response to plant density as a function of yield environment and tillering potential: A review and field studies[J]. *Frontiers in Plant Science*, Vol. 11, pp. 54. DOI: 10.3389/fpls.2020.00054, Netherlands.

[3] Jocher, G., Chaurasia, A., Qiu, J. (2023). *Ultralytics YOLOv8 [EB/OL].* Available: https://github.com/ultralytics/ultralytics, United States.

[4] Lan, M., Liu, C., Zheng, H., Wang, Y., Cai, W., Peng, Y., Xu, C., & Tan, S. (2024). RICE-YOLO: In-Field Rice Spike Detection Based on Improved YOLOv5 and Drone Images[J]. *Agronomy*, Vol. 14, No. 4, p.836. DOI: 10.3390/agronomy14040836, Switzerland.

[5] Li, R., Sun, X., Yang, K., He, Z., Wang, X., Wang, C., Wang, B., Wang, F., & Liu, H. (2025). A lightweight wheat ear counting model in UAV images based on improved YOLOv8 (PSDS-YOLOv8)[J]. *Frontiers in Plant Science*, Vol. 16, pp. 1536017. DOI: 10.3389/fpls.2025.1536017, Netherlands.

[6] Li, Y., Wang, W., Guo, X., Wang, X., Liu, Y., & Wang, D. (2024). Recognition and positioning of strawberries based on improved YOLOv7 and RGB-D sensing[J]. *Agriculture*, Vol. 14, No. 4, pp. 624. DOI: 10.3390/agriculture14040624, Switzerland.

[7] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature Pyramid Networks for Object Detection[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017: pp. 936–944. DOI: 10.1109/CVPR.2017.106, United States.

[8] Lin, Y., Xiao, X., & Lin, H. (2025). YOLOv8-FDA: Lightweight wheat ear detection and counting in drone images based on improved YOLOv8[J]. *Frontiers in Plant Science*, Vol. 16, pp. 1682243. DOI: 10.3389/fpls.2025.1682243, Netherlands.

[9] Liu, Z., Zhang, R., Zhong, H., & Sun, Y. (2024). YOLOv8 for Fire and Smoke Recognition Algorithm Integrated with the Convolutional Block Attention Module[J]. *Open Journal of Applied Sciences*, Vol. 14, pp. 159–170. DOI: 10.4236/ojapps.2024.141012, Switzerland.

[10] Lyu J, Li S J, Zeng M Y, Dong B S. A semi-supervised SPM-YOLOv5-based detection method for bagged citrus[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, 38(18): 204–211.

[11] Rasheed A F, Zarkoosh M. (2025). Optimized YOLOv8 for multi-scale object detection[J]. *Journal of Real-Time Image Processing*, 22: 6. DOI: 10.1007/s11554-024-01582-x.

[12] Redmon J, Divvala S, Girshick R, Farhadi A. (2016). You Only Look Once: Unified, Real-Time Object Detection[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016: 779–788. DOI: 10.1109/CVPR.2016.91.

[13] Wang X, Li C, Zhao C, et al. (2025). GrainNet: efficient detection and counting of wheat grains based on an improved YOLOv7 modeling[J]. *Plant Methods*, 21: 44. DOI: 10.1186/s13007-025-01363-y.

[14] Wu W, Yang T, Li R, et al. (2020). Detection and enumeration of wheat grains based on a deep learning method under various scenarios and scales[J]. *Journal of Integrative Agriculture*, 19(8): 1998–2008. DOI: 10.1016/S2095-3119(19)62803-0.

[15] Xu B, Chen M, Guan W, Hu L. (2023). Efficient Teacher: Semi-Supervised Object Detection for YOLOv5[EB/OL]. arXiv:2302.07577. DOI: 10.48550/arXiv.2302.07577.

[16] Xu M, Zhang Z, Hu H, et al. (2021). End-to-End Semi-Supervised Object Detection with Soft Teacher[C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021: 3060–3069.

[17] Yang G, Wang J, Nie Z, Yang H, Yu S. (2023). A lightweight YOLOv8 tomato detection algorithm combining feature enhancement and attention [J]. *Agronomy*, 13(7): 1824. DOI: 10.3390/agronomy13071824.

[18] Zhang R, Yao M, Qiu Z, Zhang L, Li W, Shen Y. (2024). Wheat Teacher: A One-Stage Anchor-Based Semi-Supervised Wheat Head Detector Utilizing Pseudo-Labeling and Consistency Regularization Methods[J]. *Agriculture*, 14(2): 327. DOI: 10.3390/agriculture14020327.

[19] Zhang Y, Xu Z, Zhang X, Li F, Cui X. (2025). Semi-supervised wheat ear detection algorithm based on improved YOLOv8[J]. *INMATEH – Agricultural Engineering*, 75(1): 630–639. DOI: 10.35633/inmateh-75-54.

[20] Zaji A, Liu Z, Xiao G, Sangha J S, Ruan Y. (2022). A survey on deep learning applications in wheat phenotyping [J]. *Applied Soft Computing*, 131: 109761. DOI: 10.1016/j.asoc.2022.109761.

[21] Zhou H, Jiang F, Lu H. (2023). SSDA-YOLO: Semi-supervised domain adaptive YOLO for cross-domain object detection [J]. *Computer Vision and Image Understanding*, 229: 103649. DOI: 10.1016/j.cviu.2023.103649.