BGE-YOLOv11: A LIGHTWEIGHT MODEL FOR DETECTING MAIN STEM NODES IN SOYBEAN
/
基于大豆主茎节点检测的 BGE-YOLOv11 轻量化模型研究

Hao BI, Xiuying XU*)[1], Yiting LIU, Yanxu JIAO, Lingfeng ZHU
College of Engineering, Heilongjiang Bayi Agricultural University, Daqing/ China;
Tel: +86 13945925608; E-mail: xuxiuying@byau.edu.cn
Corresponding author: Xiuying Xu
DOI: https://doi.org/10.35633/inmateh-79-71

Keywords: *soybean main stem node detection; BGE-YOLOv11; lightweight model; deep learning*

ABSTRACT
The number of main stem nodes in soybean is an important phenotypic trait for evaluating plant growth, varietal characteristics, and breeding potential. However, automatically detecting soybean main stem nodes under field conditions remains challenging because of plant occlusion, variations in illumination, and complex backgrounds, while manual counting is inefficient for large-scale phenotypic data acquisition. Therefore, this study proposes BGE-YOLOv11, a lightweight model for detecting soybean main stem nodes in complex field environments. Using an image dataset of soybean main stem nodes collected under field conditions, the model was developed by integrating a bidirectional feature pyramid network (BiFPN), Global–Local Self-Attention (GLSA), and EfficientHead into the YOLOv11n baseline to enhance multiscale feature representation and detection performance. The proposed model was evaluated in terms of precision, recall, mean average precision at an intersection over union threshold of 0.5 (mAP50), and model complexity. BGE-YOLOv11 achieved a precision of 94.9%, recall of 92.9%, and mAP50 of 97.6%, representing improvements of 3.9, 4.0, and 3.0 percentage points, respectively, over the YOLOv11n baseline. Moreover, the model contained only 1.87 million parameters, required 5.6 GFLOPs, had a memory footprint of 8.0 MB, and achieved an inference time of 3.0 ms. These results demonstrate that BGE-YOLOv11 improves detection performance while retaining a lightweight architecture. Nevertheless, its performance may be affected by severe occlusion and highly variable field conditions. Future research will focus on expanding the dataset and further improving the model's generalization ability. This study provides an effective method for the automated counting of soybean main stem nodes and the analysis of phenotypic traits.

摘要
大豆主茎节点数量是反映植株生长状态、品种特征和育种潜力的重要表型性状。然而，在实际田间环境中，由于植株遮挡、光照变化以及复杂背景干扰等因素，大豆主茎节点的自动检测仍面临挑战，而人工计数方法难以满足大规模表型数据获取需求。针对复杂田间环境下大豆主茎节点快速检测问题，本文提出了一种轻量化 BGE-YOLOv11 检测模型。基于田间采集的大豆主茎节点图像数据集，在 YOLOv11n 基础模型上引入 BiFPN、全局局部自注意力（GLSA）模块和 EfficientHead 检测头，以增强多尺度特征表达能力并提升节点检测性能。采用 Precision、Recall、mAP50 以及模型复杂度指标对所提出模型进行综合评价。结果表明，BGE-YOLOv11 模型的 Precision、Recall 和 mAP50 分别达到 94.9%、92.9%和 97.6%，相比 YOLOv11n 分别提升 3.9、4.0 和 3.0 个百分点。同时，该模型参数量降低至 1.87 M，浮点计算量仅为 5.6 G，模型大小为 8.0 MB，推理时间为 3.0 ms。结果表明，BGE-YOLOv11 在提高检测性能的同时保持了良好的轻量化特性。然而，在极端遮挡和复杂多变田间环境下，模型性能仍可能受到一定影响。未来研究将进一步扩充数据集规模，提高模型泛化能力。本研究为大豆主茎节点自动计数和表型性状分析提供了一种有效方法。

INTRODUCTION
At present, China is actively developing soybean varieties with high yield, superior quality, and strong adaptability to mechanized production (*Qu et al., 2022*). Heilongjiang Province, one of China's major soybean-producing regions, primarily cultivates determinate soybean varieties.

---

[1] *Hao Bi, currently pursuing a master's degree; Yiting Liu, currently pursuing a master's degree; Yanxu Jiao, currently pursuing a master's degree; Lingfeng Zhu, currently pursuing a master's degree; Xiuying Xu, Associate Professor, Doctor of Engineering.*

During soybean breeding and variety evaluation, accurate acquisition of plant phenotypic traits is essential for characterizing varieties and selecting superior breeding materials. The number of main stem nodes is an important agronomic trait reflecting soybean plant architecture. It affects plant height, stem structure, lodging resistance, and pod spatial distribution, and is closely associated with yield formation. Therefore, it serves as an important indicator for evaluating soybean growth status and yield potential (*Qiu et al., 2006*). However, under field conditions, soybean main stem node detection is affected by inter-plant occlusion, illumination variations, and complex backgrounds, resulting in inefficient and labor-intensive manual measurements that cannot satisfy the requirements of large-scale phenotypic data acquisition in breeding programs (*Liu et al., 2026*). Therefore, computer vision-based automatic detection of soybean main stem nodes is of great significance for improving phenotyping efficiency. With the development of high-throughput plant phenotyping technologies, automated acquisition of crop structural parameters and genetic evaluation has provided new approaches for intelligent breeding (*Fiorani & Schurr, 2013*).

With the rapid development of computer vision and deep learning, image-based crop phenotyping has become an important research direction in intelligent breeding and has been widely applied in agricultural object detection, disease recognition, and phenotypic parameter extraction (*McCarthy & Raine, 2022; Kamilaris & Prenafeta-Boldú, 2018*). For soybean main stem node detection, several deep learning methods have been explored. *Zhu et al. (2020)* improved the Mask-RCNN-plus network for soybean main stem phenotyping and achieved pixel-level segmentation and node counting. However, this method requires extensive pixel-level annotation and has high training costs and slow inference speed. *Yang et al. (2026)* proposed a YOLOv8_obb-based rotated object detection method to address field occlusion, scale variation, and small-object detection problems. By incorporating attention mechanisms and multi-scale feature fusion, the method improved node localization and achieved an average precision of 89.45%. Nevertheless, its relatively large model size limits its application in resource-constrained computing environments.

Besides CNN-based approaches, Transformer-based detectors, multi-model fusion methods, and three-dimensional vision techniques have also been investigated for soybean phenotyping. *Sun et al. (2026)* proposed an RT-DETR-based Transformer detection model with wavelet feature enhancement, achieving an average precision of 91.10%. However, its high computational complexity requires substantial hardware resources. *Zhou et al. (2023)* developed an SPP-extractor framework by combining YOLOv5s and U²-Net for soybean phenotypic information extraction under dense planting conditions, but the cascaded multi-model structure increased system complexity and inference time. *Cui et al. (2025)* applied three-dimensional point cloud technology for non-destructive soybean node measurement; however, this method relies on specialized acquisition equipment, resulting in high cost and operational complexity, which limits its application in rapid field phenotyping.

To overcome these limitations, researchers have conducted extensive studies on YOLO-based agricultural object detection models by introducing attention mechanisms, multi-scale feature fusion strategies, and lightweight network architectures (*Guo et al., 2021; Zhang et al., 2024*). However, many high-performance detection models still suffer from large parameter sizes, high computational complexity, and strong hardware requirements, restricting their application in resource-limited environments. Lightweight methods, including lightweight network design (*Vijayakanthan et al., 2026*), model pruning (*Ahsaei et al., 2026*), and knowledge distillation (*Jisan et al., 2026*), can effectively reduce computational costs. Nevertheless, achieving a balance among detection accuracy, inference speed, and model complexity remains challenging in agricultural intelligent detection. YOLOv11 provides strong feature extraction capability and efficient inference performance, making it a suitable baseline for soybean main stem node detection under complex field conditions (*Khanam & Hussain, 2024*).

Based on the above analysis, this study proposes a lightweight soybean main stem node detection model, BGE-YOLOv11, based on YOLOv11n. To address small-object characteristics, complex background interference, and insufficient feature representation in field environments, BiFPN-based weighted bidirectional feature fusion is introduced to enhance multi-scale feature interaction. A Global-Local Self-Attention (GLSA) module is incorporated to strengthen node feature representation and suppress background noise. Furthermore, the original detection head is replaced with the EfficientHead lightweight detection head to reduce model complexity while maintaining detection performance. This study aims to develop a lightweight detection model that achieves a balance between accuracy and computational efficiency, providing technical support for automated soybean phenotyping, breeding material selection, and intelligent breeding applications.

MATERIALS AND METHODS

Dataset Acquisition

To construct a field soybean main stem node detection dataset at the maturity stage, field experiments were conducted at the Anda Agricultural Science and Technology Park of Heilongjiang Bayi Agricultural University, Anda City, Heilongjiang Province, China. The determinate soybean cultivar Qinnong 13 was selected and planted using a single-row ridge planting pattern.

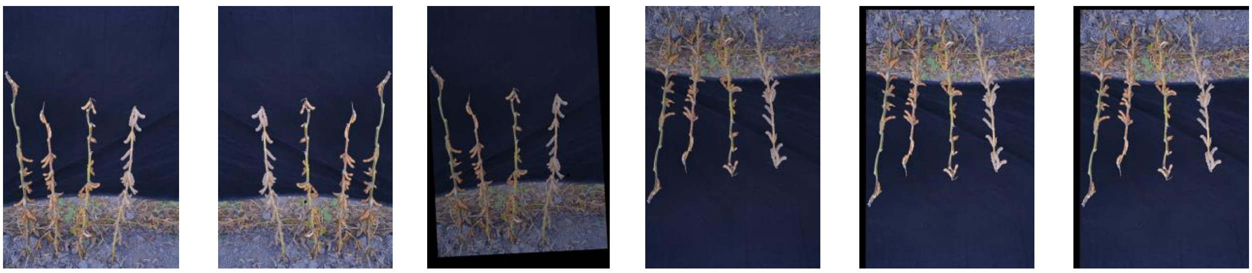
During image acquisition, field conditions involved inter-plant occlusion, interference from basal branches and senescent leaves, and complex soil backgrounds, resulting in unclear boundaries of some main stem nodes and increased annotation difficulty. To improve annotation accuracy, a local background-assisted acquisition strategy was adopted. A dark background curtain was placed behind the plants while retaining the natural soil environment at the plant base, reducing background interference while preserving field plant characteristics.

Soybean images were captured at the maturity stage using a Sony A7M3 full-frame camera. Images were collected at a fixed distance using a fixed-scale imaging protocol to ensure consistency. A total of 680 soybean plant images were obtained, with an original resolution of 4000 × 6000 pixels. All images were stored in JPEG format for subsequent annotation and model training.

Dataset Construction and Preprocessing

The collected soybean plant images were manually annotated using LabelImg. The main stem node was defined as the target category with the class name “Node”. Bounding box information, including center coordinates (x, y), width, and height, was recorded, and annotation files were generated in XML format for model training.

To prevent potential data leakage caused by augmentation, the original dataset was randomly divided into training, validation, and test sets at a ratio of 8:1:1 before data augmentation. The training set contained 544 images, while the validation and test sets each contained 68 images. Data augmentation was applied only to the training set, including brightness adjustment, rotation, horizontal flipping, vertical flipping, Cutout, random cropping, and translation. After augmentation, the training set was expanded to 1,408 images, resulting in a final soybean main stem node detection dataset containing 1,544 images. The validation and test sets maintained the original data distribution and were used for model optimization and final performance evaluation, respectively. Examples of augmented images are shown in Fig. 1.



Original image | Cutout + mirroring | Rotation + brightness adjustment | Flipping + brightness adjustment | Flipping + translation | Mirroring + rotation

Fig. 1 – Data augmentation examples for soybean plant images

BGE-YOLOv11 Model

To improve the performance of YOLOv11 for mature soybean main stem node detection under field conditions, a lightweight detection model named BGE-YOLOv11 was proposed to address challenges including inter-plant occlusion, weak small-object feature representation, and model complexity. Based on YOLOv11n, the proposed model incorporates improvements in feature fusion, attention mechanism, and detection head design. Specifically, BiFPN, a Global-Local Self-Attention (GLSA) module, and an EfficientHead detection head were introduced to enhance multi-scale feature representation and reduce computational complexity while maintaining detection accuracy. The overall architecture of BGE-YOLOv11 is shown in Fig. 2.

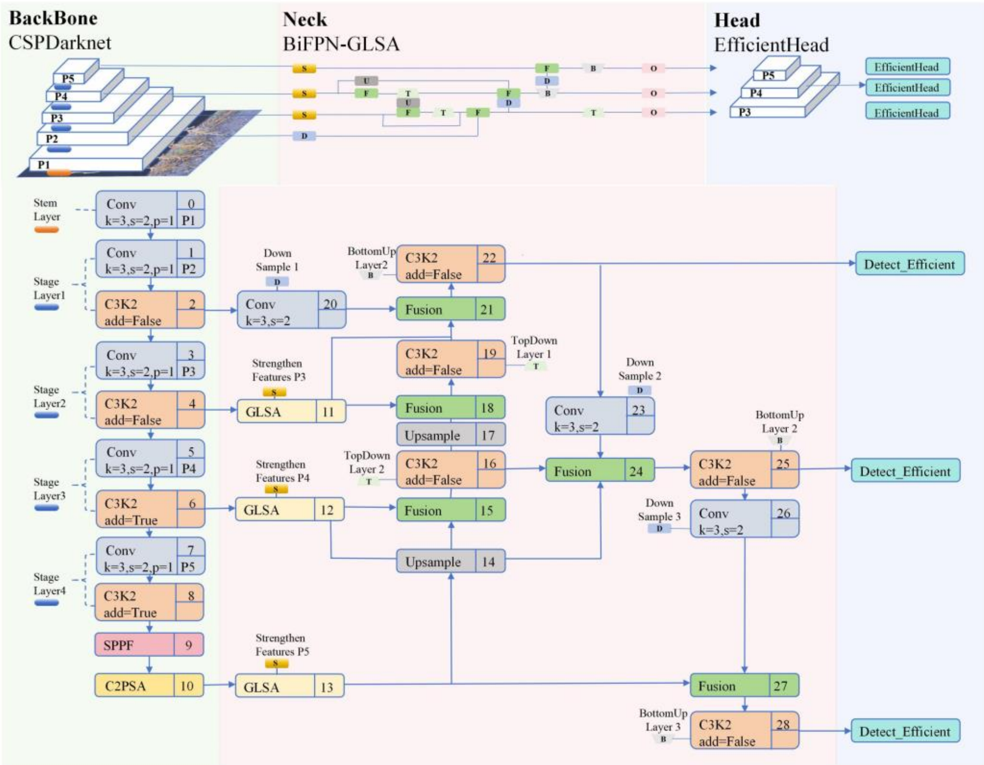


Fig. 2 – The improved BGE-YOLOv11 model structure

BiFPN Weighted Bidirectional Feature Pyramid Network

Inter-plant occlusion is a major challenge in mature soybean plants under field conditions, which may weaken multi-scale node features. The original PANet structure in YOLOv11 adopts a unidirectional feature aggregation strategy with fixed-weight feature fusion, limiting its ability to adaptively enhance occluded target features. This may result in insufficient cross-scale feature interaction and weak representation of small soybean main stem nodes.

To improve these limitations, a weighted bidirectional feature pyramid network (BiFPN) was introduced to replace the original PANet structure, forming an efficient multi-scale feature fusion module with enhanced occlusion robustness. The detailed architecture is shown in Fig. 3 (Doherty et al.,2025). By constructing bidirectional cross-scale connections and removing redundant feature nodes, BiFPN enables effective information exchange among different feature levels. Meanwhile, the learnable weighting mechanism adaptively adjusts the contribution of features from different scales, facilitating the fusion of high-level semantic information and low-level spatial details.

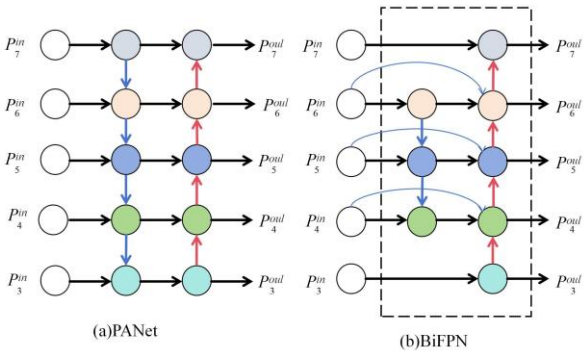


Fig. 3 – The network model structure diagram of PANet and BiFPN

During feature fusion, BiFPN adopts top-down and bottom-up pathways to achieve bidirectional information flow. Fast normalized weighted fusion is applied to reduce instability caused by feature weight differences. Compared with PANet, BiFPN improves multi-scale feature fusion quality with limited additional computational cost, thereby enhancing the feature representation of small and occluded soybean main stem nodes and improving detection performance under complex field conditions.

GLSA Global Local Self-Attention Module

Field-grown soybean plants often contain soil particles, senescent leaves, and scattered branches around the plant base, resulting in complex background interference. Meanwhile, the weak feature representation of small soybean main stem nodes makes accurate detection challenging. To enhance feature representation and suppress background noise while maintaining lightweight characteristics, a Global-Local Self-Attention (GLSA) module was embedded at the input of the neck feature pyramid.

Conventional convolutional neural networks have limited capability in modeling long-range dependencies and global contextual information, whereas standard self-attention mechanisms usually introduce high computational costs. Therefore, GLSA was adopted to improve global feature modeling capability under lightweight constraints (*Jia et al., 2026*). As shown in Fig. 4, GLSA employs a dual-branch parallel architecture. The local branch uses depthwise separable convolution to extract local detail features and preserve sensitivity to node edge and texture information. The global branch applies region-based self-attention to establish long-range dependencies among different regions with low computational overhead.

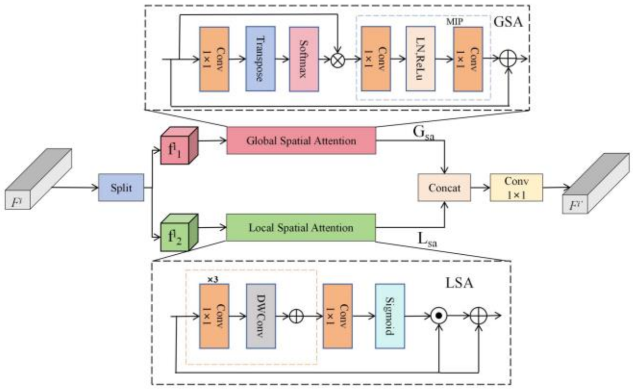


Fig. 4 – GLSA structure diagram

By embedding GLSA into the P3, P4, and P5 multi-scale feature layers before feature fusion, the model enhances target feature representation and reduces interference from complex field backgrounds. This design strengthens the feature responses of small soybean main stem nodes and provides more discriminative features for subsequent multi-scale feature fusion.

EfficientHead Lightweight Detection Head

The original YOLOv11 detection head has limited multi-scale feature alignment capability and inefficient channel information utilization, which may affect feature representation and localization performance. To improve detection performance while maintaining lightweight characteristics, an EfficientHead lightweight detection head was introduced to replace the original structure (Fig. 5) (*Jiang et al., 2025*).

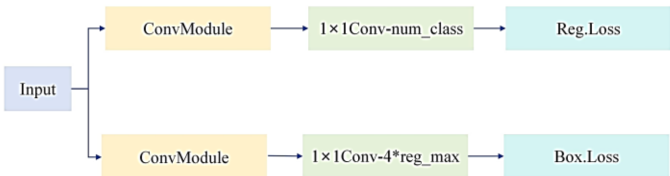


Fig. 5 – EfficientHead structure diagram

The proposed detection head integrates multi-scale feature alignment, lightweight channel attention, and dynamic localization decoding mechanisms. These designs enhance feature fusion, strengthen important soybean main stem node features, and suppress background interference. Without significantly increasing computational cost, EfficientHead improves the detection capability of weak-texture and occluded soybean main stem nodes while maintaining model lightweight characteristics.

Experimental Environment

The experiments were conducted using the PyTorch deep learning framework. The operating system was Windows 10, with an Intel® Core™ i7-14700HX CPU and an NVIDIA GeForce RTX 4060 Laptop GPU. The system was equipped with 16 GB of RAM and 8 GB of GPU memory. The software environment included Python 3.9, PyTorch 2.2.0, and CUDA 12.6. During model training, the input image size was set to 640 × 640 pixels, with a batch size of 16 and 200 training epochs.

The stochastic gradient descent (SGD) optimizer was adopted, with an initial learning rate of 0.01. Pre-trained weights were used for all models, and Mosaic augmentation was disabled during the final 10 epochs.

Model-Evaluation Index

For soybean main stem node detection, precision (P), recall (R), average precision (AP), mean average precision (mAP), and inference time were adopted to evaluate model performance. The corresponding calculation formulas are as follows:

$$P = \frac{TP}{TP+FP} \tag{1}$$

$$R = \frac{TP}{TP+FN} \tag{2}$$

$$AP = \sum_{i=1}^{N} P(i)\Delta R(i) \tag{3}$$

$$mAP = \frac{\sum_{i=1}^{k} AP_i}{k} \tag{4}$$

where TP, FP, and FN represent the numbers of true positive, false positive, and false negative detections of soybean main stem nodes, respectively. N denotes the number of samples, and k represents the number of categories.

To further evaluate the node counting performance of the proposed model, mean absolute error (MAE), root mean square error (RMSE), mean bias (Bias), and coefficient of determination (R^2) were introduced for quantitative analysis. The calculation formulas are as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{5}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{6}$$

$$Bias = \frac{1}{N}\sum_{i=1}^{N} (\hat{y}_i - y_i) \tag{7}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \tag{8}$$

where y_i represents the manually counted number of soybean main stem nodes, $\hat{y}_i$ denotes the number of nodes predicted by the model, and N represents the number of samples.

RESULTS AND ANALYSIS

Comparison of Detection Performance of Basic Detection Network

To validate the effectiveness of BGE-YOLOv11, it was compared with representative object detection models, including YOLOv5n, YOLOv6n, YOLOv7n, YOLOv8n, YOLOv9t, YOLOv10n, YOLOv11n, YOLOv12n, and RT-DETR-L. All models were trained and tested on the same soybean main stem node dataset. FLOPs, number of parameters, precision (P), recall (R), average precision (AP), memory footprint, and inference time were selected as evaluation metrics. The results are shown in Table 1.

Table 1

Comparison of different object detection models for soybean main stem node detection

Model	FLOPs [G]	Params [M]	P [%]	R [%]	mAP50 [%]	Memory footprint [MB]	Inference time [ms]
YOLOv5n	5.8	2.18	90.80	88.70	94.30	9.1	1.4
YOLOv6n	11.5	4.15	90.30	88.20	94.00	16.9	1.5
YOLOv8n	6.8	2.68	90.90	89.20	94.60	11.1	2.3
YOLOv9t	6.4	1.73	90.50	89.00	94.30	7.8	3.5
YOLOv10n	6.5	2.26	91.70	87.50	94.30	11.3	2.5

Model	FLOPs [G]	Params [M]	P [%]	R [%]	mAP50 [%]	Memory footprint [MB]	Inference time [ms]
YOLOv11n	6.3	2.58	91.00	88.90	94.60	10.7	2.7
YOLOv12n	7.80	2.50	91.30	88.10	94.30	10.6	3.7
RTDETR-L	100.6	28.44	87.00	84.10	89.20	117.8	16.3
BGE-YOLOv11	5.6	1.87	94.90	92.90	97.60	8.0	3.0

As shown in Table 1, BGE-YOLOv11 achieved a precision of 94.9%, a recall of 92.9%, and an mAP50 of 97.6%, providing superior detection performance among the compared models. In terms of lightweight characteristics, BGE-YOLOv11 required only 5.6 G FLOPs, 1.87 M parameters, and 8.0 MB memory footprint. Meanwhile, its inference time was lower than YOLOv9t, YOLOv10n, YOLOv12n, and RT-DETR-L.

Compared with the baseline YOLOv11n, BGE-YOLOv11 improved precision, recall, and mAP50 by 3.9, 4.0, and 3.0 percentage points, respectively, while reducing FLOPs by 0.7 G and parameters by 0.71 M. These results demonstrate that BGE-YOLOv11 achieves an effective balance between detection accuracy and model lightweight characteristics.

Ablation Test

An ablation study was conducted to evaluate the contribution of each improvement module to the YOLOv11n-based soybean main stem node detection model. Based on the constructed soybean main stem node dataset, different model variants were compared to verify the effectiveness of each module. The results are shown in Table 2.

Table 2

Ablation test results

Model	BiFPN	GLSA	EfficientHead	FLOPs[G]	Params[M]	P[%]	R[%]	mAP50[%]	Memory footprint [MB]	Inference time[ms]
YOLOv11n	-	-	-	6.3	2.58	91.00	88.90	94.60	10.7	2.7
B-YOLOv11	✓	-	-	6.3	1.92	92.00	89.20	95.10	8.1	2.1
G-YOLOv11	-	✓	-	8.6	3.73	91.80	90.40	95.50	15.5	3.1
E-YOLOv11	-	-	✓	5.1	2.31	92.20	89.90	95.60	9.6	2.0
BG-YOLOv11	✓	✓	-	6.7	2.07	91.30	90.10	95.20	8.8	2.8
BE-YOLOv11	✓	-	✓	5.2	1.73	91.80	89.90	95.10	7.3	2.3
GE-YOLOv11	-	✓	✓	7.4	3.48	91.50	89.00	94.60	14.4	3.2
BGE-YOLOv11	✓	✓	✓	5.6	1.87	94.90	92.90	97.60	8.0	3.0

Note: ✓ indicates that the module is used; - indicates that the module is not used.

To investigate the contribution of each module in BGE-YOLOv11, BiFPN, GLSA, and EfficientHead were individually incorporated into the YOLOv11n baseline model. The ablation results are shown in Table 2.

The single-module experiments indicate that all three modules improved model performance to different degrees. After introducing BiFPN, the mAP50 increased to 95.10%, while the parameters decreased from 2.58 M to 1.92 M, demonstrating improved feature fusion capability with reduced model complexity. The integration of GLSA increased recall and mAP50 by 1.5 and 0.9 percentage points, respectively, indicating enhanced feature representation under complex field backgrounds. With EfficientHead, the mAP50 increased to 95.60%, while FLOPs and inference time decreased to 5.1 G and 2.0 ms, respectively, showing that this module effectively reduces computational cost while maintaining detection accuracy.

When all three modules were combined, BGE-YOLOv11 achieved the best overall performance. Compared with YOLOv11n, BGE-YOLOv11 improved precision, recall, and mAP50 by 3.9, 4.0, and 3.0 percentage points, reaching 94.90%, 92.90%, and 97.60%, respectively. Meanwhile, the parameters were reduced from 2.58 M to 1.87 M, FLOPs decreased from 6.3 G to 5.6 G, and memory footprint decreased from 10.7 MB to 8.0 MB, with only a 0.3 ms increase in inference time. These results demonstrate that BGE-YOLOv11 enhances detection accuracy while reducing model complexity, achieving an effective balance between detection performance and lightweight characteristics.

Model Recognition Results

Based on the model comparison results, YOLOv8n, YOLOv11n, YOLOv12n, and the proposed BGE-YOLOv11 were selected for soybean main stem node detection experiments. Quantitative performance evaluation and visual comparison experiments were conducted to further analyze the detection capability of different models. Three representative field scenarios were considered, including fully visible soybean plants, interference from senescent leaves and branches near the plant base, and inter-plant occlusion. The samples used for both visualization analysis and quantitative node counting evaluation were collected from field experiments. All models were evaluated using the same detection settings, with a confidence threshold of 0.5 and an IoU matching threshold of 0.45.

Quantitative Performance Statistical Test

To further evaluate the soybean main stem node counting capability of the proposed model under practical field conditions, an independent node counting validation dataset was constructed. The dataset contained 44 field-collected images, including 108 soybean plants and 1,384 main stem nodes. The node numbers were manually counted and recorded as ground truth for comparison with model predictions. This dataset was not involved in model training and was only used to assess the node recognition and counting performance under field conditions.

It should be noted that the Precision, Recall, and mAP50 values in Table 1 were calculated from the model test set and used to evaluate overall object detection performance. In contrast, the evaluation metrics in Table 3 were obtained from the independent field counting validation dataset and were used to assess the practical node counting capability of the models. Since these two evaluations represent different tasks and experimental scenarios, their results should not be directly compared.

Table 3

Quantitative evaluation of soybean main stem node counting performance

Model	TP	FN	FP	Total Ground-Truth Nodes	Detection Success Rate	Miss Rate	False Alarm Rate
BGE-YOLOv11	1333	51	149	1384	96.32%	3.69%	10.05%
YOLOv11n	1312	72	171	1384	94.80%	5.20%	11.52%
YOLOv8n	1312	72	166	1384	94.80%	5.20%	11.22%
YOLOv12n	1297	87	136	1384	93.71%	6.29%	9.50%

Note: TP, FN, and FP represent correctly detected, missed, and falsely detected nodes, respectively. Total ground-truth nodes = TP + FN. Detection success rate = TP/(TP+FN), miss rate = FN/(TP+FN), and false alarm rate = FP/(TP+FP).

As shown in Table 3, BGE-YOLOv11 achieved superior performance in soybean main stem node counting. The proposed model correctly detected 1,333 nodes, with a detection success rate of 96.32% and a miss rate of 3.69%. Compared with YOLOv11n, YOLOv8n, and YOLOv12n, BGE-YOLOv11 increased the number of correctly detected nodes by 21, 21, and 36, respectively, while improving the detection success rate by 1.52, 1.52, and 2.61 percentage points and reducing the miss rate by 1.51, 1.51, and 2.60 percentage points.

Regarding false detection control, BGE-YOLOv11 achieved a false alarm rate of 10.05%, which was lower than YOLOv11n and YOLOv8n but slightly higher than YOLOv12n. This result indicates that the proposed model improves feature extraction and node recognition under complex field conditions, achieving a better balance between detection accuracy and false detection control.

Visualization Effect Comparison Test

To further evaluate the detection performance of the proposed model under complex field conditions, the results under different scenarios were visualized and compared, as shown in Fig. 6. Green bounding boxes indicate correctly detected targets, blue bounding boxes represent missed targets, and red bounding boxes denote false detections.

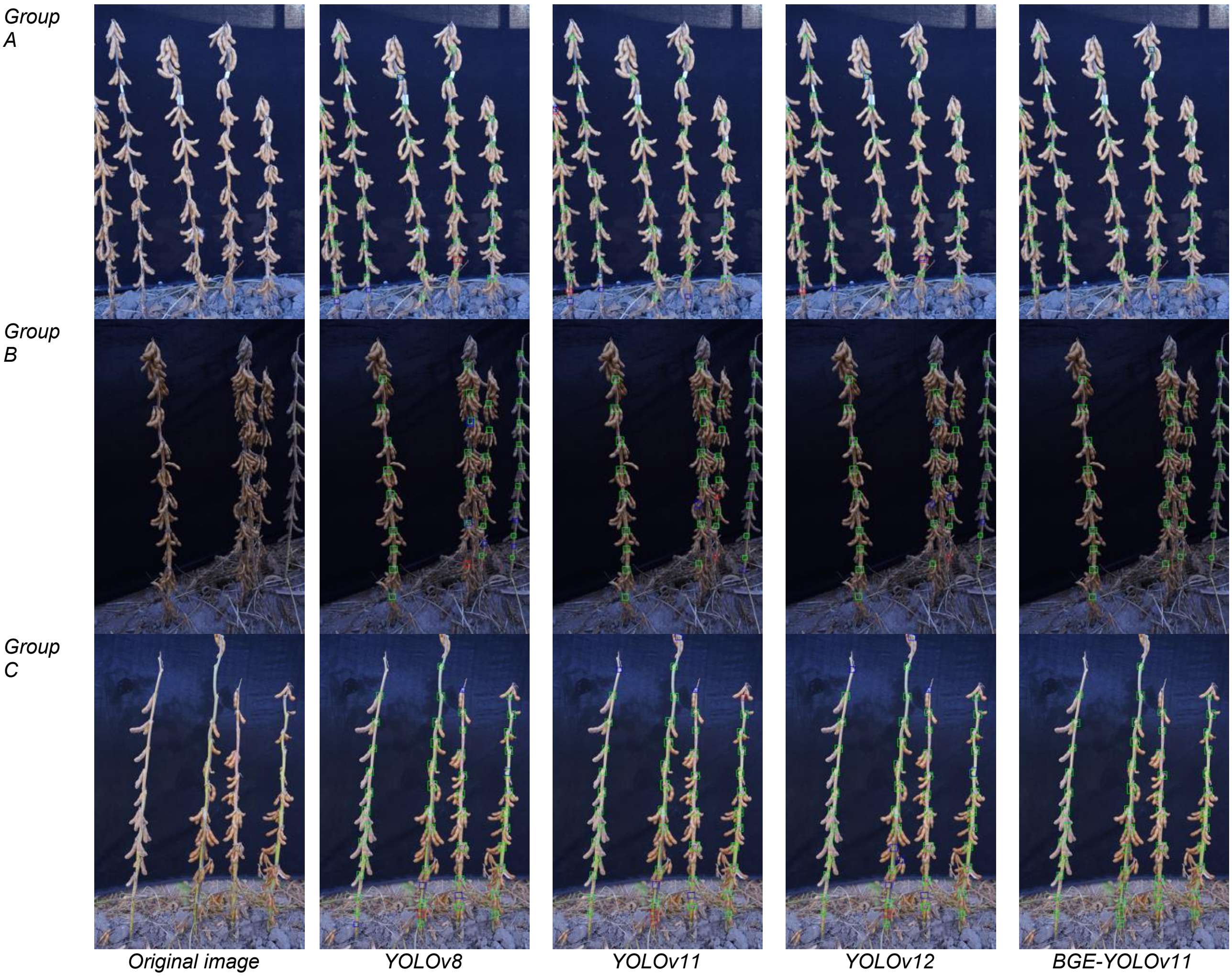


Fig. 6–Comparison of detection effects of four network models

As shown in Fig. 6, in Group A, YOLOv11n produced missed and false detections when soybean main stem nodes were located near image boundaries. YOLOv8n and YOLOv12n showed two or more detection errors, whereas BGE-YOLOv11 generated only one false detection at the basal stem region, demonstrating better boundary localization capability.

Under the severe occlusion conditions in Group B, BGE-YOLOv11 maintained relatively stable detection performance despite strong inter-plant overlap. However, missed detections still occurred in areas with densely distributed pods, while the baseline models exhibited more missed and false detections. In Group C, complex background interference from senescent leaves and scattered branches affected detection performance. Nevertheless, BGE-YOLOv11 achieved more reliable detection results, whereas the baseline models showed obvious false and missed detections.

Furthermore, BGE-YOLOv11 generally produced higher confidence scores than the baseline models, indicating improved feature extraction capability and more reliable node detection. Figure 7 shows enlarged views of the false and missed detection regions in Fig. 6. Compared with YOLOv11n, BGE-YOLOv11 improves node feature representation through enhanced feature extraction and multi-scale feature fusion, reducing missed detections under complex field conditions and achieving a better balance between detection accuracy and false detection control.

YOLOv11

BGE-YOLOv11

Fig. 7 – Partial detection results of partial amplification schematic diagram

Prediction Error Evaluation of Soybean Main Stem Node Number

To further evaluate the soybean main stem node counting capability of the proposed model under field conditions, an independent dataset containing 44 field-collected images was used for node number estimation experiments. This dataset was not involved in model training or parameter optimization. The manually counted node numbers were used as ground truth, and MAE, RMSE, Bias, and R^2 were adopted to evaluate the estimation performance of YOLOv11n and BGE-YOLOv11. The results are presented in Table 4.

Table 4

Comparison of soybean main stem node number estimation performance among different models

Model	MAE	RMSE	Bias	R^2
BGE-YOLOv11	2.09	2.82	-0.86	0.96343
YOLOv11n	3.07	3.71	-2.57	0.93895

As shown in Table 4, BGE-YOLOv11 achieved better performance in soybean main stem node number estimation. Compared with YOLOv11n, BGE-YOLOv11 reduced MAE and RMSE by 0.98 and 0.89, respectively, and its Bias value was closer to zero, indicating lower estimation error and higher prediction stability. Meanwhile, the R^2 value of BGE-YOLOv11 increased from 0.93895 to 0.96343, demonstrating a stronger agreement between the estimated node numbers and the manually counted ground-truth values.

CONCLUSIONS

In this study, YOLOv11n was selected as the baseline model. By introducing the BiFPN module to enhance multi-scale feature fusion, incorporating the GLSA global-local self-attention module to improve small-object feature representation, and adopting the EfficientHead lightweight detection head to reduce computational cost, a lightweight BGE-YOLOv11 model was developed for soybean main stem node detection and counting. Experimental results demonstrated that the proposed model achieved an mAP50 of 97.6% under complex field conditions, with only 1.87 M parameters and an inference time of 3.0 ms. It improved detection accuracy while maintaining favorable lightweight characteristics, providing technical support for the automated acquisition of soybean phenotypic information.

Currently, this study mainly focuses on main stem node detection in mature soybean plants. In future work, the proposed method will be further extended to other phenotypic parameters, such as stem diameter and plant height, as well as different growth stages. Moreover, correlation models between phenotypic traits and yield-related characteristics will be established to provide data support for precision breeding.

ACKNOWLEDGEMENT

This work was supported by the Doctoral Research Startup Fund Project of Heilongjiang Bayi Agricultural University [grant number XDB202501].

REFERENCES

[1] Ahsaei, S., Raji, M., & Ghavami, B. (2026). RAP: A reliability-aware pruning framework for deep neural networks. *Journal of Systems Architecture, 176*, 103829. https://doi.org/10.1016/j.sysarc.2026.103829

[2] Cui, D. H. (2025). *Research on extraction methods of mature soybean plant phenotypic parameters based on point cloud data* [Master's thesis, Northeast Agricultural University]. https://doi.org/10.27010/d.cnki.gdbnu.2025.000329

[3] Doherty, J., Gardiner, B., Kerr, E., & Siddique, N. (2025). BiFPN-YOLO: One-stage object detection integrating bi-Directional feature pyramid networks. *Pattern Recognition, 160*, 111209. https://doi.org/10.1016/j.patcog.2024.111209

[4] Fiorani, F., & Schurr, U. (2013). Future scenarios for plant phenotyping. *Annual Review of Plant Biology, 64*, 267–291. https://doi.org/10.1146/annurev-arplant-050312-120137

[5] Guo, R., Yu, C. Y., He, H., Zhao, Y. J., Yu, H., & Feng, X. Z. (2021). Detection method for pod number of single soybean plant using improved YOLOv4 algorithm. *Transactions of the Chinese Society of Agricultural Engineering, 37*(18), 179–187.

[6] Jia, B., Wang, X. C., Shi, Y. Y., Zhang, X. L., Xu, Z., Wang, J. H., Yang, H. H., & Qian, D. W. (2026). Efficient and lightweight model for detecting juvenile fish feeding behavior using feed pellet key point analysis. *Computers and Electronics in Agriculture, 240*, 111141. https://doi.org/10.1016/j.compag.2025.111141

[7] Jiang, T. T., Li, L., Zhang, Z., Yu, X., Zhu, Y. Q., Li, L. M., Liu, Y. D., Bai, Y. L., Tang, Z. Q., Liu, S. B., Zhang, Y., Duan, Z., Yin, D. M., & Jin, X. L. (2026). YOLO-light-pruned: A lightweight model for monitoring maize seedling count and leaf age using near-ground and UAV RGB images. *Artificial Intelligence in Agriculture, 16*, 164–186. https://doi.org/10.1016/j.aiia.2025.10.002

[8] Jisan, M. K., Hoque, K. E., Azhar, T., & Newton, M. H. (2026). Rice leaf disease classification using resource-efficient deep learning models via response-based knowledge distillation. *Applied Soft Computing, 199*, 115356. https://doi.org/10.1016/j.asoc.2026.115356

[9] Kamilaris, A., & Prenafeta-Boldú, F.X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture, 147*, 70–90. https://doi.org/10.1016/j.compag.2018.02.016

[10] Khanam, R., & Hussain, M. (2024). YOLOv11: An overview of the key architectural enhancements. *arXiv*. https://doi.org/10.48550/arXiv.2410.17725

[11] Liu, F., Wu, Q., Han, Z., Zhao, L., Pang, S., & Wang, S. (2026). Automatic stem phenotyping in soybean using keypoint detection. *Engineering Applications of Artificial Intelligence, 165*, 113300. https://doi.org/10.1016/j.engappai.2025.113300

[12] McCarthy, A., & Raine, S. (2022). Automated variety trial plot growth and flowering detection for maize and soybean using machine vision. *Computers and Electronics in Agriculture, 194*, 106727. https://doi.org/10.1016/j.compag.2022.106727

[13] Qiu, L. J., Chang, R. Z., Wang, K. J., Wang, X. M., Guan, R. X., Li, Y. H., Yang, T., Zhang, B., Zhao, L. M., Hao, C. Y., Yuan, C. P., Guo, Y., & Dong, Y. S. (2006). *Descriptors and data standard for soybean germplasm resources*. China Agriculture Press.

[14] Qu, H. L., Jiang, Z., Li, J., & Zhao, L. (2022). Development status of global soybean industry and suggestions for China's soybean industry development. *Soybean Science & Technology*, (5), 28–33, 39.

[15] Sun, L. Q., Zhang, Q. Y., He, Y. H., Guo, K., Yang, C., Su, W. H., & Tian, Z. X. (2026). Multi-phenotype measurement and yield prediction of soybean plants based on RT-DETR. *Transactions of the Chinese Society of Agricultural Engineering, 42*(3), 192–200.

[16] Vijayakanthan, G., Nasser, M., Algbear, A. F., Alosaimi, H. M., Ibrahim, A. O., & Ravi, V. (2026). Explainable and lightweight deep learning models for tea leaf disease detection: A systematic review. *Smart Agricultural Technology, 14*, 102042. https://doi.org/10.1016/j.atech.2026.102042

[17] Yang, Y. X., Li, J. Y., Shi, W. Q., Qi, L. Q., & Zhang, W. (2026). Research on soybean main stem node recognition based on improved YOLOv8-obb. *Journal of Chinese Agricultural Mechanization, 47*(1), 79–86. https://doi.org/10.13733/j.jcam.issn.2095-5553.2026.01.012

[18] Zhang, Q.Y., Fan, K. J., Tian, Z., Guo, K., & Su, W. H. (2024). High-precision automated soybean phenotypic feature extraction based on deep learning and computer vision. *Plants, 13*(18), 2613.

[19] Zhou, W., Chen, Y. J., Li, W. H., Zhang, C., Xiong, Y. J., Zhan, W., Huang, L., Wang, J., & Qiu, L. J. (2023). SPP-extractor: Automatic phenotype extraction for densely grown soybean plants. *The Crop Journal, 11*(5), 1569–1578.

[20] Zhu, R. S., Li, S., Yan, X. H., Yan, Z. Z., Yu, J. L., Shi, J., Su, X. Y., Cao, Y. Y., Xin, D. W., Li, Y., Jiang, H. W., Zhao, Z. Q., Zhang, Z. G., & Chen, Q. S. (2020). Automatic precise acquisition of soybean main stem phenotypes based on semantic segmentation and its application in introgression line selection. *Chinese Journal of Oil Crop Sciences, 42*(1), 61–70. https://doi.org/10.19802/j.issn.1007-9084.2020028