

Article

Not peer-reviewed version

---

# Ginseng Seed Quality Detection Based on YOLO-GS

---

[Lixin Hou](#), [Ning Wei](#), Mengke Wang, Xiaoran Yu, Wenshuang Tu, [Jing Zhou](#), [Hongjun Gu](#)\*

Posted Date: 9 December 2025

doi: 10.20944/preprints202512.0660.v1

Keywords: ginseng seeds; object detection; deep learning; YOLO-GS



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Ginseng Seed Quality Detection Based on YOLO-GS

Lixin Hou, Ning Wei, Mengke Wang, Xiaoran Yu, Wenshuang Tu, Jing Zhou and Hongjun Gu \*

College of Information and Technology, Jilin Agricultural University, Changchun 130118, China

\* Correspondence: guhongjun@jlau.edu.cn

## Abstract

Seed quality is a crucial factor in determining yield before sowing. Ginseng seeds undergo several processes before sowing, including picking, washing, and germination. The germination process is susceptible to damage or failure, which can directly impact the final yield of subsequent cultivation. Therefore, precise and reliable quality inspection and screening must be done before sowing to ensure a high germination rate. Based on YOLOv11n, this study proposes the YOLO-GS model to test the quality of ginseng seeds. Firstly, a SELP module was designed to enhance the network's ability to focus on the key features of ginseng seeds and improve the model's detection accuracy. Secondly, the Channel Prior Convolution Attention (CPCA) mechanism was introduced into the backbone network to dynamically assign attention weights to the feature map in both the channel and spatial dimensions, thereby enhancing the network's ability to extract features from the target. Thirdly, the C3k2 structure in the backbone was improved to account for both local feature extraction and global dependency modeling, thereby enhancing the model's accuracy. Finally, a Convolutional Attention Module (CloFormerAttnConv) based on the multi-frequency position-sensitive attention mechanism in C2PSA was introduced to achieve a dual perception of local details and global semantics while maintaining computational efficiency and improving feature extraction capabilities. The experimental findings demonstrated that the YOLO-GS model attained 97.7% mAP@0.5, with Precision, Recall, F1-Score, and mAP@0.5:0.95 reaching 96.7%, 96.4%, 90.5% and 90.3%, respectively. The model has only 4.2 million parameters. When deployed on the Jetson edge device, the model inference time is 0.6ms, providing an effective solution for real-time target detection tasks in the application of seed quality assessment of ginseng. In conclusion, the YOLO-GS model will be applicable for the precise detection of ginseng seed quality.

**Keywords:** ginseng seeds; object detection; deep learning; YOLO-GS

## 1. Introduction

Ginseng is a valuable herb widely used in health foods, medicines, and cosmetics [1]. In recent years, demand for high-quality ginseng has continued to increase. China is the world's largest ginseng producer, cultivating around 50% of the global total, with an annual output of 5,000–6,000 tonnes of dried ginseng [2]. However, this is still not enough to meet market demand. A significant challenge in ginseng cultivation is the issue of continuous cropping. After harvesting, it is necessary to wait 10 to 20 years before replanting; otherwise, problems such as burnt whiskers and rotting roots can arise, severely diminishing yields and hindering the industry's sustainable development [3]. To alleviate the contradiction between supply and demand, the ginseng industry is now actively promoting artificial breeding, where seed quality has become a crucial factor in determining yield. The quality of germinated ginseng seeds is primarily reflected in germination success or failure and defects, which directly impact the crop's germination rate and, consequently, the final yield. Currently, two primary methods for assessing seed quality are the ginseng citizen experience method and the seed opening rate method. The empirical identification method is challenging to promote widely because it requires the destruction of seeds to observe the internal germ's morphological characteristics. In contrast, the seed opening rate method is non-destructive and relies on visual

observation, but it can be affected by subjective bias. A fast, precise, and non-destructive method is needed to effectively detect the quality of ginseng seeds before sowing.

In recent years, machine vision technology has been widely adopted in agriculture, leading to significant advancements in crop inspection. Kurtulmuş and Kavdir et al. [4] proposed a computer vision algorithm based on conventional color images to accurately detect corn cobs, achieving an accuracy rate of 81.6%. Wu et al. [5] introduced a method based on image analysis and a support vector machine that uses image processing techniques to extract six typical feature parameters from each corn kernel, thereby improving the classification accuracy of corn seeds by optimizing the classification model. Tu et al. [6] developed the AIseed software, which utilizes machine vision and deep learning techniques to automatically extract and analyze 54 phenotypic traits, including shape, color, and texture, for high-throughput seed phenotyping. It also included seed quality detection and prediction models, demonstrating high efficiency and accuracy in assessing seed quality across plant species. Fan et al. [7] developed a novel method for nondestructively evaluating maize seed viability, using a grouped hyperspectral image fusion strategy and identifying key biochemical indicators (catalase activity and malondialdehyde content). Their two-stage model achieved 90% accuracy, offering both high performance and interpretability.

Although traditional machine vision methods can achieve a certain level of precision, they are limited by their slow detection speed and the requirement to handle targets individually. Deep learning-based detection algorithms show significant advantages over traditional object detection methods in terms of both accuracy and efficiency. These algorithms are divided into single-stage and two-stage types. Two-stage methods Faster RCNN and Mask RCNN first generate candidate regions and then perform fine classification and regression to achieve high-precision detection [8,9]. For instance, Wu et al. [10] employed transfer learning in conjunction with Faster RCNN to count wheat grains, achieving an average accuracy of 91% across 178 test images post-threshing. Li et al. [11] proposed an automatic detection method for hydroponic lettuce seedlings based on an improved Faster RCNN and HRNet version. This method effectively improved the detection accuracy of dead and double-planted seedlings by introducing focus loss and RoI Align, achieving an Average Precision of 86.2%. Deng et al. [12] developed an automatic recognition and counting model based on a convolutional neural network for determining the size of rice main branches. Integrating the feature pyramid network (FPN) into Faster RCNN achieved accurate separation and counting of all branches in 159 rice panicle images, with an overall accuracy of 99.4%. Tan et al. [13] proposed the RiceRes2Net model based on an improved version of Cascade RCNN. This model can accurately detect and recognize rice panicles and growth stages in complex field environments, achieving an average accuracy of 96.42%.

Single-stage methods, such as the YOLO series and SSD [14], primarily achieve target localization and classification directly via an end-to-end approach. They have high processing speeds and are ideal for real-time performance tasks. Lu et al. [15] developed a method for continuously sorting drilled lotus seeds online. Their approach uses YOLOv3 for hole detection and algorithms for online sorting control and distinguishing hole attributes. During actual machine operation, this method achieved an impressive accuracy of 95.8% in sorting a mix of defective and standard samples. This work offers a valuable reference for the online sorting of crops that depend on local features. For instance, Wang et al. [16] designed an online automatic cotton seed sorting device based on the YOLOv5 model, which can swiftly identify and remove damaged and moldy cotton seeds with 99.6% accuracy. Jiao et al. [17] combined  $\mu$ CT technology with an improved YOLOv7-Tiny network to automatically detect invisible internal endosperm cracks produced during corn seed soaking, achieving a detection Precision of 93.8% with a model size of just 9.7 MB. This high-precision automated extraction of endosperm crack information from soaked corn improves the efficiency and accuracy of corn seed quality assessment. Li et al. [18] proposed a Cotton-YOLO algorithm based on YOLOv7 optimization that can quickly and efficiently detect foreign fibers in cotton seeds with an accuracy rate of up to 99.12%. Niu et al. [19] presented an improved lightweight YOLOv8 model for efficiently detecting four states of corn seeds: healthy, broken, moth-eaten, and mildewed. The model

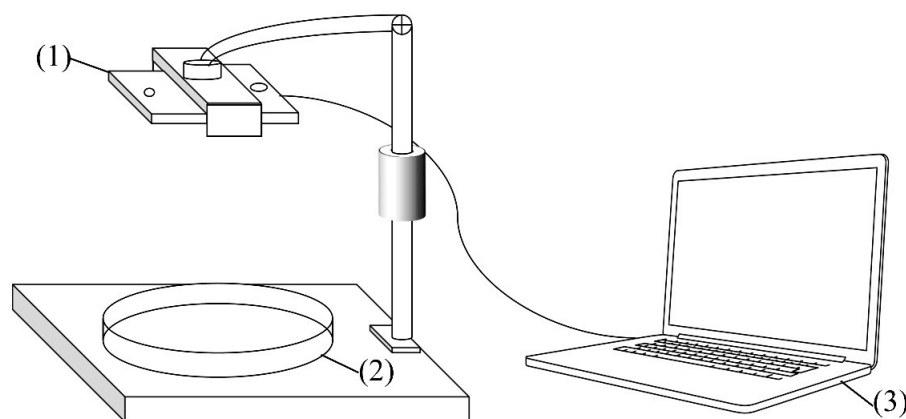
achieved an mAP@0.5 of up to 98.5%. Yi et al. [20] combined an improved YOLOv9 with binocular stereo vision, introducing DSGLSeq and MHMSI modules to significantly enhance the accuracy of fruit volume estimation in datasets such as apples, pears, pomelos, and kiwis. Systematic validation experiments conducted on major economic fruit tree varieties, including apples, pears, pomelos, and kiwifruit, demonstrate that RTFVE-YOLOv9 improved the mean Average Precision by 2.1%, 1.6%, 4%, and 3.8% respectively on the four fruit datasets compared to the baseline YOLOv9-c model. These methods have been widely applied in various real-world scenarios, meeting diverse requirements for detection speed and accuracy in different tasks. Deep learning models possess strong capabilities for feature extraction and automatic learning from end to end. They can effectively identify key semantic features from large datasets, significantly improving the models' intelligence and generalization ability. This enhancement leads to more efficient, stable, and scalable solutions for inspecting the quality of agricultural seeds.

Ginseng seeds are small and similar in appearance, especially when they fail to germinate or are damaged; their external features often resemble those of healthy seeds, making it difficult to assess their quality. This study presents the YOLO-GS model to overcome these problems. It enhances YOLOv11n by improving the integrated module in the following ways: (1) Incorporating the Channel Prior Convolutional Attention (CPCA) mechanism into the backbone network significantly enhances its ability to extract features from the target. (2) The standard convolution in the C3k2 module has been replaced with AssemFormer attention, which effectively improves the model's detection accuracy by considering both local and global information. (3) A convolutional attention module based on the multi-frequency position-sensitive attention mechanism (CloFormerAttnConv) was introduced into the C2PSA structure to enhance feature extraction while maintaining computational efficiency. (4) The SELP module combines Squeeze-and-Excitation (SE) attention with partial convolution to extract global and local features through average and max pooling. It also uses partial convolution to handle missing information adaptively, enhancing feature representation and robustness. Experiments demonstrated the model's effectiveness, achieving a detection Precision of 96.7% for ginseng seeds, 97.7% for mAP@0.5, 96.4% for Recall, 96.5% for F1-Score, and 90.3% for mAP@0.5:0.95. While maintaining high precision, it has achieved extremely high computational efficiency, meeting the speed requirements of real-time online sorting, counting and quality monitoring scenarios for modern agricultural equipment.

## 2. Materials and Methods

### 2.1. Data Collection

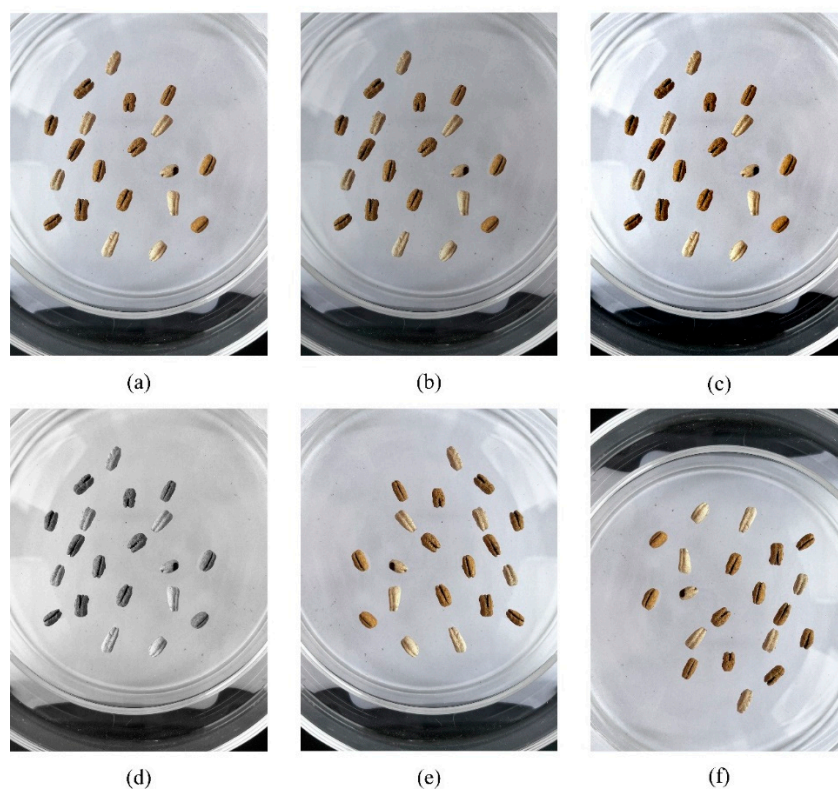
This study selected ginseng seeds harvested in August 2023 from a mountain forest in Dunhua City, Jilin Province. The seeds underwent a germination treatment, and images of the germinated seeds were captured as experimental data. A phone stand was used to mount an iPhone 13, which was positioned 14 cm away from the ginseng seeds placed in Petri dishes. The camera was oriented downward and kept parallel to the ground. Images were captured under various lighting conditions, documenting successfully germinated seeds, failed germination, and damaged seeds. A total of 1,386 images of seeds were collected, with each image containing 20 seeds. The images displayed a mix of successfully germinated seeds and those that had failed to germinate, along with a few damaged seeds. All images were stored in JPG format at 3024×4032 pixels. The equipment used for image acquisition is illustrated in **Error! Reference source not found.**



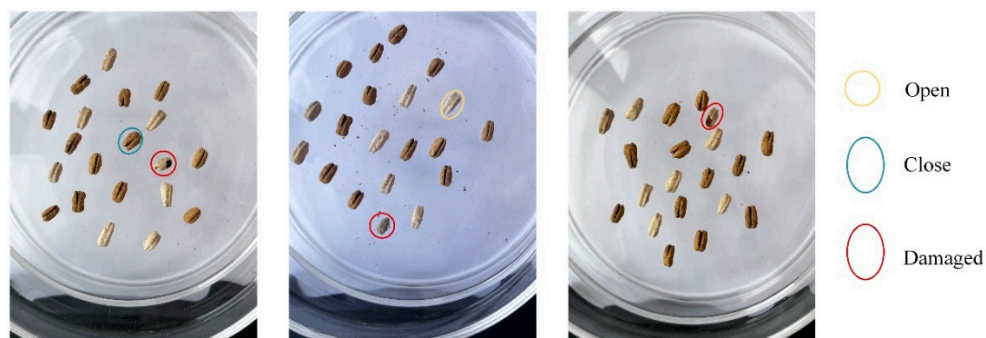
**Figure 1.** Image acquisition device (1) iphone 13 (2) petri dish (3) PC.

## 2.2. Data Preprocessing

To address the issue of insufficient data samples and to prevent overfitting and enhance the model's anti-interference ability, we applied five image enhancement operations: brightness adjustment, contrast enhancement, grayscale transformation, image flipping, and image rotation. The dataset was randomly partitioned at an 8:1:1 ratio, with the augmentation methods illustrated in **Error! Reference source not found.** We utilized Labelling software to annotate the images of ginseng seeds. Seeds were categorized into three labels: "open" for successfully germinated seeds, "closed" for those that failed to germinate, and "damaged" for seeds with visible damage. The seed categories are illustrated in **Error! Reference source not found.**



**Figure 2.** Data augmentation (a) original image (b) brightness adjustment (c) Contrast enhancement (d) gray-scale transformation (e) image flipping (f) image rotation.



**Figure 3.** Seed category.

### 2.3. YOLO-GS

This study aims to identify and evaluate the quality of germinated ginseng seeds using image detection technology. The goal is to evaluate seed viability to improve seedling emergence rates and the overall cultivation quality. Compared to other neural networks, YOLOv11n [21] integrates a lightweight design with refined capabilities for feature expression. This design reduces computational complexity while maintaining high detection accuracy. YOLOv11n strikes a superior balance between accuracy, speed, and ease of deployment, making it the baseline model of choice for this study.

The challenge in this study lies in the difficulty of visually distinguishing between ginseng seeds. We proposed the YOLO-GS model, outlined in Figure 4, to address this issue. First, we replaced the C3k2 module in the backbone network with C3k2\_AssemFormer. This enhancement improves feature extraction capabilities and sharpens the details. Second, we introduced the CPCA attention mechanism into the backbone network to enhance feature perception while maintaining a balance between model accuracy and computational efficiency. Additionally, the C2PSA\_CloFormerAttnConv replaced the C2PSA structure in the backbone network. This change helps balance convolutional efficiency with contextual awareness, increasing the model's expressive power. Finally, we introduced the proposed SELP module into the neck network. This module simultaneously captures local and global information, fully integrating channel and spatial information to prevent feature loss.



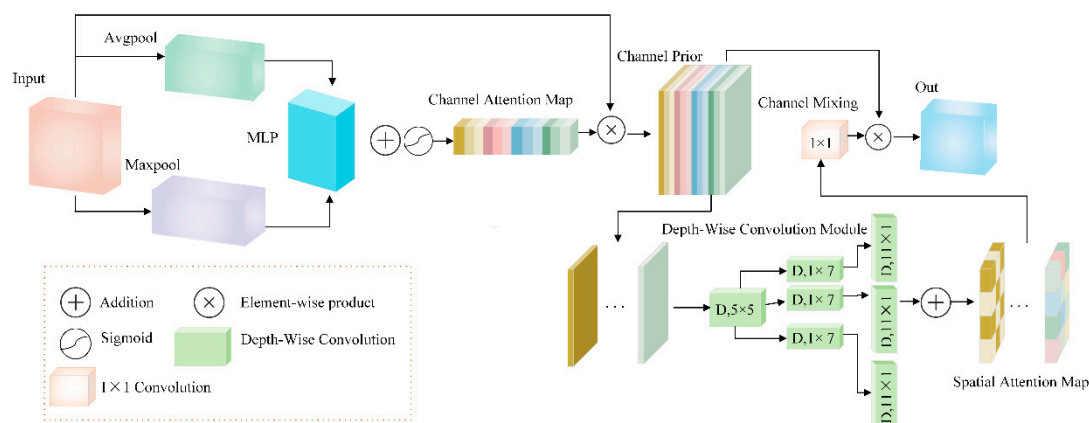
convolutions. Next, the ELU nonlinear activation function [24] further enhances feature expressiveness by better capturing nonlinear relationships, thereby improving generalization and robustness.

A second partial convolution layer then restores the channel dimension. Next, a Sigmoid function generates the final channel attention weight map, boosting the capabilities of the feature representation. Finally, during the feature fusion stage, a learnable fusion factor with an initial value of 0.5 is introduced to weight the original and enhanced features dynamically. This residual fusion strategy enables the network to adjust its ratio adaptively during training, preventing both excessive feature amplification and neglect of helpful information. This improves the model's detection accuracy and stability.

Due to the small size of ginseng seeds, their subtle morphological variations, and their susceptibility to interference from complex backgrounds, applying the SELP module to the network is essential. The module enhances the key features of ginseng seeds by dynamically adjusting the weights of critical channels, enabling a precise capture of their subtle characteristics. As a result, the model's ability to detect small to medium-sized objects is significantly improved. In this study, we integrate the SELP module into the YOLOv11n neck network, strengthening the model's focus on the critical features of ginseng seeds. This integration enhances the precision and efficiency of feature extraction while minimizing interference from redundant information. Consequently, the model achieves significantly improved recognition accuracy for ginseng seed targets and demonstrates greater robustness overall.

### 2.3.2. Channel Prior Convolutional Attention

Ginseng seeds have characteristics, such as shape and color, that make them difficult to distinguish from their background. Traditional convolutional networks have limited receptive fields, making it challenging to model long-range dependencies effectively. This limitation can result in missed detections or false positives. This study introduces the Channel Prior Convolutional Attention Module (CPCA) into the backbone network of the model [25], enhancing the expression capability of key feature regions during the feature extraction stage. The CPCA structure is shown in **Error! Reference source not found.**



**Figure 6.** Structure diagram of the Channel Prior Convolutional Attention.

CPCA is an efficient attention mechanism that enhances a model's ability to perceive key features while improving overall computational efficiency. Unlike CBAM [26], which combines channel and spatial attention while maintaining uniformity across all channels in the spatial attention map, CPCA cannot adaptively adjust to the feature requirements of different channels. The CPCA module introduces a "channel-first" strategy. Initially, it applies a channel attention mechanism (CA) to the feature map  $F$ . Then, spatial context information is aggregated via average and max pooling

operations, which are then fed into a shared multi-layer perceptron (*MLP*) to generate the feature map (*F<sub>c</sub>*). The formula is shown in (1):

$$CA(F) = \sigma(MLP(Avgpool(F)) + MLP(Maxpool)) \quad (1)$$

Here, *F* represents the input feature map,  $\sigma$  represents the sigmoid activation function, and *AvgPool* and *MaxPool* represent Average pooling and Max pooling, respectively.

Subsequently, *F<sub>c</sub>* is processed through the spatial Attention mechanism (*SA*), using multi-scale depth-separable convolutional modules to capture the spatial relationships between features while maintaining channel independence, effectively enhancing the spatial attention corresponding to each channel. Finally, the spatial attention map is generated by channel blending through 1x1 standard convolution, as shown in formula (2):

$$SA(F_c) = \text{Conv}_{1 \times 1}(\sum_{i=0}^3 \text{Branch}_i(\text{DwConv}(F_c))) \quad (2)$$

*DwConv* represents depth-separable convolution, and *Branch<sub>i</sub>*  $\epsilon$  represents the *i*<sup>th</sup> branch. A *Branch<sub>i=0</sub>* is an identity connection.

This allows for the detailed modeling of feature information across spatial and channel dimensions. The CPCA module highlights salient features of ginseng seeds and improves the model's ability to perceive boundary information while keeping computational overhead low.

### 2.3.3. C2PSA\_CloFormerAttnConv Module

In recent years, attention mechanisms have demonstrated remarkable performance in object detection tasks by enhancing models' feature extraction capabilities. However, existing methods primarily focus on modeling low-frequency global information through sparse attention while simplifying the modeling of high-frequency local features. This approach struggles to meet the demand for fine-grained feature representation when detecting small- to medium-sized objects.

Fan et al. [27] introduced CloFormer to address this issue, which fuses shared weights from convolutions and context-aware weights from Transformers [28].

This approach significantly improves the model's ability to capture local features while maintaining computational efficiency. Building on this, this study proposes an improved fusion attention module called C2PSA\_CloFormerAttnConv, which retains the C2f (Cross-Channel Partial Fusion) partial channel fusion mechanism from C2PSA and integrates the Clo Block's local context modeling capability into each channel branch. This enables the model to maintain local convolution efficiency while gaining stronger context awareness, effectively improving detection accuracy and robustness in scenarios with dense small objects. The C2PSA\_CloFormerAttnConv module is displayed in **Error! Reference source not found.**

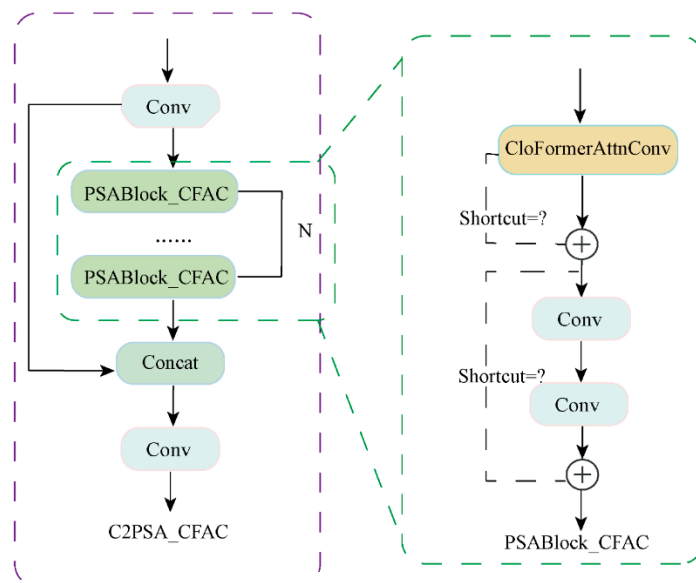


Figure 7. C2PSA\_CloFormerAttnConv module.

#### 2.3.4. C3k2\_AssemFormer Module

Traditional convolutional structures lose feature information when processing small-sized to medium-sized objects, especially in deep networks where local structures become overly compressed. This compression results in reduced localization accuracy. Although Transformers can model long-range dependencies, they struggle to express high-frequency local details and require substantial computational resources. Therefore, this study proposes the C3k2\_AssemFormer module, which combines the local modeling strengths of convolutions with the contextual modeling capabilities of Transformers. Based on the traditional C3 architecture, the C3k2\_AssemFormer module introduces the C3k component. Its larger receptive field enhances the representation of high-frequency details, such as edges and textures. The C3k2\_AssemFormer module integrates AssemFormer's [29] "local-global-local" architecture to address the limitations of convolutions in modeling long-range dependencies. The module introduces a linear self-attention mechanism [30] after local convolutions to model cross-regional context effectively. The module structure is illustrated in **Error! Reference source not found.**. The C3k2\_AssemFormer module is illustrated in **Error! Reference source not found.**

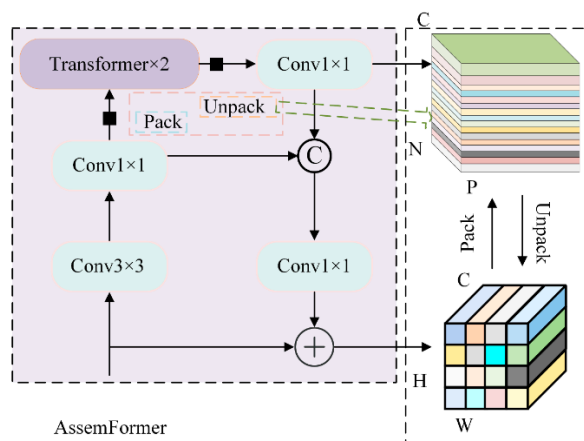


Figure 8. AssemFormer module.

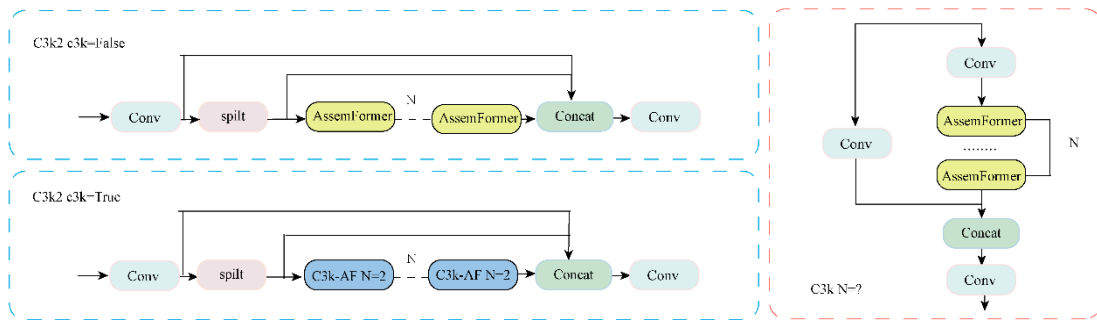


Figure 9. C3k2\_AssemFormer module.

The C3k2\_AssemFormer module enhances the model's ability to detect small-to-medium objects by integrating local convolutions and global attention. Although it requires more computing power than lightweight models, it provides superior stability and detection accuracy in complex backgrounds and scenarios with dense objects, making it ideal for ginseng seed detection.

#### 2.4. Evaluation Metrics

To validate the effectiveness of the proposed enhancement algorithm, this study comprehensively evaluated the model performance using a number of mainstream evaluation metrics. These metrics include Precision, Recall, mean Average Precision  $mAP@0.5$  intersection-over-union (IoU=0.5), the  $mAP@0.5:0.95$ , and the number of model parameters.

In binary classification tasks, each target sample can be categorized as a true positive ( $TP$ ), false positive ( $FP$ ), true negative ( $TN$ ), or false negative ( $FN$ ) based on its actual class and the predicted result. These statistics form the basis for calculating Precision, Recall, and the F1-Score.

*Precision* represents the proportion of truly positive samples among those identified as positive by the model, as shown in formula (3).

$$P = \frac{TP}{TP + FP} \times 100\% \quad (3)$$

*Recall* represents the ratio of the number of samples correctly identified as positive by the model to the total number of positive samples, as shown in formula (4).

$$R = \frac{TP}{TP + FN} \times 100\% \quad (4)$$

*Average Precision* represents the average accuracy at different recall levels, measuring a model's performance across each category. The formula is shown in (5).

$$AP = \int_0^1 P \, dR \quad (5)$$

$mAP$  primarily measures performance across all object classes.  $mAP@0.5$  represents the mean Average Precision when IoU is 0.5;  $mAP@0.5:0.95$  represents the average  $mAP$  when IoU ranges from 0.5 to 0.95 in increments of 0.05. The formula is shown in (6).

$$mAP = \frac{1}{N} \sum_{i=1}^N A P_i \quad (6)$$

*F1-Score* is used to evaluate a model's precision and completeness when detecting all targets. This metric combines *Precision* and *Recall*, with a maximum value of 1 and a minimum value of 0, as shown in formula (7).

$$F1 - Score = \frac{2 \times P \times R}{P + R} \quad (7)$$

Additionally, the number of parameters indicates the quantity of trainable weights that need to be learned within the model, reflecting its complexity and storage requirements. Comparing the differences in parameters and computational demands across models helps us evaluate the balance between performance and efficiency. Analyzing these metrics allows us to comprehensively assess a model's accuracy and robustness in target recognition tasks, measure its efficiency and resource consumption during practical deployment, and systematically compare overall algorithmic performance.

### 3. Experiments and Results

#### 3.1. Test Platform and Hyperparameter Settings

This study used the Ubuntu 22.04.3 operating system with a 6.5.0-49-generic kernel. The computer configuration included an NVIDIA GeForce RTX 4070 Ti GPU with 12 GB of memory, 32 GB of RAM, and an Intel Core i7-13700 KF processor. The PyCharm programming platform was utilized for development. The model hyperparameters were set as follows: the input image size was set to 640 x 640 pixels, the SGD optimizer was used, the batch size was set to 32, the initial learning rate was set to 0.01, and the training process lasted for 200 epochs. After multiple trials and verifications, this hyperparameter setting achieves the maximum stability and efficiency of the model, as listed in **Error! Reference source not found.**

**Table 1.** Test Platform.

Name	Configuration
CPU	i7-13700KF
GPU	Nvidia GeForce RTX 4070Ti
GPU-accelerated library	CUDA11.8
Operating system	Ubuntu22.04
Software environment	Python3.11.4

#### 3.2. Model Training Results

The improved YOLO-GS model was trained on the ginseng seed dataset, and its detection performance was evaluated using the validation set, as shown in **Error! Reference source not found.** Compared with the benchmark model YOLOv11n, YOLO-GS performed better in all evaluation metrics. The Precision, mAP@0.5 value, and F1-Score increased from 93.8%, 97.0%, and 95.0% to 96.7%, 97.7%, and 96.5%, respectively. This shows a significant improvement in overall detection performance. Further analysis revealed that the improvement in detection between the two models was relatively small for the germinated successful and germinated failed categories. However, the performance improvement was most significant for the damaged category, with a Precision improvement of 6.9%, an mAP@0.5 improvement of 2%, and an F1-Score improvement of 4%. Additionally, the YOLO-GS model maintains a detection speed of 268.7 frames per second, meeting real-time detection requirements. In summary, this model can accurately identify the quality status of ginseng seeds (successful germination seeds, failed germination seeds, and defective seeds). These results demonstrate that the improved YOLO-GS model can more easily detect complex or textured objects.

**Table 2.** Model Training Results.

Model	Category	P	R	mAP@0.5	F1-Score	mAP@0.5:0.95
YOLOv11n	all	0.938	0.963	0.970	0.950	0.897
	open	0.994	0.992	0.995	0.993	0.957
	close	0.983	0.998	0.991	0.990	0.939
	damaged	0.836	0.897	0.923	0.865	0.795

YOLO-GS	all	0.967	0.964	0.977	0.965	0.903
	open	0.998	0.992	0.995	0.995	0.959
	close	0.990	0.995	0.993	0.992	0.942
	damaged	0.905	0.905	0.943	0.905	0.808

### 3.3. Model Performance and Test Results Analysis

In the YOLO series of object detection tasks, the loss function is a key metric that measures the discrepancy between the model's predictions and the ground truth values. This loss is usually classified into two categories: training loss and validation loss. Training loss indicates how well the model fits the training dataset, while validation loss assesses the model's performance on the validation set. When both types of loss stabilize at low levels with minimal fluctuations and cease to decrease significantly, it indicates that the model has essentially converged and achieved optimal performance. **Error! Reference source not found.** and **Error! Reference source not found.** show the loss curves for the YOLOv11n and the proposed YOLO-GS models.

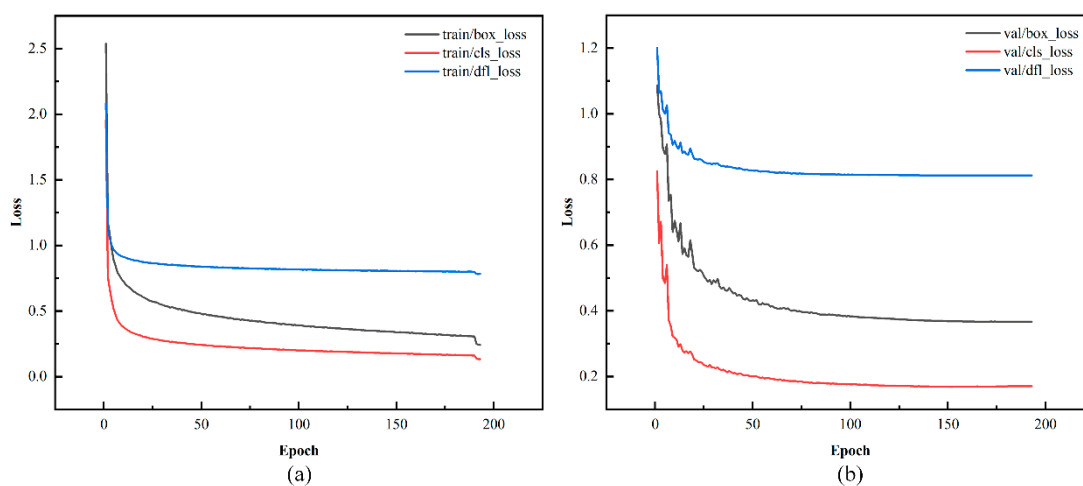


Figure 10. Loss curve of YOLOv11n (a) train loss (b) val loss.

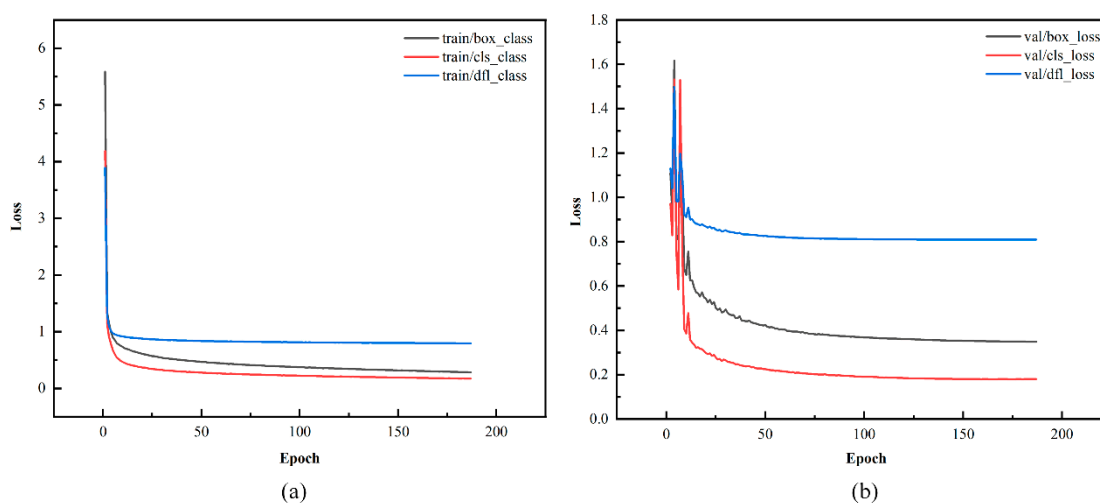


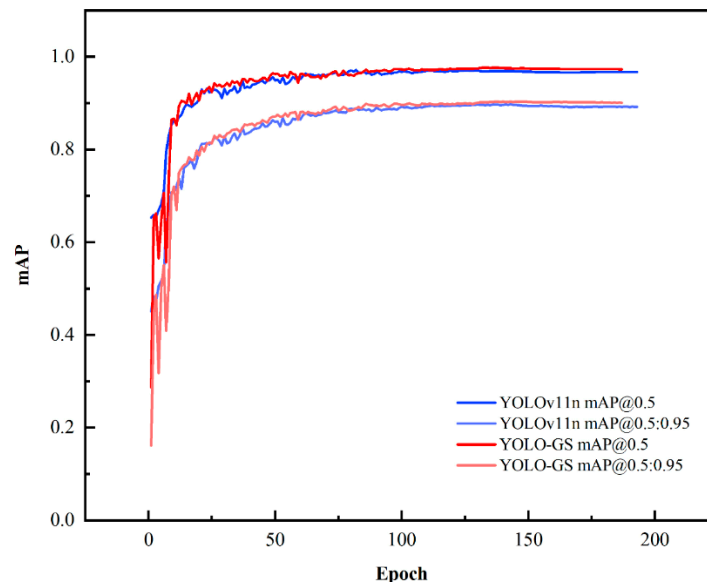
Figure 11. YOLO-GS Loss curve (a) train loss (b) val loss.

The training and validation losses for the YOLOv11n and YOLO-GS models decreased rapidly during the first 20 training rounds. After 50 rounds, the losses begin to stabilize, showing reduced fluctuations. Notably, the loss curve for the YOLO-GS model converges to a lower level more rapidly,

demonstrating less fluctuation and a smoother, more stable trend. This indicates that the YOLO-GS model accelerates the convergence process and effectively reduces differences in model losses.

mAP@0.5 measures the model's performance at a single Intersection over Union (IoU) level, evaluating its accuracy in object recognition. In contrast, mAP@0.5:0.95 encompasses multiple IoU thresholds, providing a more thorough assessment of the model's performance across various requirements for positioning accuracy. This approach more effectively reflects the model's robustness and generalization capabilities. Together, these metrics comprehensively evaluate the model's overall performance in target recognition and precise positioning.

**Error! Reference source not found.** shows the mAP@0.5 and mAP@0.5:0.95 curves for the YOLOv11n and YOLO-GS models. Both models exhibit a rapid increase in mAP values during the first 20 training rounds, followed by stabilization after 50 rounds. According to the graph, after 100 rounds of training, the YOLO-GS model had fewer curve fluctuations, and its mAP@0.5 and mAP@0.5:0.95 values were higher than those of the benchmark model, YOLOv11n. The graph indicates that after 100 rounds of training, the YOLO-GS model exhibits fewer fluctuations in its curves, and its mAP@0.5 and mAP@0.5:0.95 values are higher than those of the YOLOv11n model. Based on this analysis, we conclude that the YOLO-GS model accelerates convergence and effectively improves training stability and generalization ability.



**Figure 12.** mAP@0.5 and mAP@0.5:0.95 value variation curves of the YOLOv11n and YOLO-GS models.

### 3.4. Ablation Test

To validate the effectiveness of the YOLO-GS model proposed in this study, ablation tests were conducted by progressively integrating each component into the YOLOv11n baseline model. These tests utilized the same dataset, hyperparameters, and environmental settings for consistency. The components evaluated included the SELP module, the CPCA attention mechanism, the C3k2\_AssemFormer module, and the C2PSA\_CloFormerAttnConv structure. The ablation results are presented in Table 3. After incorporating the improved SELP module into the neck network, Recall decreases by 0.3%. However, there are increases in Precision, mAP@0.5, the F1-Score, and the mAP@0.5:0.95 value by 2%, 0.7%, 0.8%, and 0.3%, respectively. These results suggest that the enhanced SELP module improves channel attention and spatial adaptability, thereby enhancing feature representation and discrimination capabilities, and ultimately boosting the model's overall performance.

The CPCA improves Recall by 1.8%, although it decreases performance in other metrics compared to the baseline model. In contrast, the C3k2\_AssemFormer module increases Precision and

mAP@0.5:0.95 by 0.4% and 0.1%, respectively. This suggests that this module has enhanced the model's ability to identify and localize objects to some extent accurately. While there are slight declines in other metrics, the model's overall detection performance remains superior to that of the baseline model. These results indicate that the proposed improvement strategies have a positive impact on the benchmark model's performance across various metrics. Although the degree of improvement varies, these strategies enhance the model's overall performance. Compared to the YOLOv11n benchmark model, the YOLO-GS model achieves significant performance gains, with Precision, Recall, mAP@0.5, F1-Score, and mAP@0.5:0.95 increasing by 2.9%, 0.2%, 0.7%, 1.5%, and 0.5%, respectively. These results demonstrate that the YOLO-GS model effectively enhances detection accuracy and stability.

**Table 3.** Ablation Tests.

Baseline	SELP	CPCA	C3k2_ AssemFormer	C2PSA_CloForm erAttnConv	P	R	mAP@0.5	F1-Score	mAP@0.5: 0.95
√					0.938	0.962	0.97	0.950	0.897
√	√				0.958	0.959	0.977	0.958	0.9
√		√			0.936	0.980	0.964	0.957	0.886
√			√		0.942	0.959	0.968	0.950	0.898
√	√		√		0.940	0.983	0.961	0.961	0.894
√	√	√			0.924	0.979	0.957	0.951	0.885
√		√	√		0.926	0.949	0.951	0.937	0.881
√	√	√	√		0.938	0.963	0.975	0.950	0.905
√	√	√	√	√	0.967	0.964	0.977	0.965	0.903

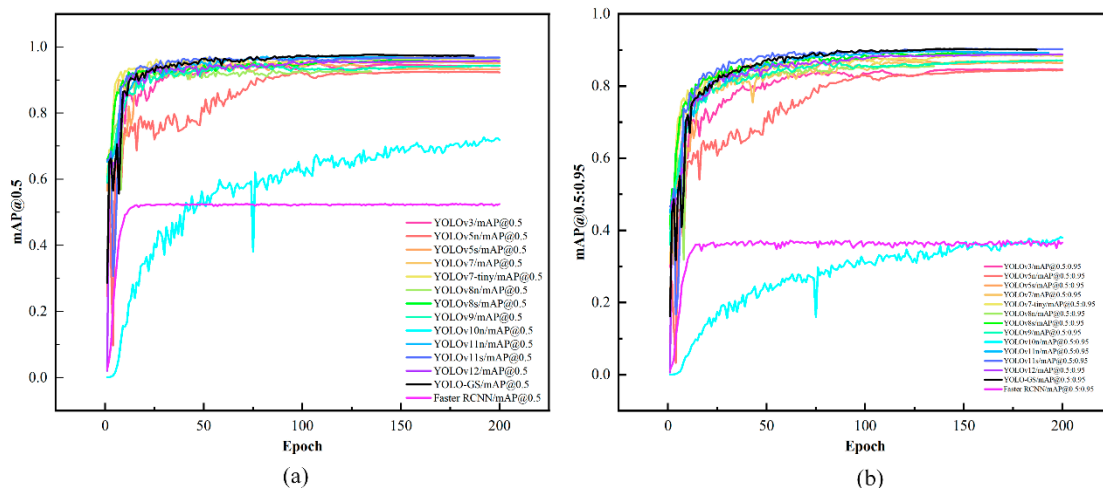
### 3.5. Performance Comparison with Other Models

To further evaluate the performance of the YOLO-GS model, this study conducted comparative trials with various existing object detection models, including one-stage and two-stage algorithms. All trials were performed using the same dataset, test environment, and hyperparameter settings.

Five mainstream performance metrics were selected for this experiment: mAP@0.5, mAP@0.5:0.95, Precision, Recall, and F1-Score. These metrics comprehensively measure each model's performance in terms of detection accuracy, false detection control, missed detection ability, and overall robustness. The results of the comparative tests are presented in **Error! Reference source not found.** and **Error! Reference source not found.**

**Table 4.** Comparative Trials.

Model	Precision	Recall	mAP@0.5	F1-Score	mAP@0.5:0.95
YOLOv3	0.922	0.948	0.943	0.934	0.846
YOLOv5n	0.893	0.941	0.925	0.916	0.844
YOLOv5s	0.953	0.914	0.951	0.933	0.879
YOLOv7	0.908	0.976	0.951	0.941	0.870
YOLOv7-tiny	0.944	0.973	0.961	0.958	0.884
YOLOv8n	0.930	0.919	0.931	0.924	0.858
YOLOv8s	0.937	0.954	0.959	0.945	0.892
YOLOv9	0.928	0.952	0.943	0.940	0.871
YOLOv10n	0.924	0.960	0.964	0.942	0.878
YOLOv11n	0.938	0.962	0.970	0.950	0.897
YOLOv11s	0.944	0.946	0.967	0.945	0.903
YOLOv12	0.941	0.967	0.957	0.954	0.888
Faster RCNN	-	0.412	0.534	-	0.367
YOLO-GS	0.967	0.964	0.977	0.965	0.903



**Figure 13.** Comparison of model checking performance (a) mAP@0.5 (b) mAP@0.5:0.95.

YOLO-GS demonstrates significant advantages across multiple metrics. It achieves the highest value of 97.7% in the mAP@0.5 metric, which is 3.4%, 5.2%, 2.6%, 2.6%, 1.6%, 4.6%, 1.8%, 3.4%, 1.3%, 0.7%, 1.0%, 2.0%, and 44.3% higher than YOLOv3, YOLOv5n, YOLOv5s, YOLOv7, YOLOv7-Tiny, YOLOv8n, YOLOv8s, YOLOv9, YOLOv10n, YOLOv11n, YOLOv11s, YOLOv12, and Faster RCNN models, respectively. This fully demonstrates the excellent detection accuracy of this model.

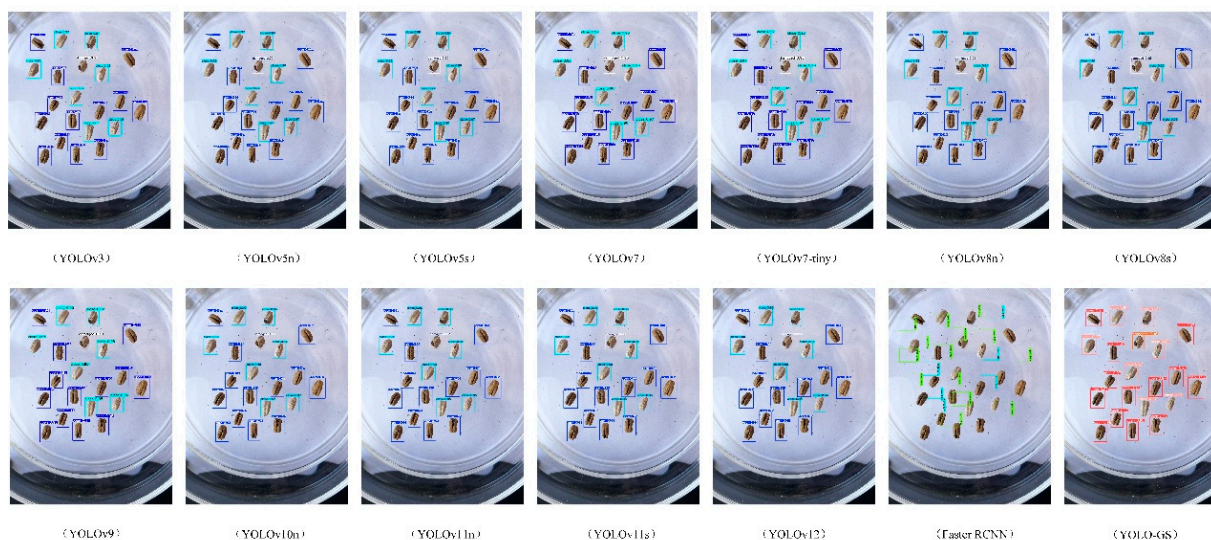
In the more comprehensive metric of mAP@0.5:0.95, compared with other benchmark models, YOLO-GS and YOLOv11s tie for the highest value at 90.3%, demonstrating the model's excellent stability and generalization ability under different IoU threshold conditions.

Regarding Precision, YOLO-GS outperforms all the models used for comparison, indicating a superior ability to suppress false positives. YOLO-GS's Recall is slightly lower than YOLOv7's and YOLOv7-tiny's (1.2% and 0.9%, respectively). It still outperforms all other models, demonstrating its strong target coverage capability.

YOLO-GS achieves the highest value in the F1-Score metric, which comprehensively evaluates Precision and Recall, outperforming all other models. These metrics demonstrate that YOLO-GS performs better in object detection tasks, balancing detection accuracy, recognition capability, and model robustness. It effectively meets the detection requirements for ginseng seeds.

### 3.6. Comparison of detection results with other models

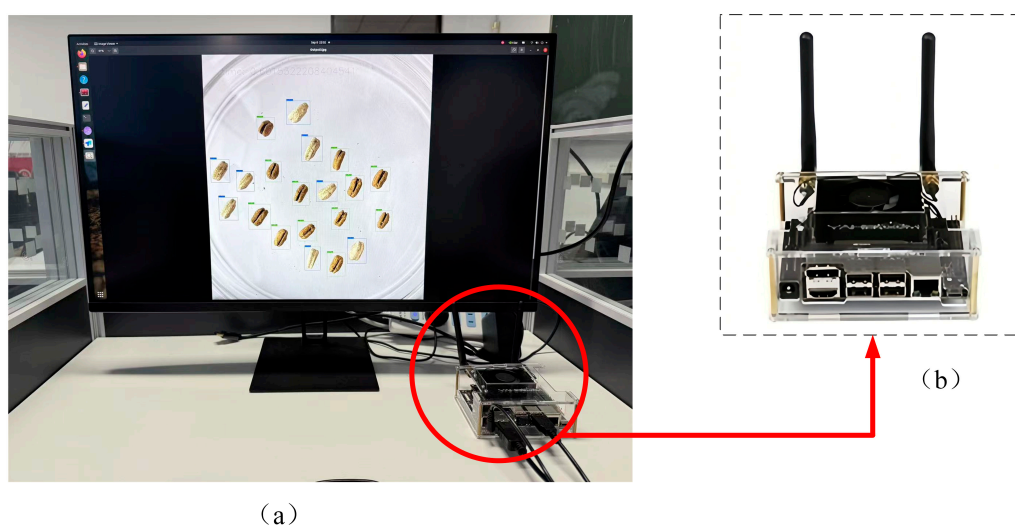
The performance of the YOLO-GS model in object detection is visually illustrated in Figure 14. This figure compares detection images from the 14 models listed in **Error! Reference source not found.** YOLO-GS accurately identifies ginseng seeds in various states and maintains stable detection performance under dense seed clusters and blurred textures. It significantly outperforms other models in detecting damaged seeds and minor defects, delivering superior overall detection results.



**Figure 14.** Detection results of different models under different lighting conditions.

### 3.7. Performance Evaluation of Model Deployment

To validate the real-time performance and deployability of the proposed YOLO-GS model on edge devices, this study deployed the model on the NVIDIA Jetson Orin Nano embedded platform (NVIDIA Corporation, Santa Clara, CA, USA). The experimental setup is illustrated in Figure 15, and the results of the inference performance are displayed in Figure 16. The primary software configuration for the deployment experiment was Python 3.8 and PyTorch 1.8. The hardware platform consisted of an Arm Cortex-A78AE central processing unit (CPU) and a 32-core tensor graphics processing unit (GPU), simulating real-world operating conditions in resource-constrained scenarios. The YOLO-GS model maintains the original detection accuracy, with a parameter size of approximately 4.2 M and an inference time of around 0.6 ms. This indicates that the model is suitable for deployment in real edge terminal detection systems, providing a viable technical reference for human seed quality identification and automatic sorting.



**Figure 15.** Inference Results of Different Models on Embedded Platforms (a) Deployment of the integrated embedded platform (b) NVIDIA Jetson Orin Nano.



**Figure 16.** The reasoning results of different models (a) Original image (b) Inference results based on YOLOv11n (c) Inference results based on YOLO-GS.

#### 4. Discussion

Ginseng seeds are challenging to sort efficiently and accurately before sowing due to their small size and subtle interspecific variations. Additionally, minor damage during germination can easily be mistaken for the seeds' natural texture. Traditional manual inspection methods, which rely on visual observation, struggle to maintain consistent accuracy in recognizing these seeds. To overcome these challenges, this study developed the YOLO-GS model, specifically designed for scenarios involving fine-grained, small objects. This model enhances the network's ability to represent subtle textures and edge information, achieving high-precision and non-destructive detection of ginseng seed quality.

Currently, deep learning-based object detection models demonstrate significant potential for accurately recognizing crop appearance and defects. Jie Li et al. [31] developed the multi-task detection model YOLO-MTP based on YOLOv8s. This model retains the object detection branch and introduces a semantic segmentation branch, enhancing the feature extraction capabilities of its backbone network through dynamic convolutions and multi-scale attention mechanisms. By combining uncertainty weighting with gradient conflict optimization strategies, this model achieved a mAP@0.5 of 96.7% and a Recall of 92.3% in potato quality inspection tasks, outperforming multiple state-of-the-art models overall. However, potatoes have relatively large target dimensions and distinct defect characteristics. In contrast, the YOLO-GS model proposed in this study maintains high detection performance even when applied to ginseng seeds, which are smaller in size and exhibit more subtle defects.

The Jujube-YOLO model, proposed by Lijun Wang et al. [32], focused on the challenging task of identifying cracks in jujube fruits. This model incorporates a DCSE module and an RCM structure with C3k2 into its backbone network. Additionally, it uses a multi-branch channel attention mechanism in the neck region to enhance feature extraction capabilities for detecting subtle cracks. As a result, the model achieves stable recognition even in complex natural environments, achieving

Precision, Recall, mAP@0.5, and F1-Score of 98.37%, 94.96%, 97.65%, and 96.63%, respectively. Compared to defects like cracks in jujube fruit, which exhibit distinct structural features, ginseng seeds display subtler appearance variations with non-structural defect morphologies. The YOLO-GS model effectively identifies minute differences in surface texture. Yuzhuo Zhang et al. [33] proposed a detection method based on an optimized S2ANet combined with edge detection thresholds. Through uniform preprocessing of seed orientation, the model's average accuracy improved to 95.6%. It achieved accuracy of 95.08% for crack recognition and 95.75% for non-crack recognition. When tested on mixed samples, the model attained an Average Precision of 95.91% and a Recall rate of 94.8%. In contrast, ginseng seeds present smaller detection targets with more minute defects. The YOLO-GS model demonstrated robust performance, achieving 97.7% for the mAP@0.5 and 90.3% for the mAP@0.5:0.95, all while maintaining a lightweight design. This research has achieved precise identification of different quality states of ginseng seeds. The model has a small number of parameters and fast inference speed, meeting the real-time detection requirements on embedded platforms.

This study has certain limitations. The experimental data primarily consist of unobstructed seed images that are positioned in a single orientation. As a result, the method may not perform well with seeds that are in random orientations, have significant adhesion, or contain foreign objects, which is common in actual production scenarios. To improve robustness and generalization, future work will involve expanding the dataset by collecting images from multiple angles, under various lighting conditions, and against diverse backgrounds using mechanized equipment. Structural optimization and knowledge distillation techniques will further compress the model to meet the requirements of embedded device deployment, enabling lower-cost, real-time edge detection. Ultimately, this method can integrate with automated sorting systems to enable online seed grading and screening, advancing ginseng cultivation toward intelligent and precision management.

## 5. Conclusions

This study created a dataset of ginseng seeds after they had germinated. A significant amount of experimental data was collected and analyzed to determine the characteristics of different seed categories post-germination. The data were annotated, and various data augmentation techniques were applied. The dataset was divided into training, testing, and validation sets in a ratio of 8:1:1. This dataset enables various deep learning algorithms to effectively learn the features of ginseng seeds.

To address quality variation issues in post-germination ginseng seeds, this study proposes YOLO-GS, an improved object detection model based on YOLOv11n. The backbone network incorporates a Channel Prior Convolutional Attention mechanism to improve its ability to extract features. The C3k2\_AssemFormer module enhances target detection accuracy, while the C2PSA\_CloFormerAttnConv module improves context awareness and model robustness. Additionally, the SELP module in the neck network enhances key feature extraction, significantly improving object detection performance. Experimental results demonstrate that the YOLO-GS model outperforms other models, achieving 97.7% mAP@0.5 and 90.3% mAP@0.5:0.95. These results validate the model's effectiveness in terms of detection accuracy and generalization performance. This model boasts lightweight characteristics, including a low parameter quantity and fast inference speed, which enable efficient deployment on edge devices to meet the stringent requirements of real-time performance and low resource consumption in practical applications.

**Author Contributions:** Conceptualization, L.H. and N.W.; methodology, L.H. and X.Y.; software, N.W. and M.W.; validation, M.W.; formal analysis, N.W. and W.T.; investigation, N.W. and X.Y.; resources, X.Y. and H.G.; data curation, M.W. and W.T.; writing—original draft preparation, L.H. and N.W.; writing—review and editing, J.Z. and H.G.; visualization, W.T.; supervision, J.Z. and H.G.; project administration, H.G.; funding acquisition, L.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by Jilin Province Science and Technology Development Plan Project (No.20230101343JC).

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** The data that support the findings of this study are available in the main text and the supplementary materials of this article. Additional data are available from the corresponding author upon request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

CBAM	Convolutional Block Attention Module
SE	Squeeze-and-Excitation
PConv	Partial Convolution
CPCA	Channel Prior Convolutional Attention
CA	Channel Attention
SA	Spatial Attention
FN	False Negative
FP	False Positive
IoU	Intersection over Union
TP	Ture Positive
P	Precision
R	Recall
AP	Average Precision
mAP	mean Average Precision
YOLO	You Only Look Once

## References

1. Yoon, S.J.; Sukweenadhi, J.; Khorolragchaa, A.; Mathiyalagan, R.; Subramaniyam, S.; Kim, Y.J.; Kim, H.B.; Kim, M.J.; Kim, Y.J.; Yang, D.C. Overexpression of Panax ginseng sesquiterpene synthase gene confers tolerance against *Pseudomonas syringae* pv. tomato in *Arabidopsis thaliana*. *Physiol Mol Biol Plants* **2016**, *22*, 485-495. <https://doi.org/10.1007/s12298-016-0384-9>.
2. Hu, L.-J.; Zhang, Z.; Wang, W.; Wang, L.; Yang, H. Ginseng plantations threaten China's forests. *Biodivers. Conserv.* **2018**, *27*, 2093-2095. <https://doi.org/10.1007/s10531-018-1530-1>.
3. Zhang, K.; Zhang, S.; Ebihara, A.; Zhou, X.; Fan, L.; Li, P.; Zhang, Z.; Wang, Y. The current research progress of ginseng species: The cultivation and application. *Cogent Food Agric.* **2023**, *9*. <https://doi.org/10.1080/23311932.2023.2216483>.
4. Kurtulmuş, F.; Kavdir, İ. Detecting corn tassels using computer vision and support vector machines. *Expert Syst. Appl.* **2014**, *41*, 7390-7397. <https://doi.org/10.1016/j.eswa.2014.06.013>.
5. Wu, A.; Zhu, J.; Yang, Y.; Liu, X.; Wang, X.; Wang, L.; Zhang, H.; Chen, J. Classification of corn kernels grades using image analysis and support vector machine. *Adv. Mech. Eng.* **2018**, *10*. <https://doi.org/10.1177/1687814018817642>.
6. Tu, K.; Wu, W.; Cheng, Y.; Zhang, H.; Xu, Y.; Dong, X.; Wang, M.; Sun, Q.J.C.; Agriculture, E.i. AIsed: An automated image analysis software for high-throughput phenotyping and quality non-destructive testing of individual plant seeds. **2023**, *207*, 107740.
7. Fan, Y.; An, T.; Yao, X.; Long, Y.; Wang, Q.; Wang, Z.; Tian, X.; Chen, L.; Huang, W. An interpretable nondestructive detection model for maize seed viability: Based on grouped hyperspectral image fusion and key biochemical indicators. *Comput. Electron. Agric.* **2025**, *239*, 111036. <https://doi.org/10.1016/j.compag.2025.111036>.

8. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137-1149. <https://doi.org/10.1109/TPAMI.2016.2577031>.
9. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *2017 IEEE Int. Conf. Comput. Vis. (ICCV)* **2017**, 2980-2988. <https://doi.org/10.1109/ICCV.2017.322>.
10. Wu, W.; Liu, T.; Zhou, P.; Yang, T.; Li, C.; Zhong, X.; Sun, C.; Liu, S.; Guo, W. Image analysis-based recognition and quantification of grain number per panicle in rice. *Plant Methods* **2019**, *15*, 122. <https://doi.org/10.1186/s13007-019-0510-0>.
11. Li, Z.; Li, Y.; Yang, Y.; Guo, R.; Yang, J.; Yue, J.; Wang, Y. A high-precision detection method of hydroponic lettuce seedlings status based on improved Faster RCNN. *Comput. Electron. Agric.* **2021**, *182*, 106054. <https://doi.org/10.1016/j.compag.2021.106054>.
12. Deng, R.; Tao, M.; Huang, X.; Bangura, K.; Jiang, Q.; Jiang, Y.; Qi, L. Automated Counting Grains on the Rice Panicle Based on Deep Learning Method. *Sensors* **2021**, *21*, 281. <https://doi.org/10.3390/s21010281>.
13. Tan, S.; Lu, H.; Yu, J.; Lan, M.; Hu, X.; Zheng, H.; Peng, Y.; Wang, Y.; Li, Z.; Qi, L.; et al. In-field rice panicles detection and growth stages recognition based on RiceRes2Net. *Comput. Electron. Agric.* **2023**, *206*, 107704. <https://doi.org/10.1016/j.compag.2023.107704>.
14. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. *Computer Vision – ECCV 2016* **2016**, 21-37. [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2).
15. Lu, A.; Guo, R.; Ma, Q.; Ma, L.; Cao, Y.; Liu, J. Online sorting of drilled lotus seeds using deep learning. *Biosyst. Eng.* **2022**, *221*, 118-137. <https://doi.org/10.1016/j.biosystemseng.2022.06.015>.
16. Wang, Q.; Yu, C.; Zhang, H.; Chen, Y.; Liu, C. Design and experiment of online cottonseed quality sorting device. *Comput. Electron. Agric.* **2023**, *210*, 107870. <https://doi.org/10.1016/j.compag.2023.107870>.
17. Jiao, Y.; Wang, Z.; Shang, Y.; Li, R.; Hua, Z.; Song, H. Detecting endosperm cracks in soaked maize using  $\mu$ CT technology and R-YOLOv7-tiny. *Comput. Electron. Agric.* **2023**, *213*, 108232. <https://doi.org/10.1016/j.compag.2023.108232>.
18. Li, Q.; Ma, W.; Li, H.; Zhang, X.; Zhang, R.; Zhou, W. Cotton-YOLO: Improved YOLOV7 for rapid detection of foreign fibers in seed cotton. *Comput. Electron. Agric.* **2024**, *219*, 108752. <https://doi.org/10.1016/j.compag.2024.108752>.
19. Niu, S.; Xu, X.; Liang, A.; Yun, Y.; Li, L.; Hao, F.; Bai, J.; Ma, D. Research on a Lightweight Method for Maize Seed Quality Detection Based on Improved YOLOv8. *IEEE Access* **2024**, *12*, 32927-32937. <https://doi.org/10.1109/ACCESS.2024.3365559>.
20. Yi, W.; Xia, S.; Kuzmin, S.; Gerasimov, I.; Cheng, X. RTFVE-YOLOv9: Real-time fruit volume estimation model integrating YOLOv9 and binocular stereo vision. *Comput. Electron. Agric.* **2025**, *236*, 110401. <https://doi.org/10.1016/j.compag.2025.110401>.
21. Khanam, R.; Hussain, M. YOLOv11: An Overview of the Key Architectural Enhancements. *Arxiv Prepr.* **2024**, *abs/2410.17725*. <https://doi.org/10.48550/arXiv.2410.17725>.
22. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. *2018 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* **2018**, 7132-7141. <https://doi.org/10.1109/CVPR.2018.00745>.
23. Liu, G.; Shih, K.J.; Wang, T.-C.; Reda, F.A.; Sapra, K.; Yu, Z.; Tao, A.; Catanzaro, B. Partial Convolution based Padding. *Arxiv Prepr.* **2018**, *abs/1811.11718*. <https://doi.org/10.48550/arXiv.1811.11718>.
24. Clevert, D.-A.; Unterthiner, T.; Hochreiter, S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *Arxiv Prepr.* **2015**, *abs/1511.07289*. <https://doi.org/10.48550/arXiv.1511.07289>.
25. Huang, H.; Chen, Z.; Zou, Y.; Lu, M.; Chen, C.; Song, Y.; Zhang, H.; Yan, F. Channel prior convolutional attention for medical image segmentation. *Comput. Biol. Med.* **2024**, *178*, 108784. <https://doi.org/10.1016/j.compbiomed.2024.108784>.
26. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. *Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part VII* **2018**, 3–19. [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1).
27. Fan, Q.; Huang, H.; Guan, J.; He, R. Rethinking Local Perception in Lightweight Vision Transformer. *Arxiv Prepr.* **2023**, *abs/2303.17803*. <https://doi.org/10.48550/arXiv.2303.17803>.

28. Vaswani, A.; Shazeer, N.M.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. *Arxiv Prepr.* **2017**, *abs/1706.03762*. <https://doi.org/10.48550/arXiv.1706.03762>.
29. Yi, F.; Wen, H.; Jiang, T. ASFormer: Transformer for Action Segmentation. *Arxiv Prepr.* **2021**, *abs/2110.08568*. <https://doi.org/10.48550/arXiv.2110.08568>.
30. Wang, S.; Li, B.Z.; Khabsa, M.; Fang, H.; Ma, H. Linformer: Self-Attention with Linear Complexity. *Arxiv Prepr.* **2020**, *abs/2006.04768*. <https://doi.org/10.48550/arXiv.2006.04768>.
31. Li, J.; Gong, G.; Zhang, J.; Liu, X.; Jiang, S.; Wu, X.; Xie, H.; Gu, L.J.J.o.F.C.; Analysis. Multi-task learning based approach for potato edibility detection and defect segmentation. **2025**, 108420.
32. Wang, L.; Wang, S.; Wang, B.; Yang, Z.; Zhang, Y.J.E.S.w.A. Jujube-YOLO: a precise jujube fruit recognition model in unstructured environments. **2025**, 128530.
33. Zhang, Y.; Lv, C.; Wang, D.; Mao, W.; Li, J.J.C.; Agriculture, E.i. A novel image detection method for internal cracks in corn seeds in an industrial inspection line. **2022**, *197*, 106930.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.