agriculture

MDPI

Article

Evaluating the Seedling Emergence Quality of Peanut Seedlings via UAV Imagery

Guanchu Zhang [1], Qi Wang [1], Guowei Li [2], Dunwei Ci [1], Chen Zhang [3] and Fangyan Ma [1,*]

[1] Shandong Peanut Research Institute, Qingdao 266100, China
[2] Institute of Crop Germplasm Resources, Shandong Academy of Agricultural Sciences, Jinan 250100, China
[3] College of Advanced Agricultural Sciences, Zhejiang A&F University, Hangzhou 311300, China
* Correspondence: mafangyan1998@163.com

Abstract

Accurate evaluation of peanut seedling emergence is critical for ensuring agronomic research accuracy and planting benefit efficiency, but traditional manual methods are limited by strong subjectivity and inconsistent batch inspection standards. In order to quickly and accurately evaluate the emergence rate and quality of peanuts, this study proposes an intelligent evaluation system for peanut seedling conditions, which is constructed based on an improved YOLOv11 combined with the Segment Anything Model (SAM) for peanut seedling emergence evaluation, using high-resolution images collected by Unmanned Aerial Vehicles as the data foundation. Experimental results show that the improved YOLOv11 model achieves a detection precision of 96.36%, a recall rate of 96.76%, and an mAP@0.5 of 99.03%. The segmentation performance of SAM is outstanding in terms of integrity. In practical applications, the detection time for a single image by the system is as low as 83.4 ms, and the efficiency of video counting is 6–10 times higher than that of manual counting. Without extensive data annotation, this method performs excellently in peanut seedling emergence quantity statistics and growth status classification, providing efficient, accurate technical support for refined peanut cultivation management and mechanical sowing quality evaluation.

Keywords: peanut seedlings; Unmanned Aerial Vehicle (UAV); seedling emergence status; image recognition

check for updates

Academic Editor: Maciej Zaborowicz

Received: 24 September 2025
Revised: 15 October 2025
Accepted: 16 October 2025
Published: 17 October 2025

Citation: Zhang, G.; Wang, Q.; Li, G.; Ci, D.; Zhang, C.; Ma, F. Evaluating the Seedling Emergence Quality of Peanut Seedlings via UAV Imagery. *Agriculture* **2025**, *15*, 2159. https://doi.org/10.3390/agriculture15202159

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

As an important global oil and cash crop [1,2], the germination rate and germination potential of peanut seeds are core indicators for measuring seed quality, and the quality of their seedling emergence directly affects yield stability and planting efficiency [3]. In large-scale breeding scenarios, traditional statistics on peanut seedling emergence mainly rely on manual on-site sampling and measurement. This method has significant drawbacks such as time lag, low work efficiency, and insufficient measurement accuracy. Moreover, subjective differences among different observers often lead to a lack of consistency in monitoring results [4].

In recent years, Unmanned Aerial Vehicle (UAV) remote sensing technology has provided strong data support for agricultural phenotyping monitoring by virtue of its advantages of high-resolution imaging and large-area coverage [5–9]. In the field of breeding, a notable advantage of crop monitoring using image data is that algorithms are relatively easy to implement, and the monitoring method based on video data is more in

line with actual field application scenarios [10–12]. However, the leaves of seedlings at the seedling stage are small in morphology [13]. When traditional image analysis methods handle complex scenarios of peanut seedlings, they not only require large datasets with detailed manual annotations but also suffer from limitations such as insufficient multi-scale feature extraction, weak anti-interference ability, and low computational efficiency, resulting in large errors in seedling emergence rate statistics. In addition, there are only a few published studies on real-time monitoring of peanut seedlings: Lin et al. [14] proposed a real-time peanut video counting model to efficiently and quickly measure the number of peanut seedlings, which realizes real-time counting based on video streams on UAV platforms by combining improved YOLO-V5s object detection and DeepSort object tracking algorithms. Zhang et al. [15] proposed a peanut seedling recognition model MS-YOLOv8 to address the problem of low recognition accuracy caused by uneven sizes of crop seedlings and changes in growth conditions under salt-alkali stress, using close-range remote sensing of UAVs to quickly identify and count peanut seedlings. The above methods focus on the accuracy of peanut seedling quantity but ignore the evaluation of seedling quality, and they require data collection at a fixed height, which cannot fully meet the actual needs of researchers. In recent years, semi-supervised or unsupervised learning has gradually been applied in agriculture [16–19]. Huang et al. [20] proposed a new method for estimating the size of strawberry canopies using a machine learning approach combining YOLOv11 and SAM. This method does not require extensive annotated datasets needed by traditional and supervised segmentation methods, making it a convenient and efficient way to perform segmentation [21].

This study in terms of innovation in feature fusion module, a Selective Boundary Aggregation (SBA) module is designed to redesign the Neck network of YOLOv11n, which enhances the model's ability to capture multi-scale features of peanut seedlings, thereby enabling accurate detection of seedlings at different UAV flight heights; second, regarding the integration of detection and segmentation for quality evaluation, the Segment Anything Model (SAM) is introduced for the first time to perform segmentation evaluation of peanut plants within the detection boxes output by YOLOv11n, and on this basis, realizing the transition from "single quantity statistics" to "quantity-quality integrated evaluation"; third, in terms of reducing dataset dependence, by combining SAM and the improved YOLOv11n, the model only needs a simply constructed dataset to achieve high-precision monitoring, which reduces the cost of data preparation; finally, for the verification of practical application value, through comparative experiments with mainstream object detection algorithms under different flight heights and field scenarios, the effectiveness of the improved algorithm in both quantity counting and quality evaluation is verified, and this model further provides efficient and objective technical support for refined management of peanut cultivation, quality evaluation of mechanical sowing, and breeding research.

2. Materials and Methods

2.1. Data Collection

The peanut seedling images and videos used in this study were collected from experimental fields at the Laixi Experimental Station of Shandong Peanut Research Institute in Laixi City, Qingdao, Shandong Province, China, from May to June 2025, representing the peanut varieties Huayu 9117, Huayu 9520, and Huayu 9139, respectively, the experimental field situation is shown in Figure 1. All images were collected during two time periods: 8:00 AM to 10:00 AM and 2:00 PM to 4:00 PM, in clear, windless, or slightly windy weather conditions without clouds.

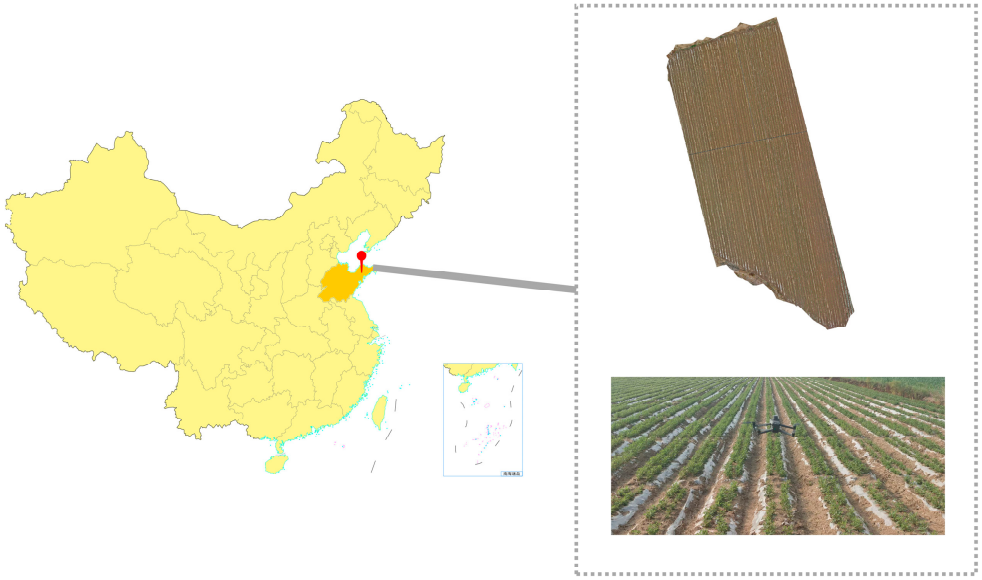


Figure 1. Overview of the peanut experimental field.

RGB images and videos were acquired using the DJI Mavic 3 Multispectral Edition (SZ DJI Technology Co., Shenzhen, China). The UAV image collection process included regional positioning, flight path planning, image capture, and post-processing. The UAV flight height was set to 1–12 m. For image collection, a planar route planning mode was adopted, with images focusing on vertical overhead shooting to ensure data consistency and high resolution. The camera parameters were set as follows: 20 million pixels in resolution, the original image resolution obtained is 5280 × 3956, ISO-100, aperture: f/2.8, exposure time 1/1000 s. During flight, the lateral and longitudinal overlap rates of adjacent images were both 90% to ensure the accuracy of image stitching and feature recognition. Video data were collected by manual control of the UAV flying at a constant speed at different heights for recording. In addition, the UAV was equipped with an RTK positioning system, which can provide centimeter-level positioning accuracy to ensure flight stability and maintain consistent image angles during collection. This minimizes the potential impact of motion blur or angular deviation on image quality and accuracy. Finally, the seedling image samples obtained through this scheme are shown in Figure 2.

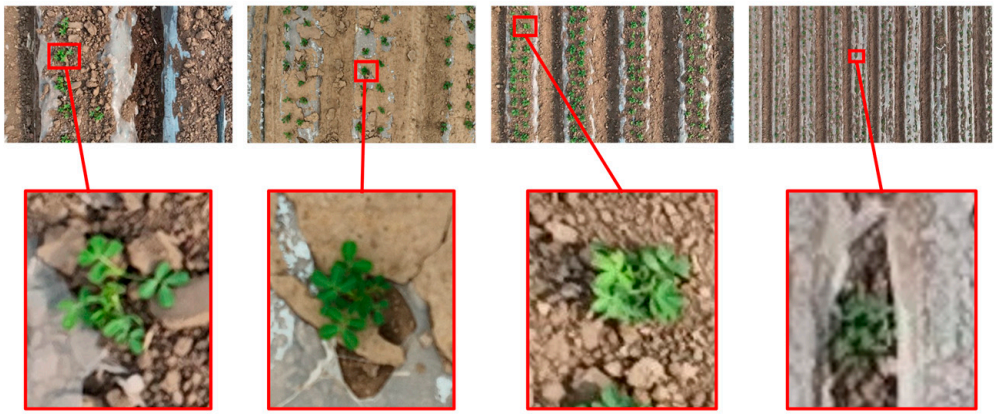


Figure 2. Seedling images acquired at different heights.

2.2. Dataset Construction

The captured peanut images were screened to remove blurred images, retaining a total of 943 images as the original dataset. All peanut images were semi-automatically annotated using T-rexlabel online annotation tool, with the annotation range being the

minimum enclosing rectangle of peanut seedlings. After annotation, TXT files containing peanut seedling labels and rectangle frame position information were automatically generated. Eventually, 943 labeled images and 943 TXT data files were obtained, containing 48,662 labels, the label distribution results are shown in Figure 3.

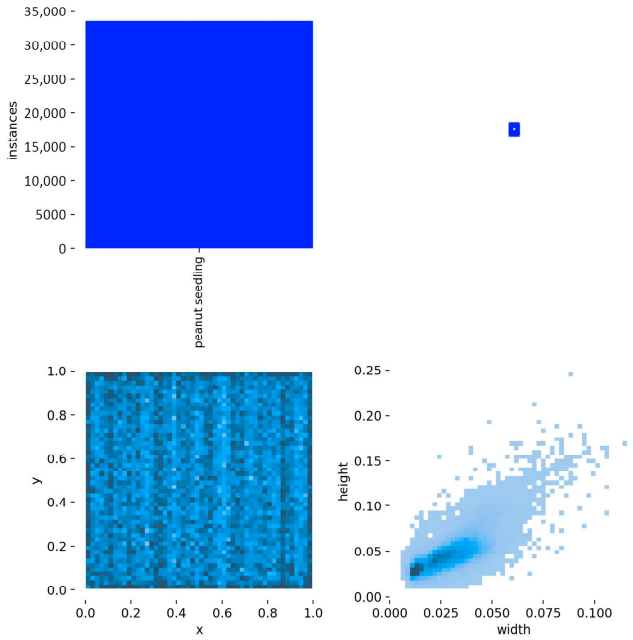


Figure 3. Statistics of annotated objects and anchor boxes.

The annotated dataset was divided into a training set of 660 images, a test set of 189 images, and a validation set of 94 images. To increase data diversity and generalization ability, online data augmentation methods such as hue perturbation, translation augmentation, scaling augmentation, horizontal flipping, and mosaic augmentation were applied to the training set for expansion.

2.3. Software and Hardware Configuration

For training, the AutoDL computing power cloud service platform was used, with the server configured as follows: CPU: 14 vCPU Intel(R) Xeon(R) Gold 6330 CPU @ 2.00 GHz; GPU: NVIDIA RTX 3090 24 GB independent graphics card; memory: 64 GB. The deep learning framework adopted was PyTorch 2.5.1, with CUDA version 12.4. For running the detection system, the device specifications were: CPU: Intel(R) Core (TM) i5-14400F @ 2.50 GHz; GPU: NVIDIA RTX 4060 Ti 8 GB independent graphics card; memory: 2 TB; operating system: Windows 11 Pro.

2.4. Improved YOLOv11 Object Detection Model

In this study, plant detection results are used as prompts to generate initial masks. However, seedling-stage plant targets are small in size with indistinct features, often disturbed by shadows and adhesion, leading to poor accuracy and robustness of methods based on feature extraction or rule-based segmentation. This poses challenges to the model's ability to extract target features. YOLOv11 is a model in the Ultralytics YOLO series of real-time object detectors. Compared to previous versions, YOLOv11 has improved backbone and neck architectures, enhancing feature extraction capabilities to achieve more accurate object detection and handle more complex tasks. YOLOv11 introduces two new modules, C3k2 and C2PSA, and continues the NMS-free training strategy from YOLOv10, enabling end-to-end object detection and further improving performance and flexibility [22].

In this study, the YOLOv11n model was used as the basis, and its structure was designed and improved to adapt to peanut seedling recognition tasks.

In field environments, peanut seedlings present complex image backgrounds and a high proportion of small targets, leading to limitations in feature fusion of the YOLOv11 model. This easily causes detail loss and edge blurring, resulting in the loss of key information. These defects not only increase the number of model parameters but also weaken its ability to detect small targets and handle complex backgrounds. Therefore, a more efficient feature fusion mechanism is needed to overcome these shortcomings. The SBA module introduces a bidirectional fusion mechanism between high-resolution and low resolution features, which enables more complete information transfer between features and further enhances the effectiveness of multi-scale feature fusion. At the same time, through an adaptive attention mechanism, the weights of features are adaptively adjusted according to the different resolutions and contents of the feature maps, in order to better capture the multi-scale features of the target. In this study, the neck network of the model was redesigned by integrating the Selective Boundary Aggregation (SBA) module for multi-scale feature fusion, enhancing the model's ability to capture features at different resolutions and addressing issues of object boundary blurring in images, as well as redundancy and inconsistency in feature fusion. The module adopts an adaptive attention mechanism, adjusting feature weights according to feature resolution and content, thereby improving the model's ability to effectively capture multi-scale features of targets. The structure of the model is shown in Figure 4.

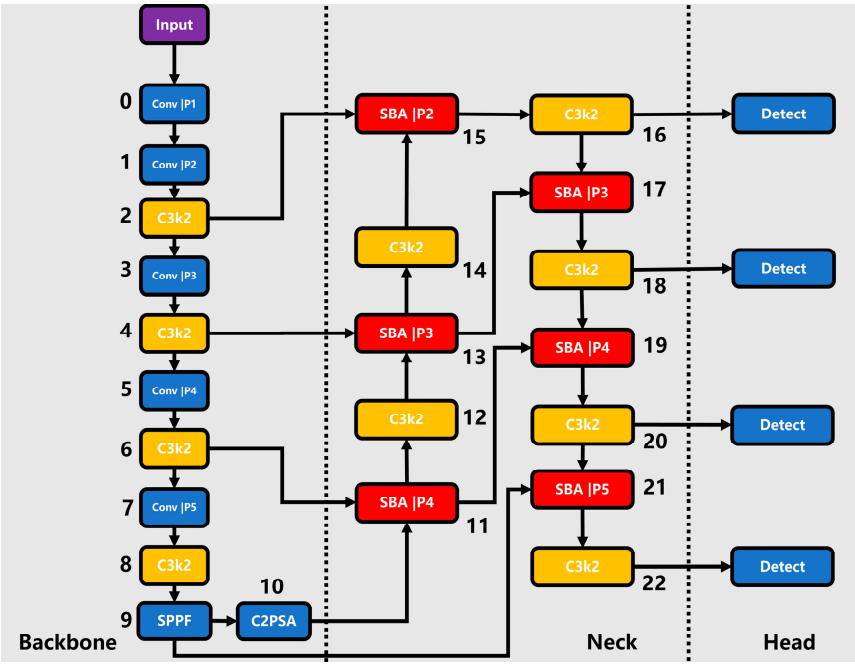


Figure 4. Structure diagram of the improved YOLOv11 model backbone network.

The SBA module selectively aggregates boundary information and semantic information to depict finer-grained object contours and recalibrate object positions. It is designed with a novel Re-Calibration Attention Unit (RAU) block, which adaptively extracts mutual representations from two inputs before fusion [23]. As shown in Figure 5, shallow and deep information is input into two RAU blocks through different methods to compensate for the lack of spatial boundary information in high-level semantic features and the lack of semantic information in low-level features.

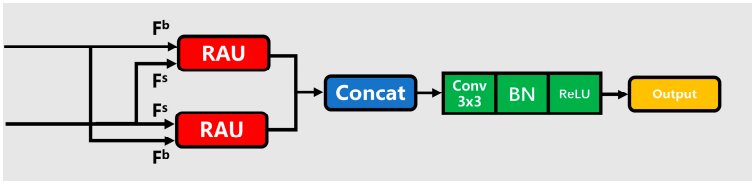


Figure 5. Schematic diagram of the SBA module structure.

Finally, the outputs of the two RAU blocks are concatenated after a 3 × 3 convolution. This aggregation strategy enables robust combination of different features and refines coarse features. The processing process of the RAU block function is shown in Equations (1) and (2), and the process is illustrated in Figure 6:

$$T_1' = W_\theta(T_1), T_2' = W_\phi(T_2) \tag{1}$$

$$P(T_1, T_2) = T_1' \odot T_1 + T_2' \odot T_2 \odot (\ominus(T_1')) + T_1 \tag{2}$$

where T_1 and T_2 are input features. Two linear mappings and sigmoid functions W_θ and W_ϕ are applied to the input features, reducing the channel dimension to 32 to obtain feature maps T_1' and T_2'. $\odot$ denotes element-wise multiplication, and $\ominus$ refines imprecise and rough estimates into accurate and complete prediction maps through an inverse operation of subtracting feature T_1', with the 1 × 1 convolution operation serving as the linear mapping process.

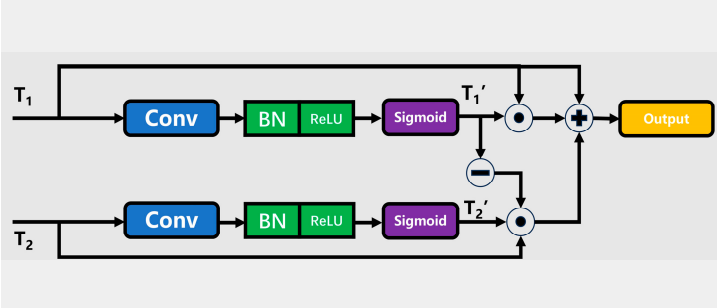


Figure 6. Schematic diagram of the RAU block structure.

Thus, the process of SBA is shown in Equation (3):

$$Z = C_{3\times3}(Concat(P(F^s, F^b), P(F^b, F^s))) \tag{3}$$

where $C_{3\times3}$ denotes a 3 × 3 convolution with batch normalization and a ReLU activation layer. $F^s \in \mathbb{R}^{\frac{H}{8}\times\frac{W}{8}\times32}$ contains deep semantic information after fusing the third and fourth layers of the encoder, while $F^b \in \mathbb{R}^{\frac{H}{4}\times\frac{W}{4}\times32}$ is the first layer from the backbone, which is rich in boundary details. *Concat* represents the concatenation operation along the channel dimension, and $Z \in \mathbb{R}^{\frac{H}{4}\times\frac{W}{4}\times32}$ is the output of the SBA module.

2.5. Peanut Seedling Segmentation Model

The Segment Anything Model (SAM), proposed by Meta, is a "segment everything" model [24] that achieves strong zero-shot generalization without extensive retraining. SAM can use points, bounding boxes, and masks as input prompts, ensuring seamless and efficient mask generation throughout the process. SAM consists of three components: an image encoder, a prompt encoder, and a fast mask decoder. The image encoder uses a Vision Transformer (ViT) with minimal adjustments for processing high-resolution inputs. The prompt encoder considers two sets of prompts: sparse prompts (points, boxes, text) and

dense prompts (masks). Points and boxes are represented by adding positional encodings to learned embeddings for each prompt type; free-form text uses an off-the-shelf CLIP text encoder. Dense prompts (i.e., masks) are embedded via convolution and added element-wise to the image embedding. The mask decoder efficiently maps image embeddings, prompt embeddings, and output tokens to masks. Specifically, the image encoder first encodes the input image to obtain the image feature representation (Image_embeddings). Then, the prompt encoder encodes user-provided prompts (such as points, boxes, masks) to generate corresponding Prompt embeddings. Next, the Mask decoder fuses Image_embeddings and Prompt embeddings: Image_embeddings are mixed with Dense embeddings, while Sparse embeddings are concatenated with Mask tokens and IoU tokens to form new tokens. These new tokens and Image_embeddings undergo a TwoWayTransformer module, updating information in images and tokens through multiple Self-attention and Cross-attention operations. Finally, the updated tokens are split into mask tokens and IoU tokens (where IoU tokens represent the confidence of each mask). Image information is restored to the original image size via transposed convolution and multiplied with mask tokens to generate final masks. The overall process is shown in Figure 7.

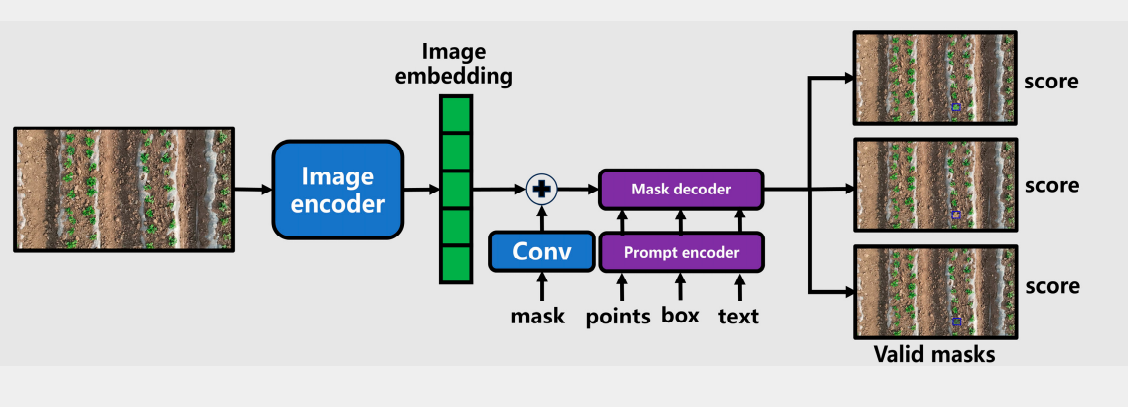


Figure 7. SAM Model Processing Flow.

2.6. Peanut Seedling Evaluation Program

The peanut seedling condition analysis system is developed in Python (3.12), integrating computer vision and deep learning technologies, specifically designed for detection, classification, and statistical analysis of peanut seedlings. The system provides intuitive operation via a Graphical User Interface (GUI) and can process multiple input types, including images, videos, or image folders. The core functions of the system are based on the YOLOv11 model for peanut seedling object detection, combined with Segment Anything Model (SAM) for precise segmentation, enabling identification of peanut seedlings and extraction of features such as area, length, and aspect ratio. Through a seedling condition calibration function, users can select representative samples to define criteria for classifying strong seedlings, normal seedlings, weak seedlings, and other types, after which the system automatically classifies detected peanut seedlings according to these criteria. For video input, the system supports setting counting lines to implement dynamic counting, tracking quantity changes in different types of peanut seedlings. The system provides rich visualization results, including original images, detection box markers, segmentation masks, background-removed images, and statistical charts, with results that can be updated in real-time or saved. Final analysis results are displayed in text and chart formats, supporting export as CSV files for further analysis. The overall flow diagram of the program is shown in Figure 8.

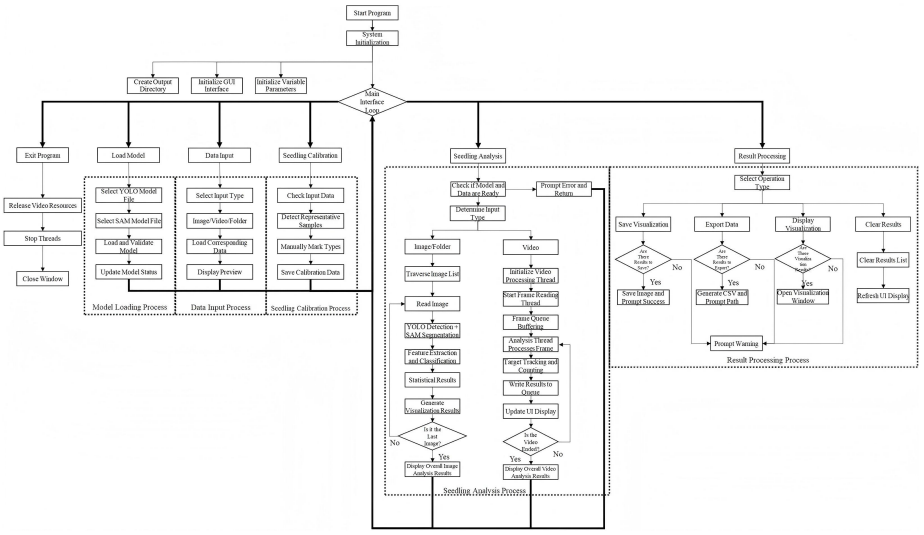


Figure 8. Flowchart of the peanut seedling counting and emergence quality evaluation software.

2.6.1. Counting Function

In static image analysis mode, the peanut seedling counting system adopts a direct statistical method based on object detection. A pre-trained YOLOv11 model scans the entire image to identify the precise positions of all peanut seedlings, and the system obtains the total number of seedlings in the image by counting these detection boxes. During detection, the system assigns a unique spatial identifier to each recognized peanut seedling, generated based on the plant's position coordinates and geometric features in the image. This identification mechanism ensures that even with partially overlapping or densely distributed plants, the system can accurately distinguish adjacent individuals, avoiding duplicate counting or omissions. For potentially overlapping bounding boxes in detection results, the Non-Maximum Suppression (NMS) algorithm is used to retain the bounding box with the highest confidence and eliminate redundant detections.

Video counting is based on multi-target tracking and regional triggering mechanisms, achieving precise counting by fusing object detection, motion prediction, and state judgment. In the object detection stage, the YOLOv11 model processes input video frames, performing feature extraction and prediction to generate an initial detection set D_t, as shown in Equation (4):

$$D_t = \{d_i\}_{i=1}^{m}, d_i = \left(x_1^{(i)}, y_1^{(i)}, x_2^{(i)}, y_2^{(i)}\right) \tag{4}$$

The center point coordinates are calculated as shown in Equation (5):

$$(c_x^{(i)}, c_x^{(i)}) = (\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}) \tag{5}$$

To improve detection quality, the system performs NMS processing, eliminating redundant detections by calculating the Intersection over Union (IoU) between bounding boxes. The IoU calculation between two bounding boxes is shown in Equation (6):

$$IoU(d_i, d_j) = \frac{Area(d_i \cap d_j)}{Area(d_i \cup d_j)} \tag{6}$$

A Kalman filter is then initialized for each detected target, defining a state vector to describe the target's position and velocity information, as shown in Equation (7):

$$\tau_t = \left[c_x, c_y, v_x, v_y\right]^T \tag{7}$$

where v_x, v_y represents velocity.

The state transition matrix F is used to estimate the target's position in the next frame $\hat{\tau}_t$, as shown in Equation (8):

$$\hat{\tau}_t = F\tau_{t-1}, F = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8}$$

where Δt is the frame interval time.

The covariance matrix is updated as shown in Equation (9):

$$P_t = FP_{t-1}F^T + Q, Q = \sigma_a^2 \begin{bmatrix} \frac{\Delta t^4}{4} & 0 & \frac{\Delta t^3}{2} & 0 \\ 0 & \frac{\Delta t^4}{4} & 0 & \frac{\Delta t^3}{2} \\ \frac{\Delta t^3}{2} & 0 & \Delta t^2 & 0 \\ 0 & \frac{\Delta t^3}{2} & 0 & \Delta t^2 \end{bmatrix} \tag{9}$$

where σ^a is the process noise intensity, typically set to 1.0.

When new observation data arrives, the Kalman gain is calculated to balance the credibility of predicted and observed values, obtaining the optimal state estimate, as shown in Equation (10):

$$K_t = P_t H^T (HP_t H^T + R)^{-1} \tag{10}$$

where $H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ is the observation matrix; $R = diag(5, 5)$ is the observation noise covariance.

State update is performed as shown in Equation (11):

$$\tau_t = \hat{\tau}_t + K_t(z_t - H\hat{\tau}_t), z_t = \begin{bmatrix} c_x, c_y \end{bmatrix}^T \tag{11}$$

To associate detected targets across consecutive frames, the Hungarian algorithm is used for data matching. A cost matrix is first constructed, as shown in Equation (12):

$$C_{ij} = \lambda \cdot D_{mahalanobis} + (1 - \lambda) \cdot (1 - IoU) \tag{12}$$

where $D_{mahalanobis} = \sqrt{(d_i - \hat{\tau}_j)^T \sum^{-1} (d_i - \hat{\tau}_j)}$.

The cost comprehensively considers the spatial distance (Mahalanobis distance) and geometric similarity (IoU) between targets. The optimal detection-tracking matching scheme is found by solving the assignment problem.

A virtual counting region with two horizontal lines is set in the video frame. When a target enters the counting region, the system triggers the counting mechanism. To avoid duplicate counting, each target is marked as counted or uncounted. The system maintains multiple counters to record the quantity of different types of targets. The lifecycle of trackers is dynamically managed: trackers that fail to match detections are given a "grace period"; if no re-matching occurs within a specified number of frames, the tracker is removed. Meanwhile, newly detected targets are initialized as new trackers.

Counting results are saved as structured data, recording the filename of each image, detection timestamp, and total number of plants. The system also generates visualized detection marked images, annotating all recognized peanut seedlings with bounding boxes on the original image and displaying statistical numbers in the image corner. The application running interface is shown in Figure 9.

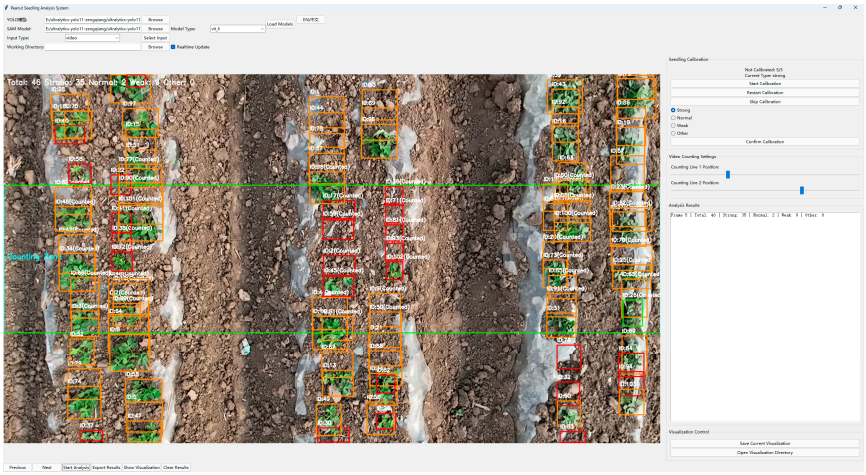


Figure 9. Video evaluation interface of the peanut seedling counting and emergence quality evaluation software.

2.6.2. Seedling Condition Evaluation

The peanut seedling condition evaluation and calibration system completes quantitative analysis of plant growth status through multi-step collaboration. In the calibration phase, the program first triggers a sample collection process: the YOLOv11 model detects peanut seedling bounding boxes in the image, and the SAM is then called to generate a binary mask for each bounding box. The program automatically selects 5 representative samples (2 smallest, 2 largest, and 1 median based on area distribution) and stores them in a list for subsequent interactive calibration. During interactive calibration, users select classification labels via the interface, and feature vectors are bound to manual labels for storage. After all samples are calibrated, the calibration data is converted into a tabular format and saved as a CSV file.

In actual evaluation, classification decisions are made based on calibration data. The program first checks for the presence of calibration data: if empty, fixed threshold rules are applied, where strong seedlings are defined as area >5000, normal seedlings as area >3000, and weak seedlings as area >1000. Strong, normal, and weak seedlings are marked with different colors. Analysis results include six subgraphs: original image, detection results, segmentation masks, classification visualization, background-separated image, and statistical charts, forming a complete evaluation report. The evaluation result report is shown in Figure 10.

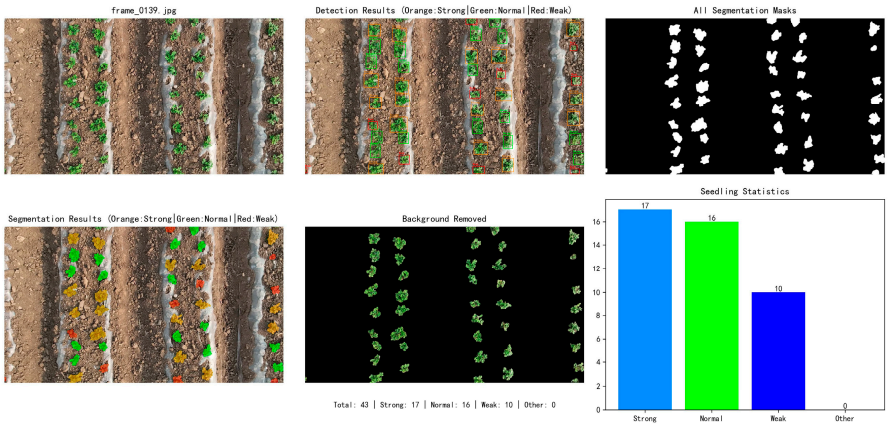


Figure 10. Results of peanut seedling counting and emergence quality evaluation.

2.7. Evaluation Metrics

This study adopts Precision (P), Recall (R), F1-score (F1), mean Average Precision (mAP), parameter count (Params), Mean Absolute Error (MAE), Mean Bias Error (MBE) and Root Mean Squared Error (RMSE) as evaluation metrics [2,5,7,25–29], with formulas as follows:

$$P = \frac{TP}{TP + FP} \tag{13}$$

$$R = \frac{TP}{TP + FN} \tag{14}$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{15}$$

$$AP = \int_0^1 P(R)dR \tag{16}$$

$$mAP = \frac{\sum_{n=1}^{N} AP(n)}{N} \tag{17}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{18}$$

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \tag{19}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{20}$$

where TP (True Positives) refers to samples correctly identified as peanut seedlings, FP (False Positives) refers to samples incorrectly identified as peanut seedlings, FN (False Negatives) refers to instances incorrectly detected as non-peanut seedlings when they are actually peanut seedlings, y_i is the true value, $\hat{y}_i$ is the predicted value, n is the sample size.

3. Results and Analysis

3.1. Performance Verification of the Proposed YOLOv11 Model

To enable the GPU to better train the algorithm, the batch size was set to 32, 300 epochs were executed, and the image size was set to 320. The remaining parameters were consistent with the original parameters in the YOLOv11 model. The curves show that after 300 epochs of training with the above settings, the losses of boxes, objects, and classifications decreased steadily. However, it is evident that the curves showed continuous growth in precision, recall, and mAP after 100 epochs of training. The training result curves are shown in Figure 11.

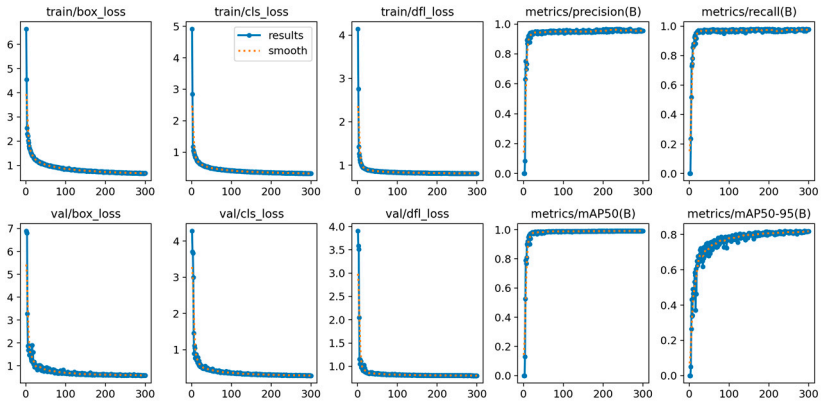


Figure 11. Training and testing loss curves.

To determine the performance of the algorithm, confusion matrices and curves were used to summarize the prediction results. The most prominent finding from the confusion matrix is that the misjudgment rate of peanut seedlings by the improved YOLOv11 algorithm was 2%, the confusion matrix is shown in Figure 12. Precision, recall, precision-recall, and F1-score as functions of confidence are shown in Figure 13. It can be seen from the figure that the model maintains a good F1-score for peanut seedlings and all categories in the medium confidence interval. By selecting an appropriate confidence threshold (around 0.410), the model achieves a good balance between precision and recall. In the high confidence region, the precision is excellent, indicating a low error rate and superior performance in identifying categories such as peanut seedlings. To ensure the correctness of recognition results, a threshold with a confidence of approximately 0.937 can be set. Under the IoU = 0.5 standard, the high mAP@0.5 (0.990) indicates that the model can not only accurately identify seedlings and reduce misjudgments but also cover most real seedlings and reduce missed detections.

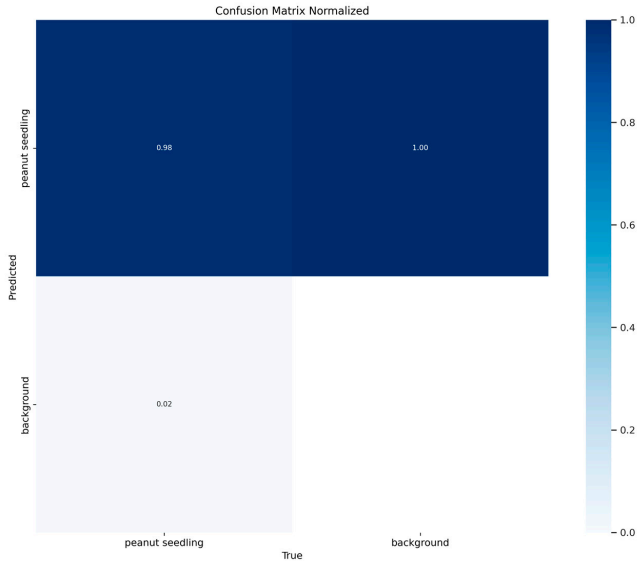


Figure 12. Confusion matrix of different categories.

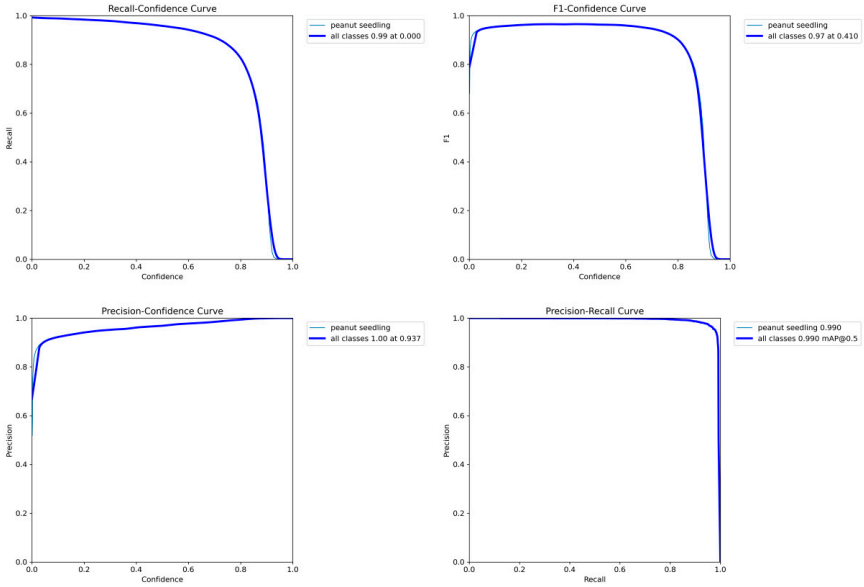
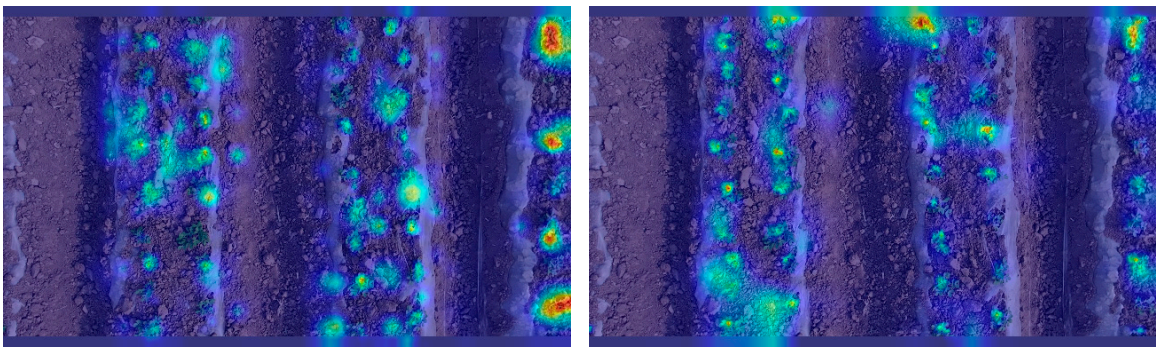


Figure 13. Precision, recall, precision-recall, and F1-score curves under different confidence levels.

3.2. Performance Comparison with Classical Algorithms

To observe the ability of the improved model's neck network to extract key features of peanut seedling targets and enhance the interpretability of the improved method, High-Resolution Class Activation Mapping (HiResCAM) was used to generate class activation heatmaps for input images. The contribution of features to the target category was directly calculated by pixel-wise multiplication of category-specific classification weights with activated feature maps, generating initial heatmaps. The initial heatmaps were interpolated to the original resolution of the input image to improve detail expression, thereby retaining high-resolution information of the target region. The heatmaps generated by HiResCAM can characterize the contribution distribution of the above improvements to the prediction output, revealing the importance of each position in feature extraction. A darker color indicates a higher response of the corresponding region in the original image to the network and a greater contribution; the position and depth of the color distribution also indicate the impact of algorithm improvements on feature extraction ability.

As shown in Figure 14, the original YOLOv11n network had insufficient feature extraction ability, incomplete extraction, and incorrect feature contributions. After using the SBA module, the contribution of correct features to the network was stronger, and more positions affected the prediction results, leading to more comprehensive and better detection performance. Additionally, comparisons were made with other models, and the results are shown in Table 1.



(**a**) (**b**)

Figure 14. Heatmaps of output feature maps (**a**) Before improvement (**b**) After improvement.

Table 1. Comparison of metrics with other models.

Model	Precision P%	Recall R%	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Parameters	Inference Speed (FPS)
YOLOv11	95.21	84.77	94.09	77.85	2.58 M	116.85
YOLOv12	96.1	87.1	94.7	78.0	5.23 M	86.0
Yolov5s_t [14]	91.2	94.0	66.0	/	25.2 M	77.0
MS-YOLOv8 [15]	91.6	94.8	97.5	72.2	1.9 M	272.2
Improved YOLOv11	96.36	96.76	99.03	82.75	3.79 M	78.17

Although the detection speed of the improved YOLOv11 decreased slightly, P, mAP50, mAP50-95, and R increased by 1.15%, 4.94%, 4.9%, and 11.99%, respectively. This indicates that integrating the SBA module into the neck network for multi-scale feature fusion can effectively retain feature information, thereby improving performance. In actual tests, the FPS for detecting a single image was 78.17, meeting the real-time recognition requirements for peanut seedlings.

3.3. Comparative Analysis Using Different Prompts

To select an appropriate prompting method for segmenting peanut seedling plants, four methods were used in the experiment: point prompting, box prompting, point + box combined prompting, and automatic grid point prompting. Among them, the Point Prompt method uses the center point of the YOLOv11 detection box as the input prompt for SAM; this method is computationally simple and fast but may result in inaccurate segmentation in densely planted areas. The Box Prompt method directly uses the YOLOv11 detection box as the input prompt for SAM, achieving high segmentation accuracy and simple implementation, but with average segmentation results for overlapping plants. The Point + Box combined prompting method integrates detection boxes and feature points, achieving the highest segmentation accuracy but requiring additional point selection logic. The automatic grid point prompting method generates uniform grid points within the YOLOv11 detection box as prompts, with high automation but large computational load. Given that peanuts are in the seedling stage and plants are small, the box prompting method can meet segmentation requirements with low computational load. As shown in the figure, the box prompting method also achieved good results in peanut seedling segmentation, accurately segmenting along the plant edges. Therefore, the box prompting method was used in subsequent experiments. The comparison of different prompting methods is shown in Figure 15.

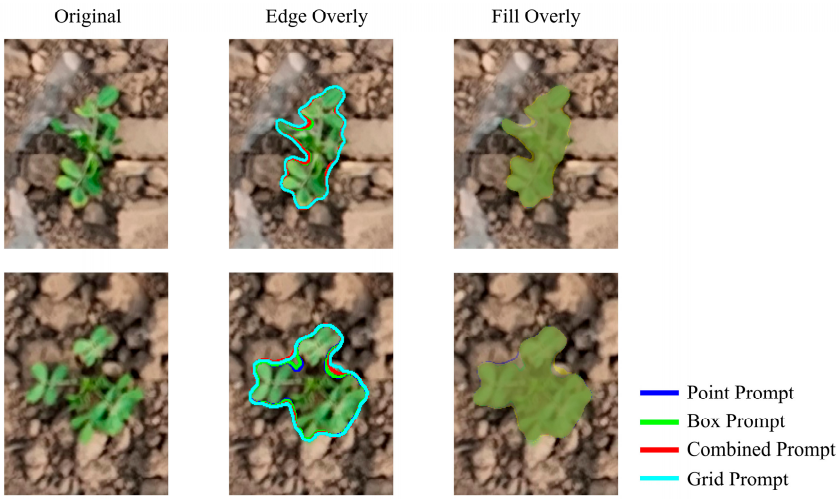


Figure 15. Comparison of different prompting methods.

3.4. Comparative Analysis with Other Segmentation Methods

To thoroughly study the performance differences between the improved YOLOv11 + SAM joint model and other methods in peanut seedling segmentation, under the premise of locating peanut seedlings using YOLOv11, the segmentation effects were compared using the unsupervised learning-based color thresholding method and K-means clustering method. Comprehensive experimental results are shown in Figure 16.

By comparing the actual segmentation results of various models, the SAM employed achieved the best performance. It featured high segmentation integrity, as SAM's feature extraction capability could capture the details of seedlings within the YOLOv11 detection boxes. Its boundaries were clear and sharp, fitting closely to the actual edges of seedlings and reducing boundary offsets. Additionally, it exhibited strong anti-interference ability, focusing on foreground features within the boxes and significantly minimizing missegmentation of background elements such as soil and stones. Moreover, it maintained consistent segmentation quality for seedlings of different sizes and heights. The color thresholding method had obvious limitations: it relied on fixed color ranges, easily missing parts of

seedlings with color variations (such as shaded leaves and tender shoots), which led to inaccurate counting and made it impossible to judge seedling emergence quality. Its boundaries were blurred with a strong jagged effect, and its performance fluctuated significantly under changing light conditions, with a noticeable decline in effectiveness in strong light or cloudy scenes. The K-means clustering method performed poorly: individual seedlings might be split into multiple clusters, leaves appeared relatively fragmented, and fine edges were easily blurred. Due to differences in seedling density and color across different images, the consistency of results was poor. At higher heights, due to limited pixels, there were numerous missed detections.

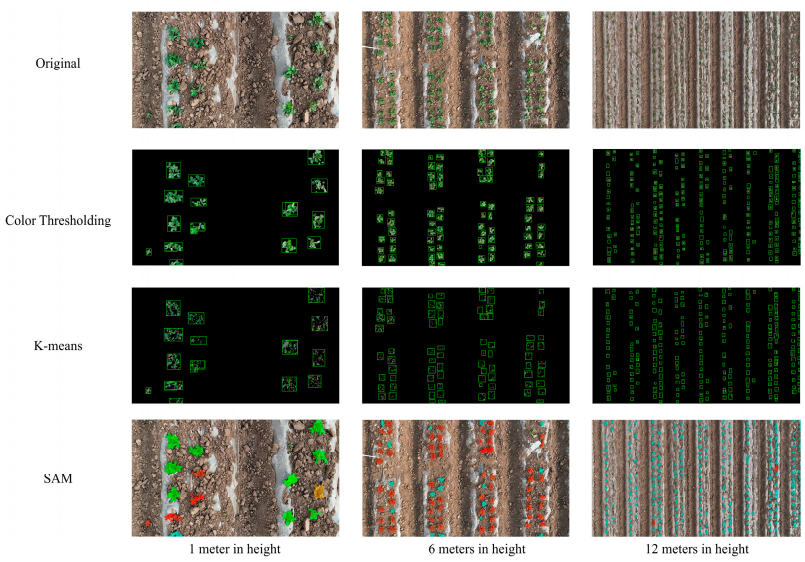


Figure 16. Comparison of different segmentation methods.

3.5. Application Performance

To verify the practical application of the proposed method in peanut seedling quality evaluation, two different experimental schemes were adopted for field experiments: manual evaluation and the method proposed in this paper, with manual evaluation as the reference to judge the accuracy of the proposed method. 10 images each for low, medium, and high heights, and 10 s of video each for low, medium, and high heights were selected for evaluation.

In terms of seedling counting efficiency, the proposed method showed significant advantages compared to traditional schemes. In traditional manual detection methods, manual counting is slow; although manual counting is relatively fast when the number of seedlings is small during low-altitude flights, the speed decreases sharply when the height increases and the number of peanut seedlings in the frame increases significantly, especially for video counting, which consumes a lot of time. The UAV scheme proposed in this paper completed the counting of over 2000 peanut seedlings in only 400 ms, which greatly reduced physical labor intensity and significantly improved counting efficiency.

To further verify the effectiveness of the proposed method in peanut seedling quality evaluation, manual quality classification was still used as the evaluation standard. Experimental data showed that the accuracy varied significantly at different heights: among different modes, the accuracy was highest at medium height, reaching 97.1%; as the number of plants in images increased, the processing time also changed, with an overall positive correlation between the number of plants and processing time; in different types of processing, image detection took less time, while video processing consumed more computing time due to limited computer resources but maintained high accuracy.

Traditional schemes face problems of low efficiency, high labor costs, and susceptibility to subjective biases. The proposed scheme significantly improved data collection efficiency: it not only eliminated the need for personnel to work in the field, reducing physical labor intensity but also achieved approximately 6 to 10 times the efficiency of traditional methods, with advantages such as historical data retrieval. The overall evaluation results are shown in Table 2. Therefore, the proposed method not only exhibits high efficiency in data collection and processing but also provides more comprehensive and objective information for experimental results, enhancing the reliability of peanut seedling condition quality evaluation.

Table 2. Comparison of evaluation results for different types of data.

Mode	UAV Height	Evaluation Precision	MAE	MBE	RMSE	Time
Single Image	Low	85.7%	0.62	−0.11	0.69	92.9 ms
	Medium	97.1%	0.68	0.54	0.74	101.3 ms
	High	96.2%	1.08	−0.91	1.3	83.4 ms
Batch Images	Low	83.1%	1.83	−0.17	2.54	124 ms
	Medium	96.4%	8.25	−3.71	9.63	126 ms
	High	92.3%	67.16	19.62	68.84	439 ms
Video	Low	85.3%	8.15	−2.03	11.47	15 s
	Medium	96.8%	21.76	21.11	24.78	87 s
	High	87.6%	183.46	149.09	223.84	149 s

4. Discussion

This study effectively addresses traditional issues in peanut seedling emergence evaluation through the in-depth integration of deep learning and UAV remote sensing technology. However, there are still several aspects that warrant improvement.

4.1. Limitations of the Dataset

Inherent differences exist in seedling morphology among different peanut varieties. The dataset in this study did not cover samples from multiple varieties or ecological regions, which may reduce the model's generalization ability when promoted across regions. In the future, it will be necessary to expand the coverage of the dataset by incorporating samples of different varieties, soil types, and weed interference levels, as well as supplementing seedling images from different growth stages (e.g., cotyledon stage, true leaf stage). This will enhance the model's adaptability to complex scenarios.

4.2. Balancing Model Performance and Deployment Efficiency

The SAM, with its powerful zero-shot segmentation capability, provides high-precision support for extracting peanut seedling contours but also poses challenges in terms of computational cost. The SAM has a large parameter size and long inference time, making real-time segmentation difficult to achieve on ordinary edge devices. Experiments showed that even on devices equipped with an RTX 4060 Ti graphics card, the segmentation time for a single image still reaches 100–200 ms. When processing video streams or large-area stitched image data, computational efficiency will become a bottleneck. In addition, although the improved YOLOv11 model enhanced detection accuracy through the SBA module, its parameter count increased by approximately 46.9% compared to the original YOLOv11n, and the inference speed decreased from 116.85 FPS to 78.17 FPS. While it still meets real-time requirements, further optimization is needed for deployment on resource-constrained mobile devices. In the future, lightweight segmentation models can be used to

replace the original SAM to reduce computational complexity while ensuring segmentation accuracy. Simultaneously, the SBA module of the improved YOLOv11 can be parameter-pruned, and redundant computations can be reduced through a dynamic channel selection mechanism to balance accuracy and efficiency.

4.3. Adaptability to Practical Application Scenarios Needs Optimization

Dynamic changes in field environments place higher demands on the stability of the evaluation system. Experimental results showed that UAV flight height significantly affects evaluation accuracy: at medium heights (e.g., 5–8 m), the balance between image resolution and coverage is optimal, with an accuracy of 97.1%. When the height exceeds 10 m, the proportion of seedling pixels decreases, easily leading to missed detections due to detail loss, and the accuracy drops to 87.6%. When the height is below 2 m, although individual plant details are clear, the image coverage is small, stitching workload is large, and counting duplication may occur due to perspective deviation. It is necessary to optimize UAV flight strategies, such as automatically adjusting height based on field seedling density, or combining multi-height image fusion technology to improve evaluation robustness through complementarity between high and low-altitude images. In addition, subjectivity remains in the calibration of seedling condition evaluation standards. The current system relies on users to select representative samples to define thresholds for strong and weak seedlings, but "standards for strength and weakness" vary among peanut seedlings in different regions and varieties. In the future, an adaptive threshold learning mechanism can be introduced to establish a dynamic evaluation model, or transfer learning can be used to migrate calibration experience from different regions to new scenarios, reducing the cost of manual intervention.

5. Conclusions

Aiming at the problems of strong subjectivity, low efficiency, and inconsistent standards in manual detection for traditional peanut seedling emergence quality evaluation, this study proposes an intelligent evaluation method based on the fusion of improved YOLOv11 and Segment Anything Model (SAM), realizing rapid and accurate evaluation of peanut seedling emergence status using high-resolution UAV images.

Based on RGB images collected by UAVs, the study constructed a complete intelligent evaluation system for peanut emergence. With YOLOv11n as the core detection framework, the Selective Boundary Aggregation (SBA) module was introduced into the neck network to enhance multi-scale feature fusion capability. This effectively solved the detection challenges caused by blurred edges and varying sizes of peanut seedlings in complex field environments, significantly improving the accuracy of target localization. Meanwhile, the powerful universal segmentation generalization performance of SAM was integrated for the first time, and the box prompting method was used to achieve precise extraction of peanut seedlings without extensive data annotation, providing a basis for seedling condition feature analysis. The system integrates counting and seedling condition evaluation functions, supports multiple input types such as images and videos, and realizes visual display of detection results and data export through a Graphical User Interface (GUI), meeting practical application needs.

Experimental verification shows that the improved YOLOv11 model performs excellently in peanut seedling detection tasks, with a precision of 96.36%, a recall rate of 96.76%, and an mAP@0.5 of 99.03%, all outperforming the traditional YOLOv11n and comparative classical algorithms. The segmentation effect of SAM is significantly better than the color thresholding method and K-means clustering method, effectively capturing detailed features such as seedling stems and leaves. In practical applications, the system's detection

time for a single image is as low as 83.4 ms, the efficiency of video counting is 6–10 times higher than that of manual counting, and the accuracy of seedling condition evaluation reaches 97.1% in medium-height scenarios. These results fully demonstrate the efficiency and reliability of the method, providing efficient and objective technical support for peanut breeding research, mechanical sowing quality evaluation, and field refined management.

Author Contributions: Conceptualization, Q.W. and F.M.; Methodology, C.Z.; Software, G.Z. and F.M.; Validation, G.Z. and G.L.; Investigation, D.C.; Resources, C.Z.; Writing—original draft, G.Z. and F.M.; Visualization, Q.W.; Supervision, Q.W. and F.M.; Project administration, G.Z.; Funding acquisition, G.Z. and Q.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China Youth Science Fund Project, grant number 32301958; Shandong Provincial Natural Science Foundation Youth Project, grant number ZR2021QC161, ZR2023QC057; Shandong Provincial Key Research and Development Program, grant numbers 2022LZGC021 and 2021LZGC026.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Gu, M.; Shen, H.; Ling, J.; Yu, Z.; Luo, W.; Wu, F.; Gu, F.; Hu, Z. Online Detection of Broken and Impurity Rates in Half-Feed Peanut Combine Harvesters Based on Improved YOLOv8-Seg. *Comput. Electron. Agric.* **2025**, *237*, 110494. [CrossRef]
2. Lin, Y.; Wang, L.; Chen, T.; Liu, Y.; Zhang, L. Monitoring System for Peanut Leaf Disease Based on a Lightweight Deep Learning Model. *Comput. Electron. Agric.* **2024**, *222*, 109055. [CrossRef]
3. Wu, S.; Ma, X.; Jin, Y.; Yang, J.; Zhang, W.; Zhang, H.; Wang, H.; Chen, Y.; Lin, C.; Qi, L. A novel method for detecting missing seedlings based on UAV images and rice transplanter operation information. *Comput. Electron. Agric.* **2025**, *229*, 109789. [CrossRef]
4. Cui, J.; Zheng, H.; Zeng, Z.; Yang, Y.; Ma, R.; Tian, Y.; Tan, J.; Feng, X.; Qi, L. Real-Time Missing Seedling Counting in Paddy Fields Based on Lightweight Network and Tracking-by-Detection Algorithm. *Comput. Electron. Agric.* **2023**, *212*, 108045. [CrossRef]
5. Cui, J.; Zhang, X.; Zhang, J.; Han, Y.; Ai, H.; Dong, C.; Liu, H. Weed Identification in Soybean Seedling Stage Based on UAV Images and Faster R-CNN. *Comput. Electron. Agric.* **2024**, *227*, 109533. [CrossRef]
6. Feng, A.; Zhou, J.; Vories, E.; Sudduth, K.A. Evaluation of Cotton Emergence Using UAV-Based Imagery and Deep Learning. *Comput. Electron. Agric.* **2020**, *177*, 105711. [CrossRef]
7. Yang, Z. Plant Recognition and Counting of Amorphophallus Konjac Based on UAV RGB Imagery and Deep Learning. *Comput. Electron. Agric.* **2025**, *235*, 110352. [CrossRef]
8. Yuan, J.; Li, X.; Zhou, M.; Zheng, H.; Liu, Z.; Liu, Y.; Wen, M.; Cheng, T.; Cao, W.; Zhu, Y.; et al. Rapidly Count Crop Seedling Emergence Based on Waveform Method(WM) Using Drone Imagery at the Early Stage. *Comput. Electron. Agric.* **2024**, *220*, 108867. [CrossRef]
9. Zeng, F.; Wang, R.; Jiang, Y.; Liu, Z.; Ding, Y.; Dong, W.; Xu, C.; Zhang, D.; Wang, J. Growth monitoring of rapeseed seedlings in multiple growth stages based on low-altitude remote sensing and semantic segmentation. *Comput. Electron. Agric.* **2025**, *232*, 110135. [CrossRef]
10. Li, B.; Xu, X.; Han, J.; Zhang, L.; Bian, C.; Jin, L.; Liu, J. The estimation of crop emergence in potatoes by UAV RGB imagery. *Plant Methods* **2019**, *15*, 15. [CrossRef]
11. Liu, M.; Su, W.-H.; Wang, X.-Q. Quantitative Evaluation of Maize Emergence Using UAV Imagery and Deep Learning. *Remote Sens.* **2023**, *15*, 1979. [CrossRef]
12. Yang, X.; Li, H.; Zhu, W.; Zuo, Y. RSHRNet: Improved HRNet-based semantic segmentation for UAV rice seedling images in mechanical transplanting quality assessment. *Comput. Electron. Agric.* **2025**, *234*, 110273. [CrossRef]
13. Bai, Y.; Shi, L.; Zha, Y.; Liu, S.; Nie, C.; Xu, H.; Yang, H.; Shao, M.; Yu, X.; Cheng, M.; et al. Estimating leaf age of maize seedlings using UAV-based RGB and multispectral images. *Comput. Electron. Agric.* **2023**, *215*, 108349. [CrossRef]
14. Lin, Y.; Chen, T.; Liu, S.; Cai, Y.; Shi, H.; Zheng, D.; Lan, Y.; Yue, X.; Zhang, L. Quick and Accurate Monitoring Peanut Seedlings Emergence Rate Through UAV Video and Deep Learning. *Comput. Electron. Agric.* **2022**, *197*, 106938. [CrossRef]
15. Zhang, F.; Zhao, L.; Wang, D.; Wang, J.; Smirnov, I.; Li, J. MS-YOLOv8: Multi-Scale Adaptive Recognition and Counting Model for Peanut Seedlings under Salt-Alkali Stress from Remote Sensing. *Front. Plant Sci.* **2024**, *15*, 1434968. [CrossRef] [PubMed]

16. Gao, X.; Zan, X.; Yang, S.; Zhang, R.; Chen, S.; Zhang, X.; Liu, Z.; Ma, Y.; Zhao, Y.; Li, S. Maize Seedling Information Extraction from UAV Images Based on Semi-Automatic Sample Generation and Mask R-CNN Model. *Eur. J. Agron.* **2023**, *147*, 126845. [CrossRef]
17. Li, Q.; Zhou, Z.; Qian, Y.; Yan, L.; Huang, D.; Yang, Y.; Luo, Y. Accurately Segmenting/Mapping Tobacco Seedlings Using UAV RGB Images Collected from Different Geomorphic Zones and Different Semantic Segmentation Models. *Plants* **2024**, *13*, 3186. [CrossRef]
18. Xu, X.; Wang, L.; Liang, X.; Zhou, L.; Chen, Y.; Feng, P.; Yu, H.; Ma, Y. Maize Seedling Leave Counting Based on Semi-Supervised Learning and UAV RGB Images. *Sustainability* **2023**, *15*, 9583. [CrossRef]
19. Yang, R.; Chen, M.; Lu, X.; He, Y.; Li, Y.; Xu, M.; Li, M.; Huang, W.; Liu, F. Integrating UAV remote sensing and semi-supervised learning for early-stage maize seedling monitoring and geolocation. *Plant Phenomics* **2025**, *7*, 100011. [CrossRef]
20. Huang, Z.; Lee, W.S.; Yang, P.; Ampatzidis, Y.; Shinsuke, A.; Peres, N.A. Advanced Canopy Size Estimation in Strawberry Production: A Machine Learning Approach Using YOLOv11 and SAM. *Comput. Electron. Agric.* **2025**, *236*, 110501. [CrossRef]
21. Liu, Y.; Jiang, L.; Qi, Q.; Xie, K.; Xie, S. Online Computation Offloading for Collaborative Space/Aerial-Aided Edge Computing toward 6G System. *IEEE Trans. Veh. Technol.* **2024**, *73*, 2495–2505. [CrossRef]
22. Sapkota, R.; Flores-Calero, M.; Qureshi, R.; Badgujar, C.; Nepal, U.; Poulose, A.; Zeno, P.; Vaddevolu, U.B.P.; Khan, S.; Shoman, M.; et al. YOLO Advances to Its Genesis: A Decadal and Comprehensive Review of the You Only Look Once (YOLO) Series. *Artif. Intell. Rev.* **2025**, *58*, 274. [CrossRef]
23. Tang, F.; Xu, Z.; Huang, Q.; Wang, J.; Hou, X.; Su, J.; Liu, J. DuAT: Dual-Aggregation Transformer Network for Medical Image Segmentation. In Proceedings of the Pattern Recognition and Computer Vision, Xiamen, China, 13–15 October 2023; Liu, Q., Wang, H., Ma, Z., Zheng, W., Zha, H., Chen, X., Wang, L., Ji, R., Eds.; Springer Nature: Singapore, 2024; pp. 343–356.
24. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.-Y.; et al. Segment Anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–3 October 2023. [CrossRef]
25. Shi, J.; Han, D.; Chen, C.; Shen, X. KTMN: Knowledge-Driven Two-Stage Modulation Network for Visual Question Answering. *Multimed. Syst.* **2024**, *30*, 350. [CrossRef]
26. Yu, S.; Baek, H.; Son, S.; Seo, J.; Chung, Y. FTO-SORT: A Fast Track-Id Optimizer for Enhanced Multi-Object Tracking with SORT in Unseen Pig Farm Environments. *Comput. Electron. Agric.* **2025**, *237*, 110540. [CrossRef]
27. Shi, H.; Li, L.; Zhu, S.; Wu, J.; Hu, G. FeYOLO: Improved YOLOv7-Tiny Model Using Feature Enhancement Modules for the Detection of Individual Silkworms in High-Density and Compact Conditions. *Comput. Electron. Agric.* **2025**, *231*, 109966. [CrossRef]
28. Han, D.; Shi, J.; Zhao, J.; Wu, H.; Zhou, Y.; Li, L.-H.; Khan, M.K.; Li, K.-C. LRCN: Layer-Residual Co-Attention Networks for Visual Question Answering. *Expert Syst. Appl.* **2025**, *263*, 125658. [CrossRef]
29. Li, H.; Li, Y.; Zhang, G.; Liu, Y.; Han, Z.; Zhang, H.; Xu, Q.; Zhao, J.; Jin, M.; Song, D.; et al. Directly Printing High-Resolution, High-Performance 3D Curved Electronics Based on Locally Polarized Electric-Field-Driven Vertical Jetting. *Addit. Manuf.* **2024**, *96*, 104579. [CrossRef]

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.