Soybean seedling detection and counting from UAV images based on an improved YOLOv8 Network

Haotian Wu[1] , Junhua Kang[1*], Heli Li[2]

[1]College of Geological Engineering and Geomatics, Chang’an University, Xi’an 710054, China -htwu@chd.edu.cn, -junhua.kang@chd.edu.cn

[2]Key Laboratory of Quantitative Remote Sensing in Agriculture of Ministry of Agriculture and Rural Affairs, Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China -lihl@nercita.org.cn

Key Words: UAV, Soybean Seedling, Object Detection, YOLO, Attention Mechanism

Abstract

The utilization of unmanned aerial vehicle (UAV) for soybean seedling detection is an effective way to estimate soybean yield, which plays a crucial role in agricultural planning and decision-making. However, the soybean seedlings objects in the UAV image are small, in clusters, and occluded each other, which makes it very challenging to achieve accurate object detection and counting. To address these issues, we optimize the YOLOv8 model and propose a GAS-YOLOv8 network, aiming to enhance the detection accuracy for the task of soybean seedling detection based on UAV images. Firstly, a global attention mechanism (GAM) is incorporated into the neck module of YOLOv8, which reallocates weights and prioritizes global information to more effectively extract soybean seedling features. Secondly, the CIOU loss function is replaced with the SIOU loss, which includes an angle loss term to guide the regression of bounding boxes. Experimental results show that, on the soybean seedling dataset, the proposed GAS-YOLOv8 model achieves a 1.3% improvement in mAP@0.5 and a 6% enhancement in detection performance in dense seedling areas, when compared to the baseline model YOLOv8s.When compared to other object detection models (YOLOv5, Faster R-CNN, etc.), the GAS-YOLOv8 model similarly achieved the best detection performance. These results demonstrate the effectiveness of the GAS-YOLOv8 in detecting dense soybean seedlings, providing more accurate theoretical support for subsequent yield estimation.

1. Introduction

Soybean is an important source of plant-based protein and represent the main oilseed consumed worldwide, which is beneficial for human health and crucial for world economic participation(Manenti et al. 2023). In the soybean production process, seedling detection and counting is an effective way to estimate soybean yield, which plays a crucial role in agricultural planning and decision-making. By accurately detecting and identifying soybean seedlings, monitoring and evaluating their growth status and distribution become feasible, enabling precise prediction and management of soybean yields. The conventional approach to assessing soybean seedling emergence has typically involved a combination of manual counting and sampling(Shuai et al. 2019). This method, however, is labor-intensive and susceptible to errors, owing to factors such as plant density, limitations of human visual perception, and the representativeness of sampling. Furthermore, it is challenging to fulfill the need for continuous spatio-temporal monitoring of large-scale fields.

With the rapid development of UAV (Unmanned Aerial Vehicle) technology and computer vision techniques, UAV images have the advantages of cost-effectiveness, flexibility, and timeliness, which are widely used in various fields of precision agriculture, particularly showing unique advantages in crop growth monitoring and yield estimation(She et al. 2024). In conjunction with object detection techniques, automatic recognition of soybean seedlings enables comprehensive monitoring and automated management of soybean cultivation areas, resulting in increased efficiency and reduced labor intensity.

However, the images captured by UAV are affected by various factors, including high flight altitude, lighting and weather conditions, complex field backgrounds, and uneven distribution of soybean seedlings. These factors can result in challenges such as small target size, image blurring, presence of shadows, seedling clustering, and occlusion in the images, which makes it very challenging to achieve accurate object detection and

counting. Most existing detection methods struggle to deal with the complexity of large-scale soybean seedling scenes and suffer from poor detection accuracy. Therefore, an intelligent and efficient method for detecting soybean seedlings is needed to promote the development of soybean yield estimation towards precision and automation.

To this end, this paper proposes an improved GAS-YOLOv8 model based on YOLOv8, aiming to overcome the challenges of soybean seedling detection in complex backgrounds. The proposed model incorporates two key improvements. Firstly, a GAM attention mechanism(Liu et al. 2021) is integrated into the Neck component to enhance its capacity for capturing features of small targets, thereby reducing the miss-detection rate. Secondly, the loss function is modified to SIOU(Gevorgyan, 2022), which addresses the problem of direction mismatch during the prediction process, enabling the model to better identify the position of soybean seedlings and improve the accuracy of soybean seedling position prediction. These two improvements will provide more accurate data support for soybean seedling detection.

2. Related Work

Crop detection is a significant issue in agricultural production management and decision-making. Traditional manual methods rely on agricultural workers to manually and visually count(Si, 2023). However, these methods are time-consuming, limiting detection efficiency, and unable to provide precise location and other relevant data, thereby restricting further analysis and decision-making. In recent years, the continuous development of computer vision technology has paved the way for object detection algorithms to address this challenge. Researchers have proposed various object detection algorithms to enhance the efficiency and accuracy of crop detection. These algorithms can be categorized into two groups: two-stage algorithms, such as R-CNN and Faster R-CNN(Ren et al. 2015), and single-stage algorithms such as YOLO(Redmon et al. 2015) and SSD(Liu et al. 2020). These methods employ training networks to discern and localize regions of crops, thereby achieving the detection and recognition of different crops in agricultural fields.

The two-stage algorithm is widely utilized to detect and locate different parts of crops, including fruits, leaves, or the entire crop. It achieves detection by randomly selecting candidate boxes, extracting features using neural networks, and subsequently feeding the features into classifiers and regressors. It has the advantages such as high accuracy, robust fault tolerance, and effective detection of small targets and shapes. For instance, Rafael A. et al.(2019) conducted a research study on the estimation model for corn seedling density using deep learning techniques based on UAV imagery. This method demonstrates the capability to accurately and reliably estimate the density of corn seedlings in field conditions. Lim et al.(2006). conducted a study on the real-time detection of mango inflorescence in orchard environments using deep neural networks. They utilized Faster R-CNN and SSD detectors to detect the collected mango inflorescence and construct an accurate and fast autonomous pollination robot system.

The single-stage detection method, known for its end-to-end training characteristics, offers advantages such as simplicity, efficiency, strong real-time performance, and the avoidance of error accumulation between stages. However, this method exhibits relatively lower positioning accuracy and faces challenges in handling small targets and dense target scenarios. To address these issues, Earl Vories et al.(2020) proposed a method to evaluate the emergence rate of cotton seedlings in cotton fields. They employed the ResNet18 network to train a processed dataset for detecting the number of cotton seedlings and canopy size in each image. The results demonstrate the significant guiding implications of this method for real-time statistics of cotton forest scores. Liu et al.(2020) proposed the YOLO-Tomato, detector, which introduced the concept of a circular detection box specifically designed for tomato detection. This detection box is capable of matching the position and size of tomatoes and is ultimately integrated into the dense architecture of YOLOv3 for tomato detection.

Compared with traditional manual counting methods, above deep learning-based object detection algorithms can achieve more accurate and efficient crop detection. However, the application of object detection in crop detection still encounters challenges and problems. Take soybean seedling detection as an example. Soybean seedling UAV images involve complex backgrounds, such as dense growth and mutual occlusion. Additionally, different crop types and growth stages can exhibit distinct characteristics, and factors like the presence of lush weeds in crop fields can affect crop identification results. These problems will have a negative impact on the detection of soybean seedlings.

In conclusion, current object detection algorithms can fulfill most crop statistics tasks, but there is still significant room for improvement in detection accuracy. Further in-depth research is needed to reduce false detection rates and missed detection rates,

both in theoretical research and practical applications. In this paper, we proposed an improved network based on YOLOv8, aiming to enhance the model's feature extraction capabilities for small targets, which to some extent addresses the challenges in recognizing densely growing, crowded, and occluded soybean seedlings.

3. Methodology

In this study, we propose an object detection model based on UAV images, called GAS-YOLOv8, using YOLOv8 as the backbone network. This model aims to improve the performance of small object detection, especially in dense and occluded areas, by employing the GAM module and SIOU loss function. The network details of our proposed model are provided in the following section.

3.1 Overall structure of GAS-YOLOv8

3.1.1 YOLOv8: YOLOv8 is the next major update version of YOLOv5 open-sourced by Ultralytics. The YOLOv8 algorithm provides a new state-of-the-art (SOTA) model that can be used for tasks such as object detection, image classification, instance segmentation, and object tracking in the field of computer vision. The YOLOv8 network consists of four modules: input, Backbone, Neck, and head. The input uses Mosaic data augmentation to expand the diversity and quantity of the training set, improving small object detection and the performance and robustness of the network model. The Backbone replaces the C3 module with the C2f module for lightweight processing, parallelizing more gradient flow branches to obtain more gradient flow information. The Neck uses FPN and PAN(Piao et al. 2021) to make features more fully fused, effectively fusing the low-level and high-level features of the network, and improving the model's perception and recognition capabilities for targets at various scales. The Head uses a Decoupled Head and Anchor Free strategy, which differs from the Coupled Head and Anchor Based strategy in YOLOv5, making it more suitable for dense detection.

3.1.1 GAS-YOLOv8: To improve the accuracy of soybean seedling detection, this paper introduces an attention mechanism to fully focus on the feature information of soybean seedlings and enhance detection precision. The global attention mechanism GAM is integrated into the 12th, 15th, 18th, and 21st layers of the Neck part of the YOLOv8 network, as shown in Figure 1. GAM is capable of magnifying global dimension interaction features while reducing information dispersion. It employs a sequential channel-spatial attention mechanism and redesigns the CBAM submodule(Sanghyun et al. 2018). GAM learns to automatically obtain the importance of each feature space and feature channel and assigns different weights based on importance to strengthen the extraction of important features and suppress irrelevant information. This enables more effective extraction of soybean seedling feature information. Furthermore, the original CIOU loss function of YOLOv8 is modified to SIOU, making the loss function smoother. By accurately calculating the similarity between the predicted bounding box and the ground truth bounding box, SIOU alleviates the problem of bounding box drift. In the soybean seedling detection task, SIOU helps the model better recognize the location of soybean seedlings, especially in dense and crowded soybean populations.

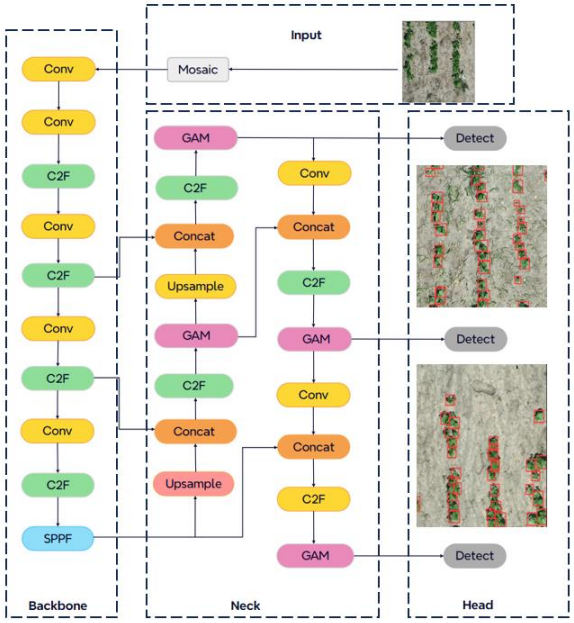

Figure 1 GAS-YOLOv8 Network Architecture

3.2 GAM

The structure of GAM is depicted in Figure 2, which consists of two modules: the channel attention module (CAM) and the spatial attention module (SAM). The input feature F1 undergoes the joint action of the channel attention submodule and the spatial attention submodule, resulting in the output feature F3. The intermediate feature map F2 and the output feature map F3 are given by Equations (1) and (2), respectively.

$$F_2 = M_c(F_1) \otimes F_1 \quad (1)$$

$$F_3 = M_s(F_2) \otimes F_2 \quad (2)$$

M_C and M_S represent channel and spatial attention feature maps respectively, and ⊗ represent element-wise multiplication.

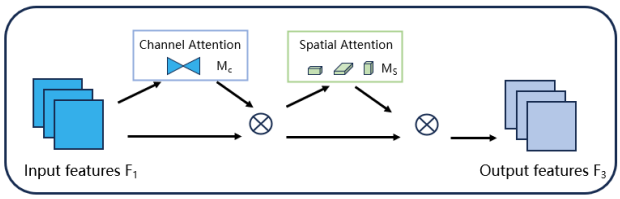

Figure 2 Global Attention Mechanism

3.2.1 Channel attention mechanism: The channel attention mechanism (CAM) is commonly used in tasks such as image classification and semantic segmentation to help the network learn the relationships between different channels and extract important channel information. The structure is shown in Figure 3. Firstly, the input feature maps undergo max pooling and average pooling operations, resulting in two feature vectors with dimensions of 1x1 and the same number of channels as the original feature maps. Subsequently, these two feature vectors are processed by a shared fully connected layer and then added together and normalized using the Sigmoid function. Finally, the normalized result is element-wise multiplied with the input feature maps, thereby incorporating the channel attention mechanism. The channel attention mechanism assigns the weights to the channels to emphasize significant channel information, helping the network automatically learn the importance of different channels. In this paper, we employed CAM in the model to assist the network in better focusing on the color and texture features of soybean seedlings, enabling more accurate differentiation between soybean seedlings, weeds, and other forms of vegetation.

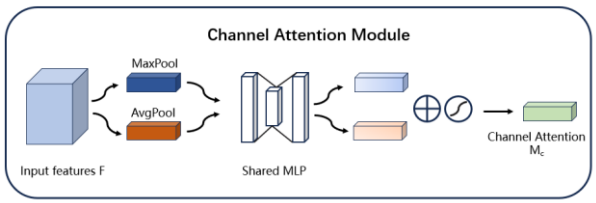

Figure 3 Channel Attention Mechanism

3.2.2 Spatial attention mechanism: The spatial attention mechanism (SAM) is commonly used in tasks such as object detection and image segmentation to help the network focus on the spatial locations of target objects. The structure is shown in Figure 4. SAM first performs average pooling and maximum pooling operations on each feature point of the input image. Unlike the CAM which operates on the height and width, SAM operates on channels. The input feature map is compressed into a feature map with the original height, width, and 2 channels. These two feature maps are then concatenated to create a new feature map, which is then convolved with a 1-channel convolutional layer and processed by the sigmoid function to obtain the weights for each feature point. These weights are element-wise multiplied with the input feature map to incorporate the spatial attention mechanism. SAM focuses on the regions of interest in the image, helping the network concentrate on important areas. In this paper, we employed SAM in the network to pay more attention to the green vegetation in the image, specifically the area of soybean seedlings, while ignoring a large amount of background information. This effectively reduces background interference and improves the accuracy of soybean seedling detection.

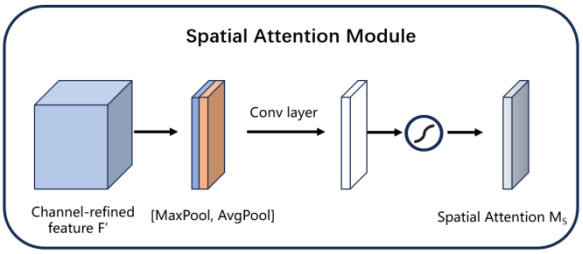

Figure 4 Spatial Attention Mechanism

3.3 SIOU loss function

There are some commonly used loss functions for object detection, such as GIOU, CIOU, DIOU, etc. These loss functions take into account the overlap area, distance, and aspect ratio between the predicted bounding box and the ground truth box. However, none of these loss functions consider the issue of orientation between the ground truth box and the predicted box. If there is a mismatch in orientation, it can lead to the predicted box wandering around, making it difficult to accurately predict the target and resulting in slower convergence and lower efficiency in object detection. To address this problem, Gevorgyan(2022) proposed a new loss function called SIOU, which consists of four components: angle loss, distance loss, shape loss, and IoU loss. The angle loss is defined as follows:

$$\Lambda = 1 - 2 \times \sin^2\left(\arcsin\left(\frac{c_h}{\sigma}\right) - \frac{\pi}{4}\right) \quad (3)$$

The distance loss is defined as follows:

$$\Delta = 2 - e^{-\gamma\rho_x} - e^{-\gamma\rho_y} \quad (4)$$

The shape cost is defined as follows:

$$\Omega = (1 - e^{-w_w})^{\theta} + (1 - e^{-w_h})^{\theta} \quad (5)$$

The IoU cost is defined as follows:

$$\mathrm{IoU} = \frac{\mathrm{Intersection}}{\mathrm{Union}} \quad (6)$$

In summary, the SIOU is defined by the following formula (7)

$$\mathrm{LLoss}_{\mathrm{SIOU}} = 1 - \mathrm{IoU} + \frac{\Delta + \Omega}{2} \quad (7)$$

SIOU redefines the angular penalty metric, enabling the prediction box to swiftly drift towards the nearest axis, subsequently requiring the regression of only a single coordinate. This effective reduction in the total degrees of freedom serves to significantly enhance both the velocity of model training and the precision of inference.

4. Experiments

4.1 Dataset

The soybean seedling dataset used in this experiment was collected in the soybean experimental field in Shijiazhuang city, which is located at a center longitude of E114.724° and latitude of N37.941°, covering an area of approximately 17 acres. The soybean seedlings in this dataset were sown on July 1st. The images were captured at a flight altitude of 12m by a UAV and manually annotated using the LabelImg software. The dataset consists of 3312 images for training and 369 images for validation. Furthermore, the dataset exhibits characteristics such as uneven distribution of seedling sowing and clustering of seedling growth. As shown in Figure 5, (a) shows sparsely sown soybean seedlings, which are relatively easier to detect, while (b) shows densely grown soybean seedlings, posing challenges for precise detection.

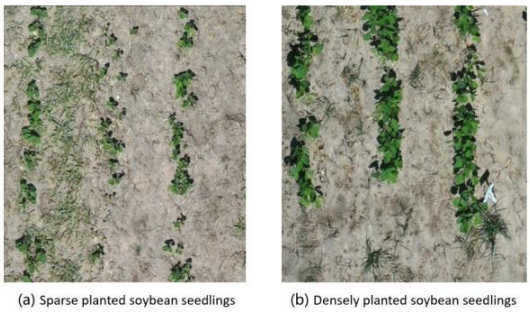

Figure 5 Soybean Seedling Growth Characteristics

4.2 Evaluation Metrics

This experiment evaluates the results using the precision, recall, and mean average precision (mAP) metrics. Precision is defined as the ratio of true positive (TP) predictions to all positive predictions (TP+FP), which represents the number of correctly predicted soybean samples out of all predicted soybean samples. It is calculated using the formula (8):

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad (8)$$

Recall is defined as the ratio of true positive (TP) predictions to all actual positive samples (TP+FN), which represents the number of soybean samples correctly predicted as soybeans. It is calculated using formula (9):

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (9)$$

Mean Average Precision (mAP) is an average precision value across multiple classes, denoted by mAP@0.5 represents the average precision for a specific class when the IoU threshold in the confusion matrix is set to 0.5. mAP@0.5 is the average of precision values for all classes, reflecting the variation of precision with recall and it is defined by formula (10). mAP@0.5:0.95 represents the average mAP across different IoU thresholds (from 0.5 to 0.95 with a step size of 0.05).

$$\mathrm{mAP@0.5} = \frac{1}{C}\sum_{i=1}^{n} \mathrm{AP@0.5_k} \quad (10)$$

4.3 Experiment Settings

This experiment was conducted on the Ubuntu 22.04.4 operating system. The GPU is the NVIDIA GeForce RTX 4090 with a memory capacity of 24209MiB The CUDA Toolkit version is 11.8, and the CUDNN version is 8.9.4. The experiments are performed using the deep learning framework PyTorch 2.0.0. The specific parameter settings are summarized in Table 1.

Parameter	Value
Input Size	512
Initial Learning Rate	0.006
Minimum Learning Rate	0.1
Learning Rate Schedule	cos
Early Stopping	50
Epochs	300
Batch Size	32
Momentum	0.937
Image Flip (Vertical)	0.5
image rotation	0.2

Table 1 Specific Experimental Parameters for GAS-YOLOv8

5. Experimental Results and analysis

5.1 Comparison with YOLOv8

In this section, a comparison was made between GAS-YOLOv8 and YOLOv8 in terms of boundary loss, classification loss, distribution focal loss, and the mAP@0.5 curve, as shown in Figure 6. The results indicate that with increasing training epochs, GAS-YOLOv8 converges faster and achieves lower values for all loss metrics compared to the baseline YOLOv8 model. In particular, the YOLOv8 model exhibits overfitting in the distribution focal loss towards the end of training, while the proposed model avoids this issue. As for mAP@0.5, the improved GAS-YOLOv8 achieves higher accuracy with increasing training epochs. These results demonstrate that the improved GAS-YOLOv8 model in this study achieves better performance in training.

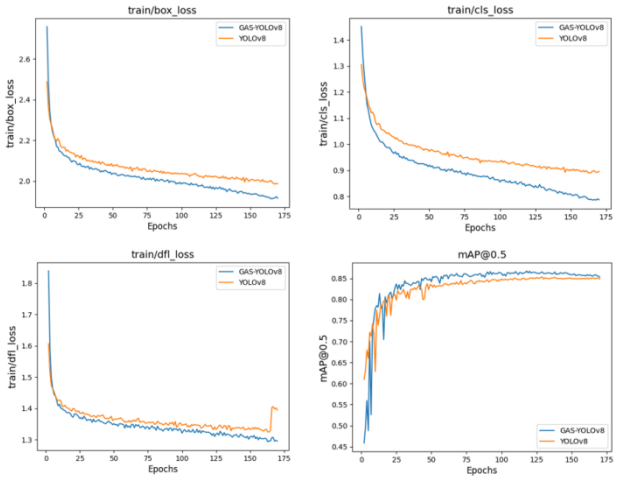

Figure 6 Comparison of Various Losses and mAP@0.5 Curves between GAS-YOLOv8 and YOLOv8

5.2 Ablation Studies

To validate the effectiveness of the two proposed improvement modules on YOLOv8, four ablation experiments were designed as shown in Table 2. The symbol "√" indicates the presence of a module in the model, while "×" indicates its absence. The experimental results show that the baseline YOLOv8n model achieves a mAP@0.5 of only 85.4%. When only the GAM module is added, the precision improves by 0.3%, recall improves by 3.1%, and mAP@0.5 increases by 1%. When only the SIOU module is added, mAP@0.5 improves by 0.9%, and there are corresponding improvements in precision and recall. When both the GAM and SIOU modules are added, mAP@0.5 reaches 86.7%, an improvement of 1.3% compared to the baseline model original, and the precision and recall increase by 1.57%and 2.78%, respectively. This indicates that GAS-YOLOv8 shows improvements in various metrics.

GAM	SIOU	P	R	mAP@0.5	mAP@ 0.5:0.95
×	×	86.1	79.02	85.4	36.4
√	×	86.41	82.13	86.4	36.5
×	√	86.35	81.47	86.3	36.8
√	√	87.67	81.80	86.7	37.0

Table 2 Ablation Experiments

5.3 Comparison with other methods

To assess the effectiveness of our proposed method in comparison with other existing methods, we have conducted experiments comparing GAS-YOLOv8 with other models, including Faster-RCNN, YOLOv5n, YOLOv5s, YOLOv7x, and YOLOv8n, as shown in Table 3. From the table, it can be observed that Faster-RCNN exhibits relatively poor detection results due to the loss of feature extraction for small objects caused by its deep network architecture. In contrast, our proposed algorithm demonstrates improvements in precision, recall, and mAP@0.5 when compared to other YOLO series algorithms. Specifically, the mAP@0.5 metric of our algorithm surpasses that of the relatively higher-performing YOLOv5s model by 0.7%. Moreover, when compared to the YOLOv8n model, which exhibits higher precision, our algorithm achieves a 0.57% improvement. Notably, the results of the mAP@0.95 comparison consistently indicate that our algorithm outperforms other

network models in terms of detection performance. In summary, the GAS-YOLOv8 model exhibits superior detection performance compared to the current mainstream models.

Models	P/%	mAP@0.5/%	mAP@0.95/%
Faster RCNN	79.32	82.38	32.23
YOLOv5n	82.90	86.0	35.50
YOLOv5s	83.03	86.0	35.70
YOLOv7x	80.20	84.6	32.51
YOLOv8n	86.10	85.4	36.40
GAS-YOLOv8	87.67	86.70	37.0

Table 3 Comparison of Accuracy Metrics across Different Models

5.4 Counting and Detection Results

5.4.1 Counting results: To evaluate the performance of the improved GAS-YOLOv8 algorithm in counting for soybean seedlings, the actual number of soybean seedlings in each experimental area was obtained by manually counting. The real soybean seedling number was then compared with the predicted seedling number, as shown in Table 4. From the table, it can be seen that the Faster RCNN model predicts only 5546 seedlings, which is 799 seedlings (9.56%) less than the 6345 seedlings predicted by the YOLOv8 network. This significant discrepancy indicates the relatively poor performance of the Faster RCNN model in counting and detection for soybean seedlings. However, the proposed GAS-YOLOv8 in this paper exhibits superior performance compared to YOLOv8. The predicted number of seedlings for GAS-YOLOv8 is 6789, which is an increase of 444 seedlings compared to the baseline model YOLOv8. The detection rate of the YOLOv8 network for seedlings is only 75.88%, whereas the improved GAS-YOLOv8 algorithm achieves a detection rate of 81.19%, representing a 5.31% improvement in detection rate. Based on the aforementioned comparative analysis, it is shown that the proposed GAS-YOLOv8 model has achieved significant results in terms of detection rate.

Models	Ground Truth	Predicted	Detection Rate
Faster RCNN	8362	5546	66.32%
YOLOv8	8362	6345	75.88%
GAS-YOLOv8	8362	6789	81.19%

Table 4 Comparison of Count Estimation among Different Network Models

To further validate the superiority of the proposed GAS-YOLOv8 model in small object detection, we divided the soybean experimental data into five areas based on the different densities. The distribution of soybean seedling density in each area is shown in Figure 7. From the figure, it can be seen that the density of seedlings in each area is mainly concentrated between 50 and 90 seedlings, with 9% of the areas having a density of over 110 seedlings. These areas exhibit dense planting, making it more challenging to achieve accurate estimation.

Figure 8 shows the detection rates for each density area. Through comparative analysis, it was found that YOLOv8 outperforms Faster RCNN in every density area. In the low-density area, YOLOv8 achieves a detection rate of approximately 80%, while in the high-density area, the prediction rate reaches around 60%. These prediction results far exceed the 65%-70% prediction rate of Faster RCNN. Furthermore, the improved GAS-YOLOv8 algorithm demonstrates significant improvements in prediction for different soybean seedling densities. The detection rates in the 70-90 and 90-110 density intervals increased by 4.41% and 5.83%, respectively. Additionally, there is a 6.23% improvement in the detection rate in the highest density interval of 110-140. These results show the proposed method exhibits certain improvements in detecting small target seedlings, particularly in scenarios involving crowded growth and mutual occlusion.

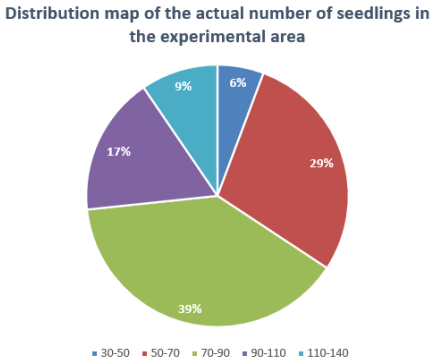

Figure 7 Distribution of True Soybean Seedling Count in the Experimental Area

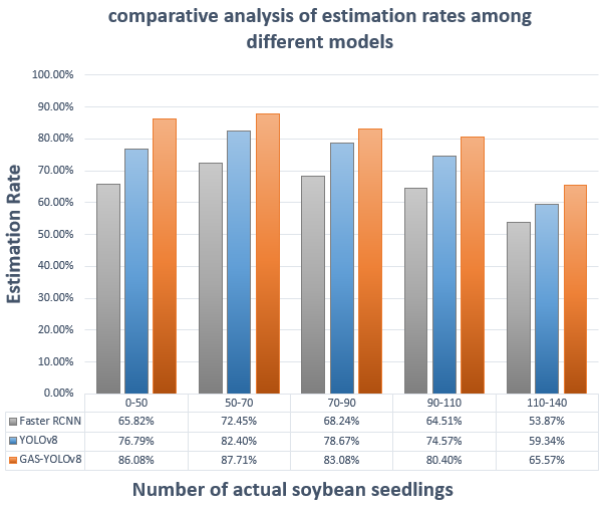

Figure 8 Comparison of Estimation Rates at Different Density Intervals across Models

5.4.2 Visualization Analysis: To visually illustrate the efficacy of the proposed algorithm in detecting dense and crowded small objects, we present a comparison of the visual detection results predicted by different models in Figure 9. From the figure, it is clear that our model detects a larger number of soybean seedlings and achieves superior detection performance. Notably, it is capable of detecting targets that the baseline model fails to identify. Particularly in the regions indicated by the arrows in Figure 9, these areas are densely populated with soybean seedling growth. In the detection of seedlings in such high-density areas, both Faster RCNN and YOLOv8 exhibit relatively poor detection performance, while our GAS-YOLOv8 model demonstrates significant advantages in the detection results. Therefore, it can be concluded that the proposed GAS-YOLOv8 model exhibits superior detection performance and higher recognition accuracy in handling densely grown and crowded soybean seedlings. This is of significant importance for soybean monitoring and management in the agricultural field, as it enables more accurate and efficient tracking of soybean growth and development.

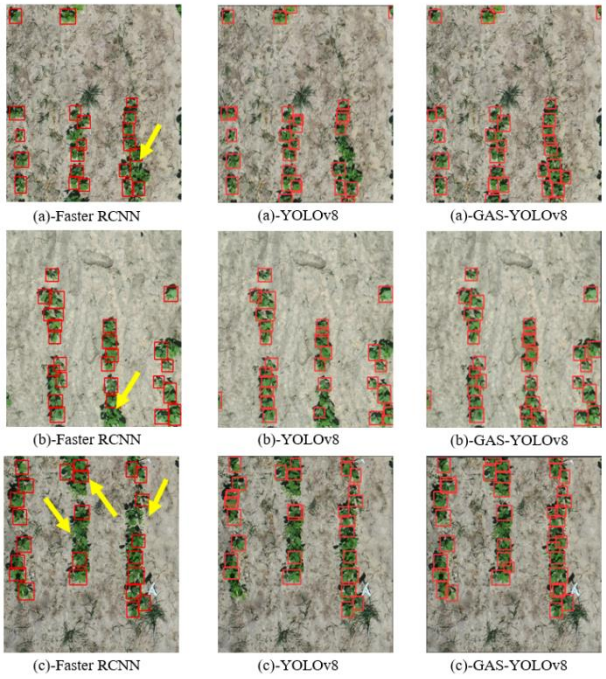

Figure 9 Visual Comparison of Detection Results from Different Models

6. Conclusion

In this paper, we present a novel soybean seedling detection network model, GAS-YOLOv8, based on the YOLOv8 architecture for UAV imagery. The proposed model incorporates a global attention mechanism to effectively capture and extract the feature information of soybean seedlings, enhancing the overall accuracy of the detection process. Furthermore, we have replaced the original loss function of YOLOv8 with the SIOU loss function to improve the precision of soybean seedling position predictions and accelerate the convergence speed during the training phase. Our experimental results demonstrate that GAS-YOLOv8 achieves a significant improvement in mean average precision (mAP@0.5) of 1.3% on a soybean dataset and a notable 5.31% improvement in detection rate. The visual results show that our model can achieve favorable performance in the task of soybean detection and counting, especially in dense and occluded areas.

Compared to other mainstream soybean seedling detection models, GAS-YOLOv8 exhibits superior detection rates, providing more reliable data support for soybean detection and yield estimation tasks. However, it is important to note that the inclusion of the GAM increases the computational complexity of the proposed network model, resulting in a larger model size. Therefore, there is still considerable room for improvement in

optimizing the proposed network model for lightweight processing to better meet the real-time requirements of modern agricultural detection applications. In future work, we plan to focus on optimizing the model architecture to reduce its computational complexity while maintaining its high detection accuracy.

7. References

Feng Aijing, Zhou Jianfeng, Vories Earl, et al. Evaluation of cotton emergence using UAV-based imagery and deep learning[J]. Computers and Electronics in Agriculture, 2020,177:105711.

Gevorgyan, Z. (2022). SIoU Loss: More Powerful Learning for Bounding Box Regression. ArXiv, abs/2205.12740.

Lim J Y, Ahn H S, Nejati M, et al. Deep neural network based real-time kiwi fruit flower detection in an orchard environment[J]. arXiv preprint arXiv:2006.04343.

Liu G, Nouaze JC, Touko Mbouembe PL, Kim JH. YOLO-Tomato: A Robust Algorithm for Tomato Detection Based on YOLOv3. Sensors. 2020;

Liu, W. et al. “SSD: Single Shot MultiBox Detector.” European Conference on Computer Vision (2015).

Liu, Y., Shao, Z., & Hoffmann, N. (2021). Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions. ArXiv, abs/2112.05561.

Manenti L ,Macholdt J ,Garcia O F , et al.Resilience of maize, wheat, and soybean cropping systems as affected by fertilization: Analysis of a long-term field network[J].Agronomy Journal,2023,115(4):2017-2029.

Piao Yongri,Jiang Yongyao, Zhang Miao, et al. PANet: PatchAware Network for Light Field Salient Object Detection[J]. IEEE trans- actions on cybernetics,2021.

Redmon, Joseph, et al. “You Only Look Once: Unified, Real-Time Object Detection.” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015): 779-788.

Ren, Shaoqing, et al. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2015): 1137-1149.

She B, Hu J, Huang L, et al.Mapping Soybean Planting Areas in Regions with Complex Planting Structures Using Machine Learning Models and Chinese GF-6 WFV Data[J].Agriculture,2024,14(2):

Shuai Guanyuan, Martinez-Feria Rafael A., Zhang Jinshui, et al. Capturing Maize Stand Heterogeneity Across Yield-Stability Zones Using Unmanned Aerial Vehicles (UAV)[J]. Sensors, 2019,19(20):4446.

Sijia Z ,Xuyang B ,Tian X , et al.Identification of Soybean Planting Areas Combining Fused Gaofen-1 Image Data and U-Net Model[J].Agronomy,2023,13(3):863-863.

Woo, Sanghyun et al. “CBAM: Convolutional Block Attention Module.” ArXiv abs/1807.06521 (2018): n. pag.