

Identification and Counting of Field Peanut Seedlings Using Improved Centernet from UAV imagery

Zhisen Wang¹, Hongyu Zhao¹, Juntao Yang^{1,*}, Mingxuan Song¹, Yirou Liu², Zhenhai Li¹, Bo Bai², Guowei Li²

¹College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China;

²Institute of Crop Germplasm Resources, Shandong Academy of Agricultural Sciences, Ji'nan 250100, Shandong Province, China.

Keywords: Object detection, Bidirectional Feature Pyramid Network, Contrastive Loss, Seedling emergence rate, UAV imagery.

Abstract

The seedling emergence rate is a crucial indicator for evaluating the growth status of crops in agricultural production and can provide valuable recommendations for subsequent crop planting and field management strategies. Currently, the determination of the emergence rate relies on manual seedling counting, which is not only labour-intensive and time-consuming, but also prone to human errors. Therefore, we utilize drone-captured images of peanut seedlings and employs deep learning networks to estimate seedling numbers. Specifically, we incorporate the BIFPN (Bidirectional Feature Pyramid Network) feature fusion module into the original Centernet model, which would combine multi-scale feature information. This modification not only enhances the accuracy of identification but also improves the localization of seedlings. To address the issue of false positives caused by complex field backgrounds in seedling recognition, we integrate the Contrastive Loss module to increase the discrepancy between positive and negative samples. The results demonstrate that the proposed method significantly enhances both precision and recall rates for peanut seedling recognition under three different scenes, compared to the original model. Furthermore, the proposed method is also applied in real peanut breeding field, fulfilling the practical requirements for emergence rate calculation.

1. Introduction

In agricultural production activities, the seedling emergence rate is an important indicator to evaluate the growth status of crops (Götz and Bernhardt, 2010, Abdin et al., 2000). The seedling emergence rate reflects the impact of soil environment, temperature, and moisture on crop emergence (Al-Mulla et al., 2014, Lindstrom et al., 1976). Combined with the planting density of crops, it provides reasonable recommendations for the proper planting of crops (Jin et al., 2017). To calculate the seedling emergence rate of crops, it is necessary to accurately obtain the number of seedlings. The traditional method of counting seedlings is mainly through manual counting. However, manual counting not only consumes excessive time and effort but also has the drawback of low accuracy, leading to a significant difference between the counted result and the actual number of peanut seedlings, which fails to meet the requirements of precision agriculture. Therefore, an efficient and accurate method for counting peanut seedlings is needed.

In recent years, computer vision has been widely applied in seedling recognition in precision agriculture (Andvaag et al., 2024, Cui et al., 2023, Bai et al., 2022, Osco et al., 2021, Tan et al., 2022). The visual features of crops can help the model capture the most distinguishable information in an image, so the ability to extract visual features directly affects the accuracy and efficiency of the model in crop recognition. With effective visual feature extraction methods, the model can extract key visual information from complex crop images, providing a reliable foundation for subsequent recognition and counting. Most researchers defined and designed appropriate visual features based on prior knowledge. Frequently-used visual features include edge, color, and texture (Liu et al., 2019, Zou et al., 2019, Jidong et al., 2016). Following this, the traditional classifiers like Support Vector Machine (SVM), Decision Trees, and K-Nearest Neighbors (KNN) based on these visual features were used to achieve the purpose of prediction and recognition (Zou et al., 2019, Genaev et al., 2019, Kuzdraliński

et al., 2017, Zhao et al., 2018). What's more, some researchers have also developed seedling models using features such as the canopy area and leaf polygons of crops, achieving good results (Gnädinger and Schmidhalter, 2017).

Due to its superior feature representation capabilities, deep learning architectures gained increasing interests in extracting and learning the visual features of crops from raw imagery. When extracting visual features of crops, hierarchical feature extraction is used, where simpler features are extracted in shallow layers and more complex features are learned in deeper layers. The advantage of this approach is that the model can automatically select the most relevant features at different levels, enhancing the model's ability to autonomously learn crop features in an end-to-end manner (Nielsen, 2015). Among them, Convolutional Neural Network (CNN) is the most typical deep learning architectures. The hierarchical structure has made it widely used in seedling prediction (Osco et al., 2021). Later, some researchers improved the CNN network to develop the Faster Region-based Convolutional Neural Network (R-CNN), which incorporates a Region Proposal Network structure for detecting crops or seedlings, achieving better results (Gao et al., 2020, Halstead et al., 2018). In the case of using the Centernet network, some researchers employed Resnet50 as the backbone network, combined with 900 images collected from the field, achieving mean Average Precision (mAP) and Average Recall (AR) of 79% and 73%, respectively (Lin and Guo, 2021).

Although the lightweight and speed improvements have enhanced the efficiency of crop recognition in static images, most datasets are collected with a single experimental background. These models cannot be directly applied to complex, high-resolution crop overlapping images (Tan et al., 2022). Therefore, we have improved the original Centernet model by adding the Bidirectional Feature Pyramid Network (BiFPN) module, which enhances the model's ability to process multi-scale features and perform well across different scenes. Additionally, we introduced the Contrastive Loss module that

* Corresponding author. Email: jtyang@sdust.edu.cn

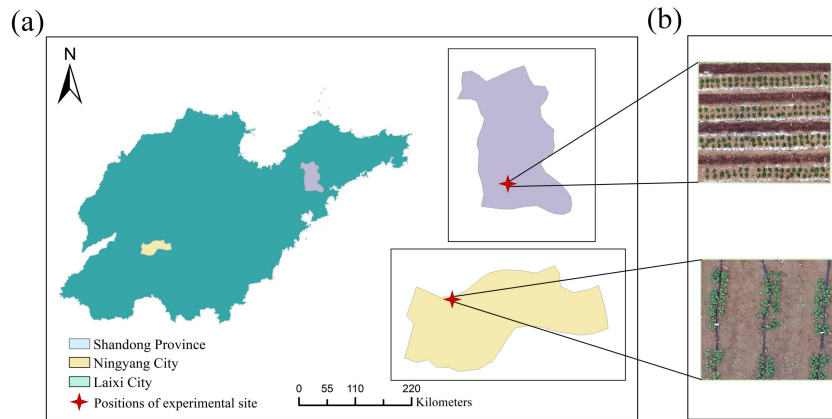


Figure 1. Study area and peanut seedling examples: (a) Location of the study area; (b) Some examples from Laixi City and Ningyang County, respectively.

Scene	Location	Date	Flight	Weather	Camera type
1	Laixi City	May 28, 2023	15m	Rainy	M3M
2		May 31, 2023	15m	Sunny	
3		May 31, 2023	25m	Sunny	
4	Ningyang City	June 10, 2024	20m	Sunny	

Table 1. Introduction to peanut seedling data in different scenes

Scene	Training set (images)	Validation set (images)	Test set (images)
1	Train1(424)	Val1(53)	Test1(53)
2	Train2(408)	Val2(51)	Test2(51)
3	Train3(144)	Val3(18)	Test3(18)
4	-	-	Test4(800)

Table 2. Statistical information of the training set, validation set, and test set in different scenes

strengthens the model's ability to recognize crops in complex surface environments by making the model focus more on the differences between positive and negative samples during training, thereby improving the overall detection performance of the model. Compared with existing methods, the improved Centernet combining the BiFPN module and Contrastive Loss, not only performs excellently in single-seedling detection but also achieves effective recognition in environments with seedling occlusion.

2. Materials and Methods

2.1 Data Acquisition and Processing

As shown in Figure 1, the data acquisition locations in this paper are Laixi City in Qingdao, Shandong Province and Ningyang County in Tai'an City, Shandong Province. The tool used for data collection was the DJI Mavic 3M drone. We created datasets for four different scenarios. The first three datasets were collected in Laixi City, where the drone flight heights were set at 15m, 15m, and 20m, respectively, with corresponding drone speeds of 1m/s, 1.2m/s, and 1.5m/s. The dataset for the fourth scenario, which focuses on peanut seedlings, was obtained from Ningyang County, with the drone flight height set to 25m and the corresponding speed set to 2m/s. Except for the first scenario, where data was acquired after rainfall, all other data were collected under clear weather conditions.

For data processing, we split the data from the first three scenarios into training, validation, and testing sets in a ratio of 8:1:1. As for the data from the fourth scenario, we did not

perform any splitting and used all the images as a test set to evaluate the generalization performance of the proposed method in real-world peanut seedling recognition.

2.2 Overall Architecture

Our proposed model incorporates two new modules: BIFPN and Contrastive Loss that is derived from the idea of contrastive clustering (Li et al., 2021), as shown in Figure 1. These two modules play an important role in improving the performance of recognition. BIFPN combines feature information with positional information, strengthening the model's ability to locate samples while ensuring recognition capability. This reduces the Offset Loss and Size Loss in the loss function, making the parameter updates more reasonable. The Contrastive Loss module with its clustering features processes the feature vectors of the positive and negative samples and reduces the differences between similar samples while increasing the differences between different samples, thus improving the model's ability to distinguish between positive and negative samples. These modules together form a more rational sample recognition structure.

2.3 Feature Fusion

To improve the model's performance in detecting peanut seedlings and predicting their positions, we incorporated the BIFPN into the Hourglass network. Compared to traditional feature fusion modules, BIFPN has the following characteristics. Firstly, BIFPN uses bidirectional feature fusion by introducing bidirectional feature paths, allowing the visual features to be fused and exchanged at each level. Specifically, it passes visual

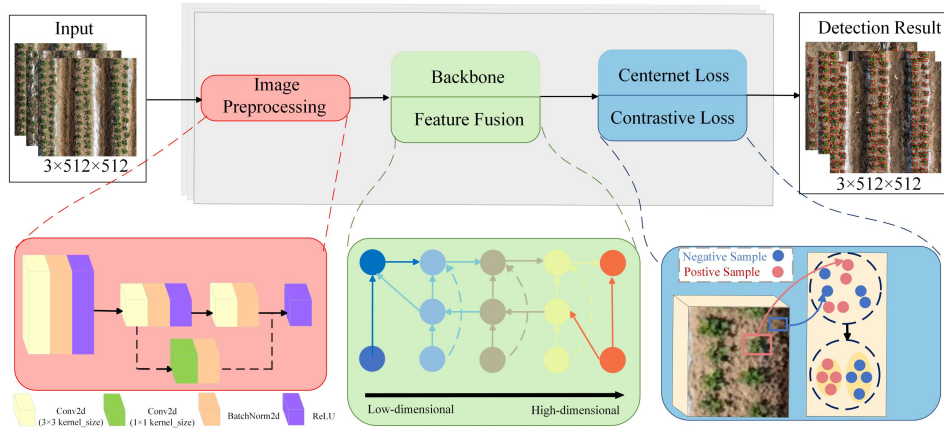


Figure 2. Flowchart for identification using the improved Centernet network.

features from the low layers, which carry more accurate positional information, to the high layers, enhancing the model's ability to locate peanut seedlings' positions. At the same time, it passes visual features from the high layers, which carry stronger semantic information, to the low layers, improving the model's ability to recognize peanut seedlings. Secondly, we removed nodes in BiFPN that only had one input edge since these nodes do not perform feature fusion and contribute little to the feature extraction network. Moreover, the presence of these nodes increases the overall computational cost and causes unnecessary resource consumption. Finally, unlike traditional feature fusion, which simply stacks feature maps, BiFPN adopts weighted feature fusion, which takes into account that feature maps with different resolutions contribute differently to the final fusion.

For the features to be fused at a certain layer P_i , the first step is to fuse with low-dimensional features P_{i+1} to obtain the position of the peanut seedlings and shallow feature information. P_i^{in} represents the features input from the backbone network. Then, the preliminary fused feature P_i^h is further fused with the high-dimensional feature P_{i-1} , thereby obtaining the deep feature information of the peanut seedlings. P_i^{out} represents the output feature after feature fusion. w_i is the corresponding weight. b_i is the corresponding bias.

$$P_i^h = \text{Conv}\left(\frac{w_1 \cdot P_i^{\text{in}} + w_2 \cdot \text{Resize}(P_{i+1}^{\text{in}})}{w_1 + w_2 + b_1}\right) \quad (1)$$

$$P_i^{\text{out}} = \text{Conv}\left(\frac{w_1' \cdot P_i^h + w_2' \cdot P_{i-1}^h + w_3' \cdot \text{Resize}(P_{i-1}^{\text{out}})}{w_1' + w_2' + w_3' + b_2}\right) \quad (2)$$

2.4 Loss function definition

Through observing the prediction results from the original Centernet model, we found that the original Centernet model might misidentify the soil background as peanut seedlings. Therefore, in the loss calculation part of the original Centernet model, we introduced the Contrastive Loss function module using a weighted approach in the loss functions (HeatMap Loss, Offset Loss, Size Loss) to enhance the model's ability to recognize and cluster different samples at the feature level. Specifically, we extracted the feature values of correctly identified peanut seedlings and land background from the current model, forming feature vectors to be used as input in the Contrastive Loss function for both positive and negative samples. By calculating the Contrastive Loss, we aimed to increase the similarity between the same samples $s_{i,i}$, while decreasing the similarity between different samples $s_{i,k}$, thus

better distinguishing between peanut seedlings and the land background. Finally, the Contrastive Loss value was weighted with the original loss value and used for backpropagation to update the network parameters, resulting in a more accurate model.

$$\mathcal{L}_c = -\log \left[\frac{\exp(s_{i,i}/\tau)}{\sum_{k \neq i} \exp(s_{i,k}/\tau) + \exp(s_{i,i}/\tau)} \right] \quad (3)$$

$$\mathcal{L} = \mathcal{L}_{\text{Heat}} + \mathcal{L}_{\text{Off}} + \mathcal{L}_{\text{Size}} + \mathcal{L}_c \quad (4)$$

3. Experimentation and analysis

3.1 Training and Evaluation Procedure

In our implementation, training and evaluation are conducted on a Windows 11 operating system equipped with an NVIDIA GeForce RTX 3050Ti graphics card and 16GB of RAM, using the PyTorch 2.1.0 deep learning framework and a Python 3.10 development environment. To obtain an initial trained model, we first select ResNet50 and Hourglass as backbone networks combined with the BiFPN feature fusion module to train the data from scenes 1, 2, and 3. The resolution of the input images is set to 512×512 pixels, the learning rate is initialized to 0.0007, the batch size is set to 8, and the backbone is trained for 120 epochs without freezing to ensure model convergence. Afterward, a Contrastive Loss module is added to fine-tune the model. At this point, the learning rate is set to 0.0004, and the backbone is frozen while training for an additional 40 epochs. During the entire training process, the Adam optimizer is used. The model's performance is evaluated every 5 epochs, and the weight file is saved every 10 epochs. These settings guide the model's training.

For accuracy evaluation, we calculate precision, recall, mean average precision, and F1-Score based on TP , FP , and FN to assess the model's accuracy and stability. The higher values these metrics, the better model performance.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$F_{1-\text{score}} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

$$\text{AP} = \sum_{k=0}^{K-1} [\text{Recalls}(k) - \text{Recalls}(k+1)] \times \text{Precisions}(k) \quad (8)$$

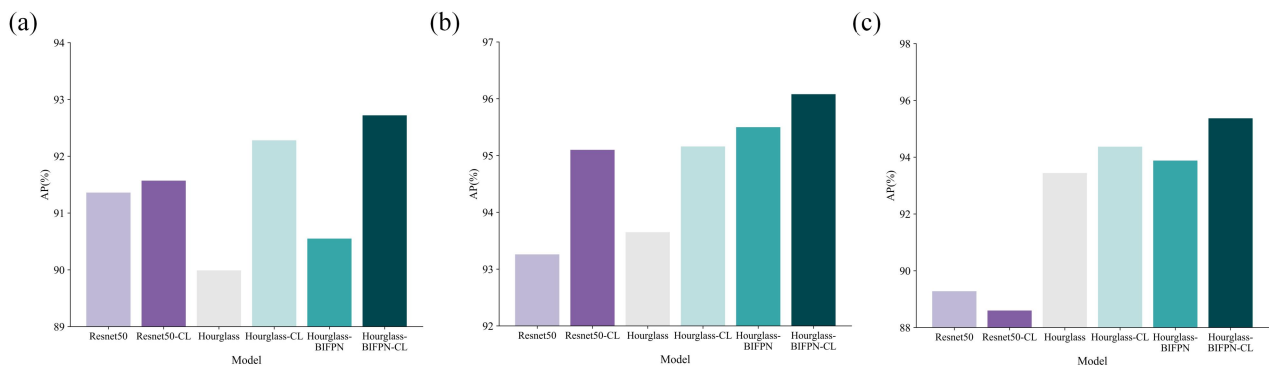


Figure 3. Comparison of the performance of different models across different scenes: (a) Scene 1; (b) Scene 2; (c) Scene 3.

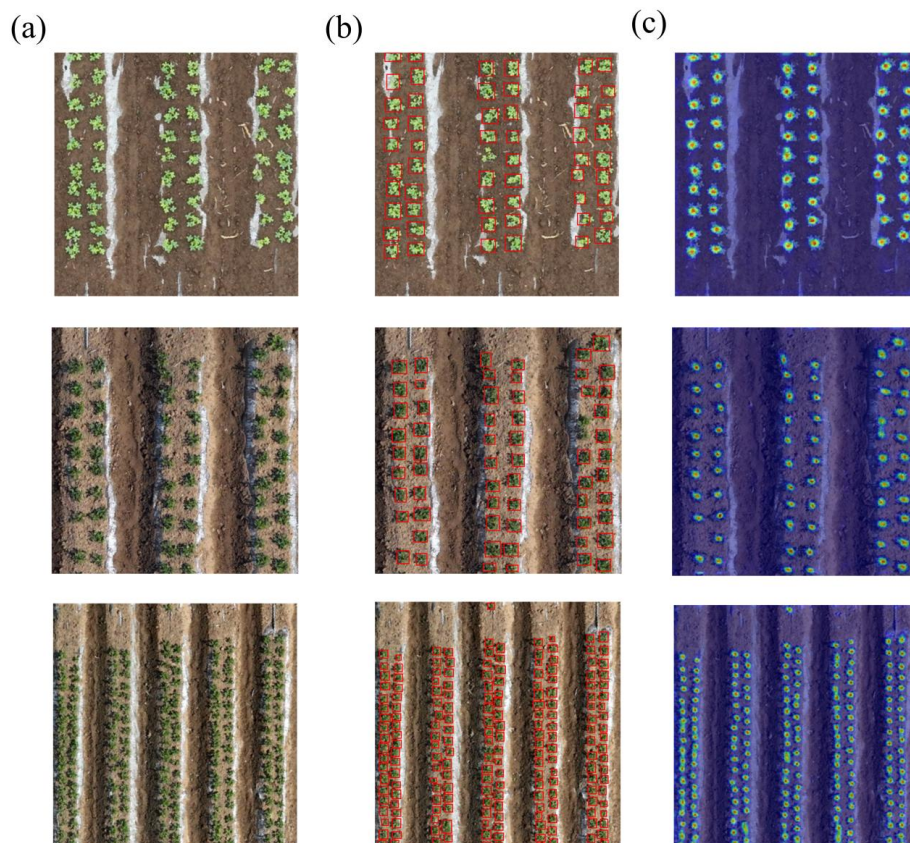


Figure 4. Peanut seedling identification performance of the model in different scenes: (a) Image to be identified; (b) Seedling identification result; (c) Seedling identification heatmap.

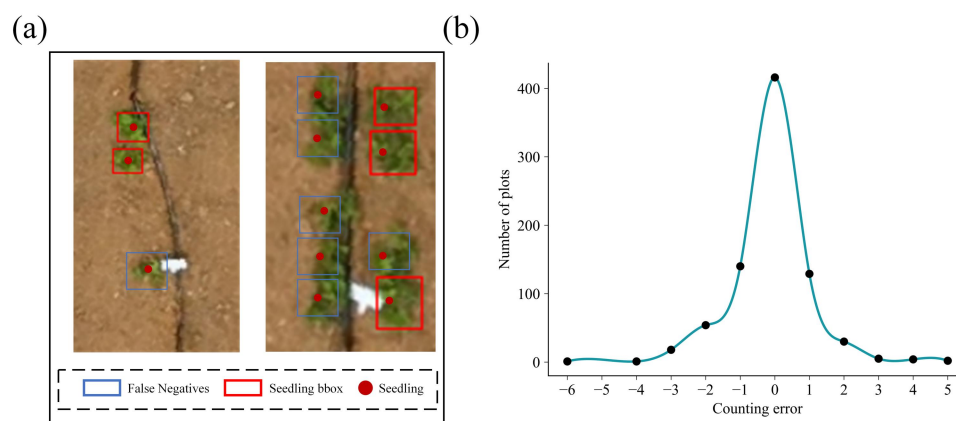


Figure 5. Examples of identification errors and curve of counting errors: (a) Examples of peanut seedling identification errors; (b) Curve of peanut seedling counting errors in the plots.

3.2 Result analysis

3.2.1 Ablation studies: We conducted ablation experiments on peanut seedling data across three different scenes to test the baseline model with and without the BIFPN feature fusion module and the Contrastive Loss module. The results indicate that both of these modules play a positive role in improving the overall performance of the model. Specifically, the BIFPN feature fusion module combines multi-dimensional feature information, allowing the final output to include both the high-dimensional features of the peanut seedlings and the low-dimensional features and positional information. This significantly enhances the model's ability to recognize and locate peanut seedlings. On the other hand, the Contrastive Loss module reduces the differences between peanut seedling features while increasing the differences between peanut seedlings and the soil background. This effectively addresses the issue of mistakenly identifying land as peanut seedlings and enhances the model's ability to recognize peanut seedlings in complex terrain conditions. Overall, the collaborative effect of the BIFPN feature fusion module and the Contrastive Loss module strengthens the stability and accuracy of the peanut seedling recognition model, proving their importance in optimizing the baseline model.

Quantitative evaluation: According to the results, the individual addition of the Contrastive Loss module and BIFPN to each baseline shows excellent performance across different datasets, especially when the BIFPN module is combined with the Hourglass backbone network, which results in an AP improvement of approximately 1.57%. When the Contrastive Loss module is added to the BIFPN-Hourglass model, the model achieves optimal peanut seedling recognition accuracy and stability. In Scene 1, the improved network shows an AP and F1-Score increase of 2.73% and 1.91%, respectively, compared to the original network. The experimental results demonstrate that the improved model has excellent peanut seedling recognition capability, which can be attributed to the combination of the feature fusion module and contrastive learning.

3.2.2 Qualitative evaluation: To provide a more intuitive demonstration of the model's performance, we randomly selected one image from the test set of each of the first three scenes and applied the trained model for detection. As shown in Figure 3, it can be observed that in all scenes, every peanut seedling was accurately identified, with each red box representing a successfully recognized seedling. In the peanut seedling recognition heatmap, the color gradually changes from yellow-green to red as it moves from the edge of the seedling to the center, while the ground surface is shown in blue. This trend in the heatmap indicates that in the model's recognition results, the confidence score for the center of the peanut seedlings is close to one. These results demonstrate that the improved Centernet model has a high confidence score for identifying peanut seedlings and performs excellently across different scenes.

3.2.3 Evaluate the seedling emergence: To verify the model's ability to identify seedlings in real-world complex terrain and its application performance, we applied the model to peanut seedlings in Scene 4, which was not involved in the training process. After testing 800 plots, it was found that the model accurately predicted the number of peanut seedlings in 414 plots (51.75%), while in 683 plots (85.375%), the predicted

number of seedlings was within one of the manual count. Only 8 plots (1%) had a prediction error greater than three seedlings. As shown in Figure 5, the statistical errors of peanut seedlings in the plots were plotted as a curve, which generally conformed to a normal distribution with zero as the mean. Finally, we selected two peanut seedling plots with typical identification errors. The main reasons for the errors were as follows: Compared to the other clearly visible peanut seedlings, the ones that were not identified are slightly blurred, leading to the occurrence of false negatives. This issue can be mitigated by appropriately adjusting the confidence interval.

4. Discussion

Centernet is an anchor-free detector that can accurately and rapidly identify and count peanut seedlings. However, due to complex soil backgrounds and weather variations, the detection performance tends to decline. Therefore, we propose a series of improved methods for detecting and counting peanut seedlings in different scenarios, which will be discussed below.

4.1 The effect of BIFPN on detection results

We incorporate BIFPN to enhance the Hourglass model's ability to extract multi-scale features of peanut seedlings, modifying the original feature map output. By this weighted feature fusion approach, the output feature map combines the advantages of both low-level and high-level features, contributing to the improvement of the model's accuracy. Compared to the baseline, BIFPN results in an AP improvement of 0.56%, 1.85%, and 0.44% in three different scenes, respectively. The improvement in the model's ability to recognize peanut seedlings means that we can set a lower confidence threshold during detection, which, to some extent, reduces the occurrence of misidentifying weeds as peanut seedlings.

4.2 The effect of Contrastive Loss on detection results

To improve the model's ability to recognize peanut seedlings in complex terrains, we adopted a weighted approach to combine Contrastive Loss with the original loss function. Contrastive Loss increases the difference between positive and negative samples while reducing the difference between identical samples, thereby enhancing the model's ability to recognize peanut seedlings. The results indicate that compared to the baseline, Contrastive Loss brought an AP improvement of 2.29%, 1.51%, and 0.93% in three different scenes, respectively. To verify whether Contrastive Loss also has similar advantages when combined with other backbone networks, we incorporated Contrastive Loss with ResNet50. The results showed that the AP improved by 0.19% and 1.84% in scene one and scene two.

5. Conclusion

We propose a method for detecting the seedling emergence rate of peanut seedlings based on the Centernet model in deep learning. The proposed method has achieved state-of-the-art results, with APs of 92.72%, 96.08%, and 95.37% in the three scenes. In scene four, the proposed method predicted the number of peanut seedlings in 85.38% of the plots with an error of one or fewer compared to the manual result. The reason for achieving such results is the incorporation of BIFPN into our Centernet network and the combination of the original loss function with Contrastive Loss. This design not only enables high-precision peanut seedling detection in complex ground surface conditions but also allows the detection of peanut seedlings at different scales. In the future, we will explore the

feature relationships between seedlings in different scenes and strive to achieve seedling statistics under various complex surface conditions with minimal annotation costs by integrating existing detection models.

Acknowledgement

This work was jointly funded by the Shandong Provincial Key Research and Development Program (grant numbers 2022LZGC021 and 2021LZGC026), the Higher Education Institutions Youth Innovation and Science & Technology Support Program of Shandong Province under Grant 2024KJH062.

References

- Abdin O, Zhou X, Cloutier D, et al. 2000. Cover crops and interrow tillage for weed control in short season maize (*Zea mays*). *European Journal of Agronomy* [J], 12: 93-102.
- Al-Mulla Y A, Huggins D R, Stöckle C O 2014. Modeling the emergence of winter wheat in response to soil temperature, water potential, and planting depth. *Transactions of the ASABE* [J], 57: 761-775.
- Andvaag E, Krys K, Shirtliffe S J, et al. 2024. Counting Canola: Toward Generalizable Aerial Plant Detection Models. *Plant Phenomics* [J], 6: 0268.
- Bai Y, Nie C, Wang H, et al. 2022. A fast and robust method for plant count in sunflower and maize at different seedling stages using high-resolution UAV RGB imagery. *Precision Agriculture* [J], 23: 1720-1742.
- Cui J, Zheng H, Zeng Z, et al. 2023. Real-time missing seedling counting in paddy fields based on lightweight network and tracking-by-detection algorithm. *Computers and Electronics in Agriculture* [J], 212: 108045.
- Gao F, Fu L, Zhang X, et al. 2020. Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN. *Computers and Electronics in Agriculture* [J], 176: 105634.
- Genaev M A, Komyshev E G, Smirnov N V, et al. 2019. Morphometry of the wheat spike by analyzing 2D images. *Agronomy* [J], 9: 390.
- Gnädinger F, Schmidhalter U 2017. Digital counts of maize plants by unmanned aerial vehicles (UAVs). *Remote sensing* [J], 9: 544.
- Götz S, Bernhardt H 2010. Produktionsvergleich von Gleichstandsamt und Normalsaat bei Silomais. *Landtechnik* [J], 65: 107-110.
- Halstead M, Mccool C, Denman S, et al. 2018. Fruit quantity and ripeness estimation using a robotic vision system. *IEEE robotics and automation LETTERS* [J], 3: 2995-3002.
- Jidong L, De-An Z, Wei J, et al. 2016. Recognition of apple fruit in natural environment. *Optik* [J], 127: 1354-1362.
- Jin X, Li Z, Yang G, et al. 2017. Winter wheat yield estimation based on multi-source medium resolution optical and radar imaging data and the AquaCrop model using the particle swarm optimization algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing* [J], 126: 24-37.
- Kuzdraliński A, Kot A, Szczerba H, et al. 2017. A review of conventional PCR assays for the detection of selected phytopathogens of wheat. *Journal of molecular microbiology and biotechnology* [J], 27: 175-189.
- Li Y, Hu P, Liu Z, et al. Contrastive clustering[C]//Proceedings of the AAAI conference on artificial intelligence.2021:8547-8555.
- Lin Z, Guo W 2021. Cotton stand counting from unmanned aerial system imagery using mobilenet and centernet deep learning models. *Remote sensing* [J], 13: 2822.
- Lindstrom M, Papendick R, Koehler F 1976. A model to predict winter wheat emergence as affected by soil temperature, water potential, and depth of planting 1. *Agronomy Journal* [J], 68: 137-141.
- Liu X, Zhao D, Jia W, et al. 2019. A detection method for apple fruits based on color and shape features. *IEEE Access* [J], 7: 67923-67933.
- Nielsen M A. Neural networks and deep learning[C]//:Determination press San Francisco, CA, USA,2015
- Osco L P, De Arruda M d S, Gonçalves D N, et al. 2021. A CNN approach to simultaneously count plants and detect plantation-rows from UAV imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* [J], 174: 1-17.
- Tan C, Li C, He D, et al. 2022. Towards real-time tracking and counting of seedlings with a one-stage detector and optical flow. *Computers and Electronics in Agriculture* [J], 193: 106683.
- Zhao B, Zhang J, Yang C, et al. 2018. Rapeseed seedling stand counting and seeding performance evaluation at two early growth stages based on unmanned aerial vehicle imagery. *Frontiers in Plant Science* [J], 9: 1362.
- Zou K, Ge L, Zhang C, et al. 2019. Broccoli seedling segmentation based on support vector machine combined with color texture features. *IEEE Access* [J], 7: 168565-168574.