Weed Classification with Drone Image Using DINOV2

Jinghong Xu, Sanlin Wang and Xinyou Dong *

College of Information Technology, Shanghai JianQiao University, China

* Corresponding Author Email: xinyoudong@gench.edu.cn

Abstract. Increasing interest in efficiency and intelligence of agriculture is driven by the growth of global population and food demand. Weed infestation is a critical factor restraining the soybean growth. Traditional weed management solutions are often inefficient as well as harmful to the environment. In this study, soybean fields weed recognition system based on DINOv2 is designed and implemented. The combination of drones captured images through all stages in Soybean lifecycle enabled precise classification and detection between soybean and weeds. The paper uses unsupervised learning and applies data augmentation methods including multi-scale cropping, color jitter, random occlusion for better generalization. Experimental results show that DINOv2 can achieve remarkable results on feature extraction and precision, while YOLOv5 is more competitive in terms of real-time efficiency with a lightweight design. This work explores how deep learning can be applied in agricultural applications and offers optimization strategies for precision agriculture. The findings contribute to reducing the application of pesticide, boosting soybean yield and contributing to sustainable agriculture.

Keywords: DINOv2, Precision Agriculture, Weed Detection, Drone Imagery, Deep Learning Models.

1. Introduction

The rapid increase of the global population has driven the development of efficient and intelligent agriculture, which has become a focus of research attention. Modern agriculture [1] relies not only on traditional cultivation techniques, but also on advanced equipment with big data technology. For instance, the integration of drones and deep learning can implement the informatization, intelligence and precision of agricultural production management. Drones quickly cruise around the fields to capture images to analyze the location of weeds in real time. Soybean is a critical food crop, and its yield is related to human survival and security. However, weeds violation is one of the major factors affecting soybean growth. Traditional weed management mainly depends on manual identification and chemical pesticide application, which is time-consuming, labor-intensive, and environmentally damaging. The deployment of UAV-based technologies has automated weeds detection in soybean fields.

2. Usage In Agriculture

2.1. Application of Deep Learning Models

In the field of crop object segmentation, deep learning models and optimization strategies play a pivotal role in processing image complexity and environment variability. Convolutional Neural Networks (CNNs) [1] and Vision Transformers (ViTs) [2] are two kinds of models. The implicit feature refining method [3] combines multi-level characteristics extracted by the attention mechanism to enhance the accuracy of edge detection. And uses feature fusion technology to detect objects, such as weeds in crop fields. However, this solution remains constrained by its high computational requirements.

Additionally, paper [4] proposes an instance segmentation model in this field which is suitable for large-scale drone image analysis. It can achieve high-precision crop detection under complex background. This research accelerates the application of drones in precision agriculture. However, this method which relies on large amount of labeled data needs to balance model complexity and hardware performance in complex scenario.

Based on the aforementioned methodology, paper [5] makes a new breakthrough in 3D leaf instance segmentation using an unsupervised pre-training model. The algorithm extracts self-attention feature to enhance its ability to recognize subtle structures and reduce dependence on labeled data. It significantly improves the model's generalization and adaptability, but the high image data collection cost is still the main limitation of real-time application.

Unmanned aerial vehicles (UAVs) [6] have become an effective tool in precision agriculture due to their data collection capabilities which are widely used in farmland monitoring and crop management. Using images taken by drones with deep learning methods can quickly detect crops and weeds, improve field management efficiency, reduce pesticide use, and increase crop yield and quality. The study [4] also shows that combined drone image with AI technology can achieve real-time detection.

2.2. 3D Modeling and Data Augmentation

With the development of precision agriculture, the significance of 3D modeling and data augmentation in crops full life cycle management has become increasingly prominent. From seed to harvest, data collection, analysis and application at each stage have promoted the development of intelligent agriculture. The studies [7, 8] demonstrate the potential of 3D reconstruction technology combined with instance segmentation in monitoring crops growth conditions. It provides comprehensive crops phenotypic information which supports real-time monitoring of crop growth. However, the high computational cost of this technology is the major limitation [5].

Moreover, data augmentation is another key point to improve model generalization. Paper [9] proposes a color calibration algorithm based on semantic style transfer, which can improve the model stability in natural environments, such as diurnal changes and various weather conditions. Paper [10] adopts augmentation strategies for random cropping, rotation, and color perturbation to further improve the robustness of the model.

2.3. Real-Time Performance and Lightweight Models in Agricultural Applications

Real-time performance is an important factor for intelligent agricultural applications. Fovea Mask [8] has shown the real-time detection advantages in green fruit instance segmentation; however, its performance can be further enhanced under complex background and non-uniform illumination conditions. The DCSN model [10] employs a lightweighted network to improve processing efficiency. Additionally, other techniques can further reduce computational requirements, such as model pruning, quantization, and knowledge distillation.

The application of deep learning technology in agriculture, particularly in the field of instance segmentation, has paved the way for agricultural intelligence. Integrated with 3D modeling, data augmentation and lightweight models can improve detection accuracy for real-time processing. Furthermore, the adoption of drones has enabled technologies mentioned above to be used in the monitoring and management of crops, such as soybean.

3. Datasets and Processing

3.1. Soybean Lifecycle Datasets

Deep Weeds [12] covers various environmental conditions and complex backgrounds, which helps the model improve its generalization. However, the dataset only contains 8 species of weeds. Weed25 [13] includes 25 species of weeds, which increases the diversity of the dataset and helps improve the model ability to recognize weeds. However, the dataset [13] still needs to add more data of diversity backgrounds and lighting complexity. Soy Images Dataset [14] is designed for soybean pest detection. It contains a variety of environmental conditions making it suitable for detection and classification tasks. But this dataset [14] only contains two categories of pests, which is limited to pest identification tasks.

The soybean full-cycle image dataset used in our study [15, 16] was collected by an UAV equipped with a high-definition camera which flies at an altitude of 30 to 50 meters and covers the whole region of the soybean. Images [15] regularly captured by drone with fixed flight path and period cross full-cycle stages of soybean growth. In addition, it is necessary to ensure that the flying parameters, such as flight speed, acquiring angle, and illumination conditions, remain consistent to ensure the comparability samples. These images not only show the detail feature of soybean, but also record the distribution information of weeds [17] in the field.

3.2. Data Augmentation

The DINOv2 [18] demonstrates feature extraction capabilities. To optimize it in soybean and weeds recognition, this paper uses the following image augmentation techniques:

- Multi-scale cropping and padding: soybeans and weeds vary significantly in size and location. Through this method, the model can learn multi-scale features;
- Color jittering: The lighting conditions are unstable in field, and the color information is easily affected by the environment. Through this method, the model can maintain stable recognition performance under varying illumination conditions;
- Random occlusion (Cutout): Soybeans in the field cause shading issues due to height differences. Through this method, the model can identify weeds even in the case of partial occlusion;

Finally, our paper adopts a strategy of mixing original images with augmented images for avoid data distribution deviation caused by excessive augmentation. By maintaining a configured ratio between original and augmented images, our algorithm guarantees that the training data is both diverse and realistic.

4. Neural Network Architectures

4.1. Overview of Model Architectures

DINOv2 [18] is an unsupervised visual feature learning model based on the Vision Transformer (ViT) which is pre-trained to extract robust visual features. The essence of DINOv2 is a multi-layer self-attention mechanism and an unsupervised learning framework which enables the model to learn intricate details and complex structures from images without labeled data. The model can detect objects with different sizes in high-dimensional space by its novel feature extraction algorithm. It lays the foundation for instance segmentation and complex image analysis.

YOLOv5 [19] is an object detection model which combines feature pyramid network (FPN) and path aggregation network (PAN) to improve the multi-scale features fusion capability. The model adopts lightweight architecture which reduces the computational costs. And it is an optimal solution for real-time detection tasks. YOLOv5 introduces the anchor box mechanism and optimizes the loss function that significantly improves the model’s calculated speed and accuracy. YOLOv8 has made many advancements based on YOLOv5, such as dynamic label allocation, adaptive anchor boxes, etc. Moreover, new loss function further optimizes the regression and classification weight distribution which increases the performance of the model in small object detection.

4.2. Advantages and Disadvantages Analysis

The unsupervised learning framework of DINOv2 greatly reduces the dependence on large-scale annotated data. It has tremendous practical value in areas where annotation is highly expensive. The self-attention mechanism can extract subtle features in complex backgrounds and is suitable for diverse applications, for example, semantic segmentation and object detection. Additionally, multi-scale features enhance the flexibility of the model for objects with different sizes.

The main benefit of YOLOv5 is its detection speed and low latency characteristics that make it well-suited for real-time tasks. The model’s lightweight design makes YOLOv5 easier to deploy on edge devices, while multi-scale features fusion improves its detection performance of small targets. But, YOLOv5 performs worse than Transformer-based models when handling complex backgrounds issues. In addition, it requires data augmentation strategies to improve robustness on large datasets. YOLOv8 introduces dynamic label allocation, adaptive anchor boxes, optimizing the training process and new loss functions, which notably increase accuracy and detection capabilities compared to YOLOv5. These improvements make YOLOv8 more adaptable for small target detection in various environments meanwhile increase training complexity and requiring more hyperparameters fine-tuning to achieve optimal performance.

Although YOLO series are expert in real-time detection, DINOv2 performs better in high-precision tasks due to its self-attention mechanism. In agricultural applications, DINOv2 can achieve reliable crop and weed classification with limited labeled data.

5. Analysis and Findings

5.1. Evaluation

This research evaluates the performance of two models on the object recognition task: DINOv2 [18] and YOLO [19]. The experiment initializes with their respective pre-trained weights to train and test on 1024×1024-pixel images.

5.2. Analysis of Model Performance

5.2.1. Performance Comparison: YOLOv5 vs. YOLOv8

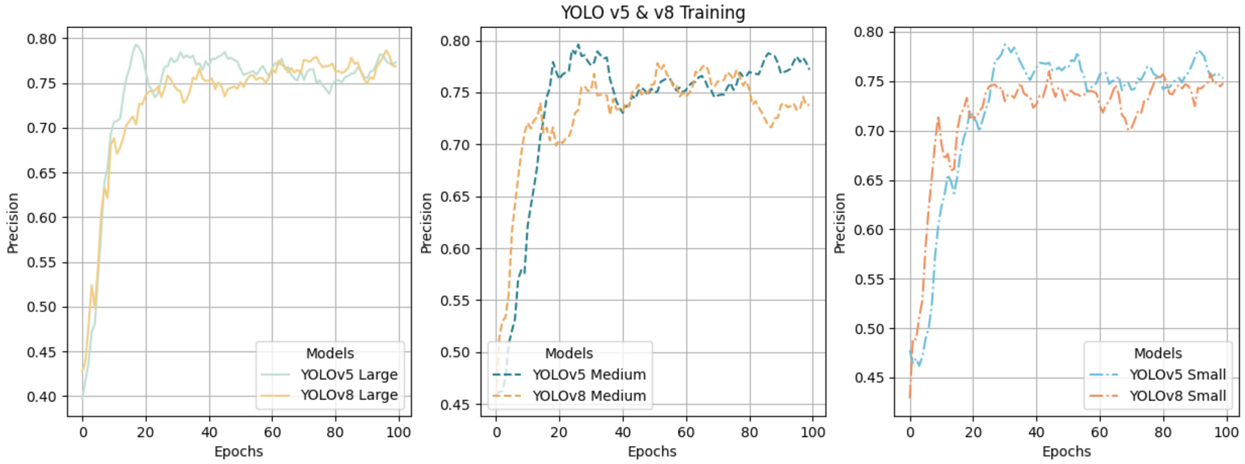

Figure 1. Performance comparison of large, medium, and small models within the YOLOv5 and YOLOv8 architectures for object detection task

As shown in Fig 1, YOLOv5 achieves faster convergence and demonstrates more stable compared to YOLOv8. The results show that YOLOv5 performs better than YOLOv8 in soybean object detection task.

5.2.2. Performance Comparison: DINOv2 vs. YOLOv5

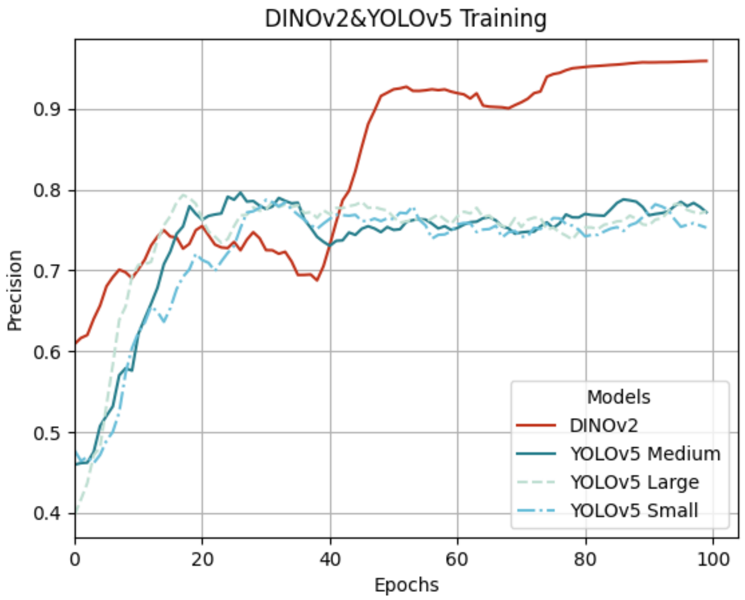

Figure 2. Performance comparison of large, medium, and small models within the YOLOv5 and YOLOv8 architectures for object detection task

As shown in Fig 2, DINOv2 achieves faster convergence and greater stability compared to YOLOv5. The experimental findings indicate that DINOv2 achieves better performance than YOLOv5 in soybean object detection task. And also, the precision of DINOv2 is much better than YOLOv5 from the Table 1 below.

Table 1. Precision comparison of DINOv2 and YOLO in object detection task

Architecture	Image Size	Evaluation Precision
DINOv2	1024×1024	0.961
YOLOv5 Large	1024×1024	0.821
YOLOv5 Medium	1024×1024	0.827
YOLOv5 Small	1024×1024	0.82

The experiments use the AdamW optimizer, train for 100 epochs, and set the batch size to 6. For reproducibility of the results, a reasonable random initialization is applied. In each experiment, 5 different random seeds are used. For each random seed, the training set is divided into 80% for training and 20% for validation. This study uses an early stopping mechanism with a patience parameter of 50 epochs. When the performance improvement plateaus, the training procedure will stop automatically.

5.3. Results Analysis

Experimental findings indicate that the DINOv2 model performs well in soybean object detection task, mainly due to its advanced architecture of unsupervised learning. DINOv2's excellent classification and segmentation capabilities not only improve object detection accuracy, but also provide support for crops monitoring in precision agriculture.

However, the results also revealed some potential problems. For example, some certain lighting conditions or plant features are similar, the model's misclassification rate may increase. These challenges lead to specific directions for optimization:

- Improve data augmentation strategies: Add more diverse illuminated images and plant feature samples to improve the model's adaptability;
- Introduce multimodal data fusion: Combine RGB and multispectral data to enhance the model’s generalization ability.

A key observation is DINOv2's faster training convergence which indicates that its significant advantage is efficient optimization. The fast convergence feature saves computing resources, meanwhile improves the usability of the model. Future research can further explore the application of optimization and multimodal data and expand its application in precision agriculture.

6. Conclusion

This study evaluates the performance of DINOv2 and YOLO series, in soybean and weed object detection task. According to experimental results, DINOv2 was finally determined to be the optimal model. The research results indicate:

- Exceptional Performance: DINOv2’s advanced architecture of unsupervised learning achieves excellent classification and segmentation capabilities, and performs outstandingly in crop monitoring tasks in precision agriculture;
- Efficient Training: DINOv2’s training convergence and optimization efficiency is significantly higher than that of the YOLO series;
- Challenges and Future Enhancements: Although DINOv2 has excellent performance, it still has misclassification problems under specific lighting conditions or when plant features are similar. In future, the robustness can be further improved by optimizing data augmentation strategies, and combining multimodal data.

In summary, DINOv2 provides important support for object detection tasks in precision agriculture due to its high accuracy and efficiency. Future research will be focused on optimizing model performance and expanding its application areas in intelligent agriculture.

References

[1] Kujawa S, Niedbała G. Artificial neural networks in agriculture [M]. 2021: 497.

[2] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C] // Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 5998 - 6008.

[3] Ma L, Wang T, Dong B, et al. Implicit feature refinement for instance segmentation [EB/OL]. arXiv, 2021.

[4] Weyler J, Quakernack J, Lottes P, et al. Joint plant and leaf instance segmentation on field-scale UAV imagery [J]. IEEE Robotics and Automation Letters, 2022.

[5] Li D, Li J, Xiang S, et al. PSegNet: Simultaneous semantic and instance segmentation for point clouds of plants [J]. Plant Phenomics, 2022.

[6] Singh D, Makanjuola WO, Misra A, et al. A review on UAV-based applications for precision agriculture [J]. Information, 2021, 10 (11): 349.

[7] Ni X, Li C, Jiang H, et al. Three-dimensional photogrammetry with deep learning instance segmentation to extract berry fruit harvestability traits [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021.

[8] Jia W, Zhang Z, Shao W, et al. FoveaMask: A fast and accurate deep learning model for green fruit instance segmentation [J]. Computers and Electronics in Agriculture, 2021.

[9] Huang H, Yang A, Tang Y, et al. Deep color calibration for UAV imagery in crop monitoring using semantic style transfer with local to global attention [J]. International Journal of Applied Earth Observation and Geoinformation, 2021.

[10] Li X, Zhou Y, Liu J, et al. The detection method of potato foliage diseases in complex background based on instance segmentation and semantic segmentation [J]. Frontiers in Plant Science, 2022.

[11] Pan F, Wang F. DCSN: A flexible and efficient lightweight network for dense cell segmentation [C] // Proceedings of the 2023 International Medical AI Conference. 2023.

[12] Olsen A, Konovalov DA, Philippa B, et al. DeepWeeds: A multiclass weed species image dataset for deep learning [J]. Scientific Reports, 2019, 9 (1): 2058.

[13] Wang P, Tang Y, Luo F, et al. Weed25: A deep learning dataset for weed identification [J]. Frontiers in Plant Science, 2022, 13: 1053329.

[14] Mignoni ME, Honorato A, Kunst R, et al. Soybean images dataset for caterpillar and Diabrotica speciosa pest detection and classification [J]. Data in Brief, 2022, 40: 107756.

[15] RoboFlow Universe. Soy-Weed-Seg Dataset [EB/OL]. [2024-11-23]. Available at: https://universe.roboflow.com/weeds-d0as0/soy-weed-seg.

[16] Smith J, Johnson A, Williams R, et al. From Seedling to Harvest: The Growing Soy Dataset for Weed Detection in Soy Crops via Instance Segmentation [J]. Agricultural Informatics and Technology, 2023, 45 (3): 123 - 134.

[17] Green P, Taylor M, Rodriguez L, et al. Transforming weed management in sustainable agriculture with artificial intelligence: A systematic literature review towards weed identification and deep learning [J]. Sustainable Agriculture Reviews, 2023, 18 (2): 245 - 260.

[18] Oquab M, Darcet T, Moutakanni T, et al. Dinov2: Learning robust visual features without supervision [EB/OL]. arXiv, 2023.

[19] Ultralytics. Ultralytics - YOLO[EB/OL]. [2024-11-23]. Available at: https://github.com/ultralytics/ultralytics.