A Weakly-Supervised Learning Approach for RGB Crop Detection Using UAV Imagery

Anonymous ECCV 2024 Submission

Paper ID #20

Abstract. This paper presents a novel weakly-supervised learning approach for crop detection in precision agriculture, leveraging RGB imagery captured via drones. The method integrates the Segment Anything model for zero-shot segmentation, DBSCAN for clustering, and Faster R-CNN with a ResNet101 backbone for object detection. The approach addresses the challenges of limited labeled data and the need for cost-effective solutions in agricultural settings. The model's performance was evaluated on a newly proposed dataset containing RGB images of vineyards, orchards, olive groves, and wheat fields. The results demonstrated high precision, recall, and F1 scores across crop types, validating the model's effectiveness in real-world scenarios. This research highlights the potential of advanced machine learning techniques to enhance crop monitoring and management, ultimately contributing to more sustainable and productive agricultural practices.

Keywords: Precision Agriculture, Weakly-Supervised Learning, Crop Detection, Segment Anything, Drones

1 Introduction

Precision agriculture represents a transformative approach to farming, leveraging advanced technologies to enhance crop production and sustainability [5, 22]. Precision agriculture addresses agricultural fields' spatial and temporal variability by integrating GPS, remote sensing, and data analytics, optimizing resource usage and maximizing yield potential. This innovative approach is crucial for meeting the growing global food demand while minimizing the environmental impact of agriculture.

Recent advances in Artificial Intelligence have revolutionized decision-making processes in precision agriculture [6]. These technologies analyze large datasets to provide predictive analytics, enabling real-time monitoring and management of crops and soil through Internet of Things devices. Despite these advancements, the high cost of technology and the need for specialized knowledge pose significant challenges to widespread adoption. Furthermore, existing approaches often rely on multispectral or hyperspectral imagery, which can be costly and inaccessible to many farmers. The scarcity of labeled data for training models further limits the applicability of these technologies in diverse agricultural settings.

Unmanned Aerial Vehicles (UAVs), commonly known as drones, have emerged as a vital tool in precision agriculture, offering a versatile platform for capturing high-resolution imagery of agricultural fields [19, 24]. With advanced sensors and imaging technologies, UAVs facilitate detailed monitoring of crop health, soil conditions, and environmental factors. Their ability to provide timely and precise data enables farmers to make informed decisions on resource allocation, pest management, and irrigation, ultimately enhancing crop productivity and sustainability. However, integrating UAV data into effective agricultural practices still faces challenges, particularly in data processing and analysis, which often require sophisticated algorithms and substantial computational resources.

In response to these challenges, this paper presents a novel weakly-supervised learning approach for crop detection with drones. *Weakly-supervised learning*, which utilizes limited labeled data to infer accurate models, offers a practical solution in scenarios where fully labeled datasets are unavailable [28]. Our method combines the Segment Anything model [15] for zero-shot segmentation, DBSCAN [9] for clustering, and Faster R-CNN [20] with a ResNet101 backbone [13] for object detection. The key advantage of this innovative approach is the use of RGB imagery, which is more accessible and cost-effective than multispectral data, making it practical and more affordable for farmers. Moreover, unlike traditional models, our method does not rely on labeled data for initial training. Instead, it generates labels through zero-shot segmentation, significantly enhancing the model's adaptability and scalability in real-world agricultural applications, where labeled data is often scarce or unavailable. This capability allows our method to address the limitations of existing work by providing an efficient and cost-effective solution for accurate crop detection and monitoring.

We evaluate our method using a newly proposed dataset, which includes diverse agricultural fields captured via drone imagery. Our experiments demonstrate the model's effectiveness in accurately detecting and classifying various crop types, even under different field conditions.

The rest of this paper is organized as follows. Section 2 reviews related work in precision agriculture, focusing on UAV applications and deep learning methods. Section 3 details our proposed approach, including the model architecture and training procedure. Section 4 presents the dataset description, experimental setup, and results. Finally, Section 5 concludes the paper and outlines potential directions for future research.

2 Related Work

Developments have notably impacted advances in precision agriculture through computer vision and remote sensing techniques. Unmanned Aerial Vehicles have become essential tools in modern farming, providing capabilities for detailed crop health monitoring and precise agricultural land mapping [16]. The adoption of UAVs in precision agriculture has been driven by their ability to enhance crop quality and reduce health hazards associated with manual pesticide application. With advanced sensors and imaging technologies, UAVs offer solutions for crop

monitoring, height estimations, pesticide spraying, and soil analysis [18]. UAVs typically consist of several vital components, including an airframe, propulsion system, power supply, control system, navigation system, and payload (e.g., cameras and sensors), which work together to perform various tasks efficiently [1, 7].

Deep learning algorithms have significantly advanced precision agriculture by leveraging data from UAVs and ground sensors. Convolutional Neural Networks (CNNs) have been particularly effective in processing imagery data to detect plant diseases, pest infestations, and nutrient deficiencies [2, 21]. These algorithms enable precise monitoring and management of crops by analyzing images, which is essential for early intervention and optimal resource allocation [17, 25, 27]. While traditional machine learning methods, such as Support Vector Machines and Decision Trees, have been fundamental, they often fall short in scalability and feature extraction. Deep learning, on the other hand, offers superior capabilities by automatically extracting detailed features from large datasets, thereby enhancing agricultural decision-making systems.

UAVs provide high-resolution, timely data for detecting subtle distinctions in crop conditions, an area where satellite imagery often falls short due to resolution limitations. Integrating UAV data with deep learning models, such as CNNs, enables precise crop and plant segmentation, improving agricultural interventions [3, 4, 11, 12]. Recent advancements include two-stage detectors like Faster R-CNN [20] and one-stage detectors like YOLO [14], which balance accuracy and operational speed for real-time applications. These models can excel in detecting various crops and plants from UAV imagery, significantly enhancing the precision of agricultural monitoring.

Our research builds on these advancements by integrating state-of-the-art machine learning techniques into a weakly-supervised learning model for crop detection. A significant advantage of our approach is its independence from labeled data for initial training. The model generates "labels" through zero-shot segmentation, making it adaptable and applicable to various types of imagery without requiring pre-existing annotations. Additionally, our approach leverages RGB imagery, which is more accessible and less costly than multispectral data. This practical and affordable solution enhances scalability and usability in real-world agricultural applications, where labeled data is often scarce or unavailable.

3 Method

The methodology developed for this research, illustrated in Fig. 1, addresses the challenges of unsupervised crop detection by integrating three advanced techniques: zero-shot segmentation, clustering, and object detection. Given the scarcity of labeled data, a weakly supervised approach was necessary. Combining these techniques forms a robust pipeline capable of generating and refining *pseudo*-ground truth data for crop detection.

The first component of our pipeline is the Segment Anything model from Meta [15], which employs a Vision Transformer (ViT-B) [8]. SAM is renowned for handling complex image data without requiring labeled training data, which

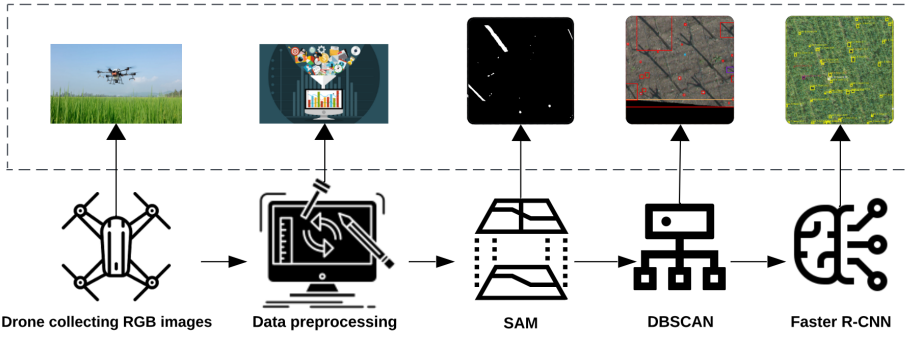

Fig. 1: Illustration of the comprehensive pipeline employed in the study, showcasing the sequential integration of three main components: SAM for initial zero-shot segmentation, DBSCAN for grouping segmented features, and Faster R-CNN with a ResNet101 backbone for precise object detection. The process begins with a drone capturing RGB images of agricultural fields, followed by preprocessing steps such as image tiling and pixel normalization, setting the stage for detailed crop type classification.

is particularly valuable for the agricultural domain, where labeled datasets are often scarce. The zero-shot capability of SAM allows it to identify potential crop regions in RGB images, making it adaptable to new or unseen crop types. This adaptability is crucial for the dynamic and varied environments typical in agriculture, where crop types and their appearances can significantly vary.

Following the initial segmentation by SAM, the next step involves clustering the detected features. The features considered in this process include the size, color, and texture of bounding boxes detected by SAM. We employ the DBSCAN algorithm [10], enhanced with PCA for dimensionality reduction [23]. DBSCAN is particularly suitable for our application because it excels in identifying clusters of varying densities, which is common in agricultural fields. PCA reduces the feature space, making the clustering process more efficient and focused on the most significant features. DBSCAN can handle complex and irregular patterns without requiring a predefined number of clusters, making it ideal for agricultural data's diverse and heterogeneous nature.

In other words, the SAM model generates the pseudo-ground truth by performing zero-shot segmentation to identify potential crop regions, followed by DBSCAN clustering to group these regions into coherent clusters. This process creates a set of labeled regions that serve as the foundation for training an object detection model. This weakly supervised approach circumvents the need for manually labeled data.

Once the clustering is complete, the next phase involves precise object detection using the Faster R-CNN framework [20] with a ResNet101 backbone [13]. Faster R-CNN is well-known for its robust performance in detecting and classifying objects within images, particularly in complex and cluttered environments like agricultural fields. Its region proposal network effectively identifies regions of interest, which are then refined and classified by the network. This capability

is crucial for accurately detecting crops that other plants or objects may partially occlude. The ResNet101 backbone is selected for its deep feature extraction capabilities. With its 101 layers, ResNet101 can capture meticulous differences between various crop types, providing a rich and detailed representation of the input images. This depth allows the model to learn fine-grained features essential for distinguishing between crops, even when the visual differences are subtle.

Integrating these components—SAM, DBSCAN, and Faster R-CNN—creates an efficient and scalable pipeline for crop detection. SAM's zero-shot segmentation provides a quick initial pass to isolate potential crop regions. DBSCAN clusters these segments based on inherent properties without needing predefined parameters. Faster R-CNN ensures precise detection and classification of crops within the clustered regions. This synergy allows for the effective handling of large-scale agricultural datasets, enhancing both the accuracy and scalability of the crop detection system.

4 Experiments

4.1 Dataset & Setting

The dataset used in this study consists of RGB images captured via drone over various agricultural fields in Foggia, Italy. This dataset includes four major crop types: vineyards, orchards, olive groves, and wheat fields, totaling 40,000 images, each sized 512×512 pixels.[1] The images were collected using a DJI AIR 2S drone, flying at an altitude of 40 meters between 9 a.m. and 1 p.m. at an estimated speed of 5 m/s. To present the dataset, Fig. 2 shows some examples of images representing different crops. To ensure robust model evaluation, the dataset was divided into three subsets: 70% for training, 15% for validation, and 15% for testing. The splitting was done *field-wise* to guarantee that tiles from the same field were not present in the training and test sets. Ground truth data was generated through zero-shot segmentation using the Segment Anything model followed by DBSCAN clustering, ensuring crop type labeling, as described in the previous section. In addition to identifying the crop types, DBSCAN identified two additional classes: rocks and weeds.

The training procedure was designed to optimize the model's performance in crop detection, involving several critical steps from data preparation to model training and validation. The comprehensive ortho-mosaic captured by the drone was subdivided into 512 × 512 pixel tiles, a crucial step for managing computational load and retaining high-resolution details. Non-crop elements, primarily roads, were removed using the LabelMe [26] annotation tool to ensure that the dataset focused exclusively on agricultural features. Additionally, the pixel values were normalized to further enhance the model's accuracy during training. The SAM model was employed in "evaluation" mode, leveraging its pre-trained capabilities to identify crop regions without additional labeled data. This step

[1] The data will be made publicly available upon acceptance.

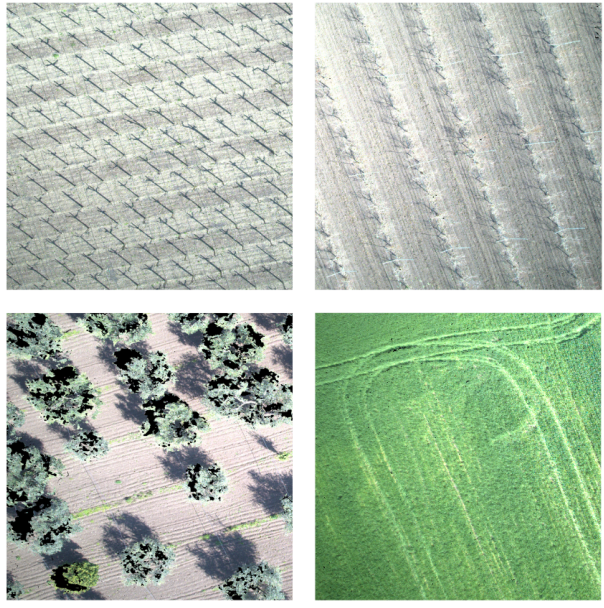

Fig. 2: Different crops from our newly proposed dataset.

took advantage of the model's robustness in zero-shot scenarios, efficiently isolating crop-related features. The DBSCAN algorithm, combined with PCA for dimensionality reduction, clustered the segmented features based on their inherent properties, effectively handling the complex patterns typical in agricultural data and grouping similar features for detailed analysis. The key hyperparameters for the experiments included an epsilon value of 0.5 and a minimum sample size of 5 for DBSCAN clustering.

The Faster R-CNN model with a ResNet101 backbone was trained with specific configurations to ensure high accuracy and efficiency. Key training configurations included using Stochastic Gradient Descent as the optimizer, with a learning rate of 0.005, momentum of 0.9, and weight decay of 0.0005. The learning rate was adjusted using `ReduceLROnPlateau` based on validation loss to ensure optimal learning efficiency. Early stopping was implemented to prevent overfitting by halting training after three epochs of no improvement in validation loss. This structured approach ensured that the model effectively learned and adapted to the diverse crop types present in the dataset, optimizing both accuracy and computational efficiency.

The experiments were conducted using an Intel Core i7 processor, 32 GB RAM, and an NVIDIA GeForce RTX 3080 Ti GPU. The software environment included PyTorch for deep learning, extensively using the `torch-vision` library for model architectures and image processing.

4.2 Results

To evaluate the model's performance in crop detection, we utilized precision, recall, and F1 score alongside Intersection over Union (IoU) to measure localization accuracy.

The proposed method combines SAM for zero-shot segmentation, DBSCAN for clustering, and Faster R-CNN with a ResNet101 backbone. Our approach shows remarkable performance in crop detection, as demonstrated in Table 1. This highlights the effectiveness of this combined approach for handling complex agricultural imagery. The baseline model, which applied DBSCAN clustering on pixel values followed by Faster R-CNN, yielded the lowest results. This baseline provided a foundation for assessing the improvements offered by our proposed approach.

To further validate the effectiveness of the primary model configuration, additional experiments were conducted using alternative combinations of neural network architectures and clustering algorithms. For instance, replacing ResNet101 with ResNet50 yielded reasonably good results, though slightly lower than the primary model. In another ablation study, DBSCAN was replaced with K-means clustering while keeping SAM and ResNet50. This configuration demonstrated lower performance metrics, indicating that DBSCAN's ability to handle irregular data patterns and varying densities is more suited for the agricultural datasets used in this study.

The results from these additional experiments, along with the baseline model, provide a comprehensive view of how different configurations impact the performance of crop detection systems. The best-performing model, utilizing SAM, DBSCAN, and Faster R-CNN with ResNet101, demonstrated superior precision, recall, and F1 scores. This highlights the model's effectiveness in handling the complexity of agricultural imagery compared to the alternatives tested.

The performance differences observed across various classes with the proposed method can be attributed to several factors. Class imbalance in the dataset, where some classes were over-represented (such as orchards) compared to others (such as wheat), likely led to the model performing better on classes with more examples during training. This imbalance could explain why the model performed worse on certain classes. Additionally, some classes exhibited higher intra-class variability, meaning the same class appeared differently under varying conditions (e.g., lighting, angle, growth stage). This variability made it more challenging for the model to correctly identify all instances of the class, leading to lower recall. Visually similar classes (e.g., vineyards and orchards) or overlapping features also contributed to errors.

To illustrate the qualitative differences between model predictions, Fig. 3 compares the detection results across different configurations. Each subfigure demonstrates the segmentation and classification output, highlighting the advantages of the SAM + DBSCAN + ResNet101 configuration in accurately detecting and classifying crops.

The analysis also reveals the advantages of SAM over direct pixel clustering, as it offers superior contextual understanding and effectively handles complex

Method	Class	Precision	Recall	F1 score
SAM + DBSCAN + ResNet101	Olive grove	0.83	0.76	0.67
	Orchard	0.80	1.00	0.91
	Vineyard	0.84	0.64	0.73
	Wheat	0.73	0.68	0.75
	Weeds	0.83	0.68	0.75
	Rocks	0.80	0.56	0.66
SAM + DBSCAN + ResNet50	Olive grove	0.75	0.69	0.61
	Orchard	0.73	0.97	0.83
	Vineyard	0.76	0.58	0.66
	Wheat	0.66	0.62	0.68
	Weeds	0.75	0.62	0.68
	Rocks	0.73	0.51	0.60
SAM + K-means + ResNet50	Olive grove	0.64	0.59	0.52
	Orchard	0.62	0.83	0.71
	Vineyard	0.65	0.49	0.56
	Wheat	0.56	0.53	0.58
	Weeds	0.64	0.53	0.58
	Rocks	0.62	0.43	0.51
DBSCAN + Faster R-CNN	Olive grove	0.53	0.46	0.50
	Orchard	0.45	0.55	0.50
	Vineyard	0.52	0.45	0.48
	Wheat	0.47	0.52	0.49
	Weeds	0.50	0.47	0.48
	Rocks	0.45	0.39	0.41

Table 1: Detection performance metrics for various model configurations and crop types at an IoU threshold of 0.5.

scenes. SAM's zero-shot segmentation captures long-range dependencies, making it more adept at segmenting overlapping crops and dense vegetation. Additionally, DBSCAN's ability to manage irregular data patterns and varying densities makes it more suitable than K-means for agricultural fields with diverse crop patterns. Finally, the deeper architecture of ResNet101 provides enhanced feature extraction, improving detection accuracy for complex visual patterns compared to ResNet50.

The integrated system, which combines segmentation, clustering, and detection, has proven to be highly effective in a real-world agricultural setting. The use of DBSCAN and ResNet101 significantly contributed to the system's success by enhancing its adaptability and accuracy, making it well-suited to handle the complexities of real-world agricultural environments.

5 Conclusion

This study implemented a robust weakly-supervised learning approach that integrates the Segment Anything model for segmentation, DBSCAN for clustering, and Faster R-CNN with a ResNet101 backbone for object detection. The results

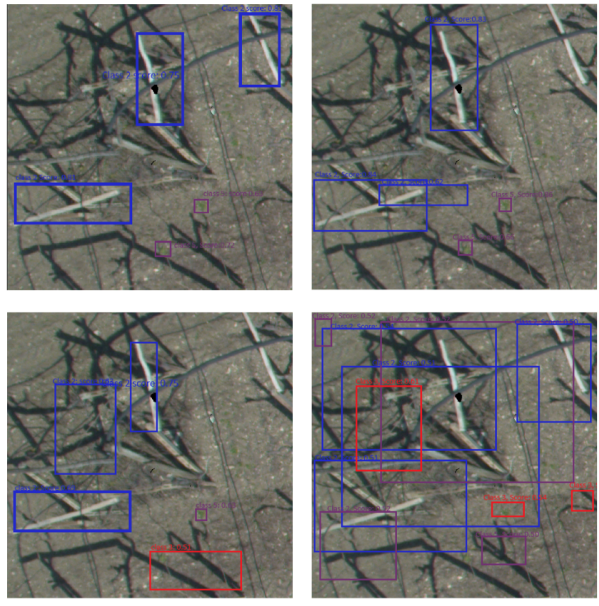

Fig. 3: Qualitative comparison between the predictions of the different models on the same tile. From top to bottom and left to right: SAM + DBSCAN + ResNet101, SAM + DBSCAN + ResNet50, SAM + K-Means + ResNet50, DBSCAN + ResNet101. The blue regions represent the orchards class, the purple regions represent the rocks class, and the red regions represent the vineyard class.

demonstrated high precision, recall, and F1 scores across different crop types, indicating the model's effectiveness in real-world agricultural settings. Integrating these advanced machine learning techniques facilitated accurate crop detection and classification, which is beneficial for enhancing agricultural productivity and management practices.

While the results are promising, it is essential to acknowledge the limitations of this study. The model's performance is validated under the specific conditions and crop types in our newly proposed dataset, which was collected using drones. Consequently, its performance may vary under different environmental conditions or with crop types not represented in the study. Nevertheless, this research marks significant progress in applying unsupervised learning techniques to precision agriculture, particularly in addressing the complexities of natural agricultural environments without requiring labeled training data.

Future research could build on this study's findings by expanding the dataset to include more diverse environmental conditions and additional crop types, thereby testing the model's robustness and adaptability. Incorporating additional environmental variables such as soil moisture, temperature, and crop health indicators could enhance the model's utility and accuracy. Additionally, exploring

other advanced machine learning and deep learning frameworks might offer improvements over the current DBSCAN and Faster R-CNN models.

In conclusion, this research underscores the potential of integrating advanced machine learning techniques in precision agriculture, paving the way for more efficient and scalable crop monitoring solutions. The application of drones for data collection has proven to be highly effective, enabling the capture of high-resolution imagery essential for accurate analysis. By addressing the challenges of labeled data scarcity and complex agricultural environments, the study contributes to the ongoing efforts to enhance agricultural practices through technology driven solutions.

References

1. Aslan, M.F., Durdu, A., Sabanci, K., Ropelewska, E., Gültekin, S.S.: A Comprehensive Survey of the Recent Studies with UAV for Precision Agriculture in Open Fields and Greenhouses. Applied Sciences **12**(3) (2022) 3
2. Bouguettaya, A., Zarzour, H., Kechida, A., Taberkit, A.M.: Deep learning techniques to classify agricultural crops through UAV imagery: A review. Neural computing and applications **34**(12), 9511–9536 (2022) 3
3. Castellano, G., De Marinis, P., Vessio, G.: Applying Knowledge Distillation to Improve Weed Mapping with Drones. In: 2023 18th Conference on Computer Science and Intelligence Systems (FedCSIS). pp. 393–400 (2023) 3
4. Castellano, G., De Marinis, P., Vessio, G.: Weed mapping in multispectral drone imagery using lightweight vision transformers. Neurocomputing **562**, 126914 (2023) 3
5. Cisternas, I., Velásquez, I., Caro, A., Rodríguez, A.: Systematic literature review of implementations of precision agriculture. Computers and Electronics in Agriculture **176**, 105626 (2020) 1
6. Coulibaly, S., Kamsu-Foguem, B., Kamissoko, D., Traore, D.: Deep learning for precision agriculture: A bibliometric analysis. Intelligent Systems with Applications **16**, 200102 (2022) 1
7. Delavarpour, N., Koparan, C., Nowatzki, J., Bajwa, S., Sun, X.: A technical study on UAV characteristics for precision agriculture applications and associated practical challenges. Remote Sensing **13**(6), 1204 (2021) 3
8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020) 3
9. Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: kdd. vol. 96, pp. 226–231 (1996) 2
10. Hahsler, M., Piekenbrock, M., Doran, D.: dbscan: Fast Density-Based Clustering with R. Journal of Statistical Software **91**(1), 1–30 (2019) 4
11. Hall, O., Dahlin, S., Marstorp, H., et al.: Classification of Maize in Complex Smallholder Farming Systems Using UAV Imagery. Drones **2**(3) (2018) 3
12. Hasan, M., Tanawala, B., Patel, K.: Deep Learning Precision Farming: Tomato Leaf Disease Detection by Transfer Learning. In: Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE) (2019) 3

13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016) 2, 4
14. Jiang, P., Ergu, D., Liu, F., Cai, Y., Ma, B.: A Review of YOLO algorithm developments. Procedia computer science **199**, 1066–1073 (2022) 3
15. 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. pp. 4015–4026 (2023) 2, 3
16. Mogili, U.R., Deepak, B.B.V.L.: Review on Application of Drone Systems in Precision Agriculture. Procedia Computer Science **133**, 502–509 (2018) 2
17. Olaniyi, E., Chen, D., Lu, Y., Huang, Y.: Generative Adversarial Networks for Image Augmentation in Agriculture: A Systematic Review. Journal of Agricultural Science and Technology **47**(3), 123–138 (2023) 3
18. Rahman, M.F.F., Fan, S., Zhang, Y., Chen, L.: A Comparative Study on Application of Unmanned Aerial Vehicle Systems in Agriculture. Agriculture **11**(1) (2021) 3
19. Rejeb, A., Abdollahi, A., Rejeb, K., Treiblmaier, H.: Drones in agriculture: A review and bibliometric analysis. Computers and electronics in agriculture **198**, 107017 (2022) 2
20. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in neural information processing systems **28** (2015) 2, 3, 4
21. Saranya, T., Deisy, C., Sridevi, S., Anbananthen, K.S.M.: A comparative study of deep learning and Internet of Things for precision agriculture. Engineering Applications of Artificial Intelligence **122**, 106034 (2023) 3
22. Sharma, A., Jain, A., Gupta, P., Chowdary, V.: Machine learning applications for precision agriculture: A comprehensive review. IEEE Access **9**, 4843–4873 (2020) 1
23. Shlens, J.: A tutorial on principal component analysis. arXiv preprint arXiv:1404.1100 (2014) 4
24. Sishodia, R.P., Ray, R.L., Singh, S.K.: Applications of remote sensing in precision agriculture: A review. Remote sensing **12**(19), 3136 (2020) 2
25. Tugrul, B., Elfatimi, E., Eryigit, R.: Convolutional Neural Networks in Detection of Plant Leaf Diseases: A Review. Journal of Agricultural Science and Technology **47**(3), 123–138 (2023) 3
26. Wada, K.: LabelMe: Image Polygonal Annotation with Python. GitHub repository (2016) 5
27. Wu, H.T.: Developing an intelligent agricultural system based on long short-term memory. Mobile Networks and Applications **26**(3), 1397–1406 (2021) 3
28. Zhou, Z.H.: A brief introduction to weakly supervised learning. National science review **5**(1), 44–53 (2018) 2