DLP353: A Benchmark Dataset for Green Fruit Detection in Natural Camouflaging Environments Using YOLOv8

Shreyan Kundu
Department of Computer Science & Engineering (AI)
Institute of Engineering & Management
University of Engineering & Management
Kolkata, India
shreyankundu@gmail.com

Nirban Roy
Department of Computer Science & Engineering
(Internet of Things and Cyber Security
including Blockchain Technology),
Institute of Engineering & Management
Kolkata, India
nirbanroy04@gmail.com

Swapnanil Adhikary
Department of Computer Science & Engineering (IT)
Institute of Engineering & Management
Kolkata, India
adhikaryswapnanil@gmail.com

Shuvam Chakraborty
Department of Computer Science & Engineering
Indian Institute of Technology
Delhi, India
mcs242004@iitd.ac.in

Susovan Jana
Department of Computer Science & Enginnering (IoT),
University of Engineering & Management
Kolkata, India
jana.susovan2@gmail.com

***Abstract*—The complexes of camouflaging environments present a challenge in computer vision, especially in the context of natural agricultural settings, where objects merge with their surroundings. This paper presents a benchmark dataset, named DLP353, which consists of 353 high-resolution images of lime fruits taken under different environmental conditions, such as lighting and occlusions. The proposed data set is a handy tool through which machine learning applications can be tested for green fruit detection and segmentation in the problem of dense green foliage. Subsequent investigations using YOLOv8 revolve around object detection/segmentation techniques to assess their capacity for overcoming the conceptual limitations imposed by the camouflaged object detection task. The comparative study revealed that although object detection is more computationally efficient, image segmentation is superior concerning precision and robustness of boundary establishment. Future works will expand the same dataset to include other types of fruits and scenarios concerning extreme and harsh weather environments and also examine advanced architectures to improve the accuracy of detection. This study offers a substantial contribution to the growing resource that works toward machine learning-based solutions in precision agriculture and beyond. It is particularly focused on areas that necessitate object detection under challenging visual conditions.**

***Index Terms*—Lime Fruit, Image segmentation, Machine learning, Computer vision, Precision Agriculture**

I. Introduction

The detection of objects in camouflaging environments poses a significant challenge in computer vision, particularly when objects blend seamlessly with their surroundings. This study addresses such challenges by focusing on the detection of green fruits, primarily limes, against dense green foliage. The difficulty of this task is compounded by the similar color and texture of the fruits and their surrounding foliage, further complicated by natural lighting variations, shadows, occlusions from overlapping leaves, and the unstructured nature of natural backgrounds. Under these conditions, traditional image processing approaches often fail to produce reliable results. To tackle these challenges, a comprehensive dataset was created, capturing real-world conditions such as natural lighting, diverse angles, and occlusions to form a robust database of lime images recorded in nature. This dataset serves as a benchmark for evaluating advanced object detection and segmentation algorithms, demonstrating their potential for precise and reliable performance in challenging scenarios. This paper highlights a comparative study of object detection and image segmentation using YOLOv8, alongside an object detection-focused analysis between YOLOv8 and Faster R-CNN. These models were trained on the custom dataset to evaluate their effectiveness in identifying green fruits under camouflaging conditions.

The results of this study not only underline the feasibility of using advanced models for object detection in visually challenging environments but also validate the robustness of the proposed dataset as a benchmark. The findings open new opportunities for extending such methods to other applications, such as detecting red objects against red backgrounds or identifying wildlife in dense foliage.

II. Related Work

The task of detecting green fruits, especially in lime plantations, poses a significant challenge due to the close visual resemblance between the fruits and surrounding leaves. This issue is further intensified by variations in natural lighting conditions, which often undermine the effectiveness of traditional image processing techniques. Earlier research sought to address this by employing hyperspectral imaging, which captures a broader spectrum of light to enhance detection accuracy. Hyperspectral imaging has proven useful in various agricultural applications, such as detecting green apples, by allowing a clearer distinction between fruits and their background foliage [1], [2].

More recent advancements in computer vision have introduced new methods for improving fruit detection. One notable effort is the development of the PlantDoc dataset by Singh et al., which focuses on identifying plant diseases in uncontrolled outdoor environments, emphasizing the importance of real-world datasets for agricultural image classification [4]. Similarly, Wachs et al. combined infrared and visible spectrum cameras to improve apple detection, demonstrating how multi-modal imaging can enhance fruit detection in outdoor agricultural environments [3]. These early studies laid the groundwork for more sophisticated algorithms capable of detecting fruits in dynamic, natural settings.

In recent years, deep learning has become the dominant approach in object detection and image segmentation tasks, including fruit detection. YOLO-based models, such as YOLOv5 and YOLOv8, have achieved significant success in real-time detection and segmentation of fruits. Liu et al. demonstrated the effectiveness of lightweight YOLOv5 models when deployed on agricultural robots, achieving high accuracy while maintaining low computational requirements [5]. Additionally, the Mask R-CNN framework has been applied to fruit detection, particularly in scenarios with occlusion, further advancing automation in agriculture [9].

In addition to deep learning techniques, the integration of multiple imaging modalities has been explored to further improve detection accuracy. For instance, Bansal et al. combined thermal and multispectral imaging to detect immature green citrus fruits, illustrating that such combinations can significantly enhance detection under challenging lighting conditions [7]. Furthermore, Chebrolu et al. developed a comprehensive dataset for agricultural robots that integrates sensor data with RGB-D and laser scanning, offering valuable insights into tasks such as fruit localization and yield estimation [8].

One persistent challenge in fruit detection is dealing with occlusions and fluctuating environmental conditions. Liu et al. utilized the Fully Convolutional One-Stage (FCOS) model to detect green apples, highlighting how lightweight object detection architectures can still perform well in dense foliage environments [10]. Recent advancements in automated harvesting integrate computer vision and machine intelligence. Jana et al. [11] proposed a method using GLCM texture and statistical color features for fruit classification. Images are preprocessed to segment fruit from the background, and feature descriptors are used to train an SVM model. Their approach achieves high accuracy and is efficient for embedded systems, demonstrating potential for real-time automated fruit sorting.

III. Methodology

This section outlines the process of dataset creation, data processing, and the implementation of advanced object detection models, YOLOv8 and Faster R-CNN, to detect green fruits against similar-colored backgrounds. The methodology emphasizes a comparative study of their performances, showcasing the robustness of the custom dataset in addressing camouflaging challenges in precision agriculture.

A. Dataset generation

A specialized dataset has been created with 353 high quality images from lime trees under various environmental settings. This dataset was developed to build a machine learning model to detect-the-green-fruits-in-their-natural-agricultural holds. Table I gives the complete specifications of the dataset.

TABLE I: Specifications Table

Subject	Agricultural Science, Computer Vision, Machine Learning
Specific subject area	Green fruit detection using image segmentation in agricultural crops
Type of data	Image, Table, Chart, Graph, Figure
How data were acquired	Data acquired using a smartphone camera, with manual labeling using LabelImg for bounding boxes and CVAT for segmentation masks.
Data format	Raw, Analyzed
Parameters for data collection	Images collected at different times of the day (morning, midday, late afternoon) under varying lighting conditions and multiple perspectives.
Description of data collection	Images collected in lime plantations over several days to capture diverse lighting and environmental conditions. Each image was manually annotated with bounding boxes and segmentation masks, ensuring compatibility with YOLOv8.
Data source location	**Location**: Parul Apartment, 151 Sahid Hemanta Kumar Basu Sarani, South Dumdum, Kolkata, West Bengal 700074 **City/Region**: South Dumdum, Kolkata **Country**: India **GPS**: Latitude: 22.6140644 N, Longitude: 88.4023184 E
Data accessibility	**Repository name**: DLP-353 **URL**: https://github.com/GHOSTCALL983/DLP-353.git

B. Data Acquisition and Annotation

The images of the dataset was taken on a smartphone, on the grounds that they shall not be out of place in practical application for agricultural use. This dataset thus captures different lighting conditions in different times of the day:

- **Morning (6-9 AM):** Low-angle sunlight creates shadows and varying lighting intensities.
- **Midday (12-3 PM):** Strong sunlight enhances contrast between fruits and foliage.
- **Late Afternoon (4-6 PM):** Diffused light poses challenges for detecting fruits in lower-contrast environments.

Fig 1 shows representative samples from the raw dataset, illustrating typical conditions under which the images were captured.

Fig. 1: Images from the dataset

Annotations were performed manually using the LabelImg tool for bounding boxes and CVAT for segmentation masks. Each image was labeled with precise bounding box coordinates (x_1, y_1) and (x_2, y_2) and segmentation masks represented as binary matrices $M \in \mathbb{R}^{H \times W}$, where:

$$M_{ij} = \begin{cases} 1, & \text{if pixel } (i,j) \text{ belongs to the fruit} \\ 0, & \text{otherwise.} \end{cases}$$

C. Data Preprocessing

The captured images were resized to a uniform dimension of 640×640 pixels to meet the input requirements of YOLOv8. Resizing was performed using the following transformation:

$$I_{\text{resized}} = \text{resize}(I, 640, 640),$$

where I represents the original image. Pixel values were normalized to the range $[0, 1]$ using:

$$I_{\text{normalized}} = \frac{I_{\text{resized}}}{255}.$$

This systematic methodology ensures robust detection of green fruits amidst complex backgrounds, addressing the challenges posed by camouflaging environments.

IV. Comparative Analysis of Object Detection and Image Segmentation

The ability to distinguish green fruits from their dense foliage background is crucial for precision agriculture applications, particularly under challenging natural conditions. In this study, two approaches, YOLOv8 for object detection and segmentation, were implemented and evaluated to assess their effectiveness in identifying lime fruits amidst complex green backgrounds. This comparative analysis provides insights into the strengths and limitations of each method, emphasizing their suitability for detecting camouflaged objects in agricultural contexts.

A. Methodology for Comparative Evaluation

Both object detection and image segmentation were performed using the DLP-353 dataset. YOLOv8's object detection mode focuses on bounding box-based localization, while its segmentation mode predicts pixel-wise masks for a more granular understanding of fruit location. The evaluation was based on standard performance metrics:

- **Mean Average Precision (mAP):** Assesses precision-recall trade-offs at multiple confidence thresholds.
- **Mean Precision:** Indicates the fraction of correct positive predictions among all detections.
- **Mean Recall:** Measures the proportion of actual instances correctly identified by the model.

The workflow for detection, training, and evaluation is summarized in Fig 5. Each stage ensures robust preparation, training, and validation to derive meaningful comparisons.

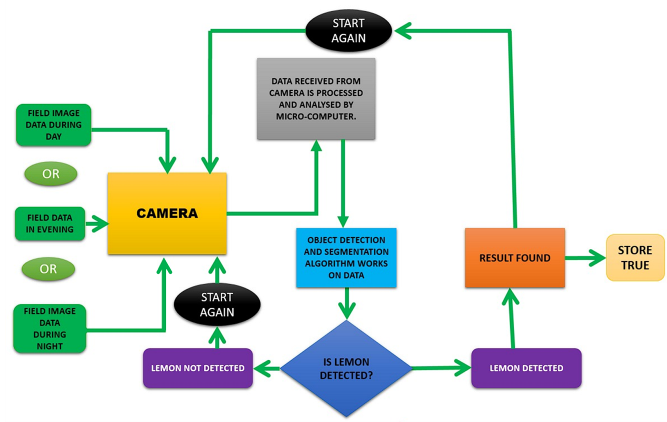


Fig. 2: Flowchart of the detection and evaluation workflow

V. Comparison of YOLOv8 and Faster R-CNN

To further evaluate the robustness of YOLOv8, a comparative study was conducted with Faster R-CNN, a widely used two-stage object detection model. Both models were trained on the DLP-353 dataset and assessed based on their capability to detect lime fruits in complex environmental conditions.

A. Implementation of Faster R-CNN

Faster R-CNN follows a two-stage detection pipeline:

- **Region Proposal Network (RPN):** The first stage involves generating region proposals where objects are likely to be present. This is achieved using an RPN, which scans the image with anchor boxes and predicts objectness scores and refined bounding box coordinates.
- **Classification and Refinement:** In the second stage, the proposed regions are refined and classified into specific categories, including the target class (lime fruit) and background.

For this study, transfer learning was applied to fine-tune Faster R-CNN using pre-trained weights. The final two layers of the network were modified to adapt to the lime fruit detection task. The model was trained with optimized hyperparameters, including a learning rate of 0.001 and batch normalization to improve convergence.

B. Visual Comparisons

Figures 3 and 4 depict the detection outputs of YOLOv8 and Faster R-CNN, respectively.

YOLO_Detection.png

Fig. 3: Detection of lime fruits using YOLOv8 (Bounding Box and Segmentation Outputs)

Faster_RCNN_Detection.png

Fig. 4: Detection of lime fruits using Faster R-CNN (Bounding Box Outputs)

C. Key Observations

- **YOLOv8:** The model provides fast real-time detection with accurate bounding boxes and segmentation masks. However, in instances where the contrast between fruits and foliage is minimal, YOLOv8's segmentation sometimes results in false positives.
- **Faster R-CNN:** The two-stage architecture allows for more refined object detection, particularly in highly occluded environments. Faster R-CNN produced more accurate bounding boxes, with fewer false positives than YOLOv8, making it more reliable for precise applications. However, due to its computational complexity, it is less suitable for real-time deployment.

VI. Workflow for Detection and Evaluation

The workflow for detection, training, and evaluation is summarized in Fig. 5. Each stage ensures a systematic approach to data collection, model training, and validation.

image.png

Fig. 5: Flowchart of the detection and evaluation workflow

This structured workflow ensures an end-to-end approach for detecting lime fruits in challenging agricultural environments. The subsequent section will provide a detailed quantitative analysis of the models' performances.

VII. Results and Analysis

This study aimed to evaluate the performance of object detection and image segmentation methods for green fruit detection using YOLOv8. The primary objective was to identify lime fruits against a challenging green foliage background. Both methods were assessed on the DLP353 dataset using metrics such as mean Average Precision (mAP), mean Precision, and mean Recall. This section presents a comparative analysis of these two approaches, incorporating the corresponding evaluation metrics, visual outputs, and key observations.

A. Evaluation Metrics and Comparative Performance

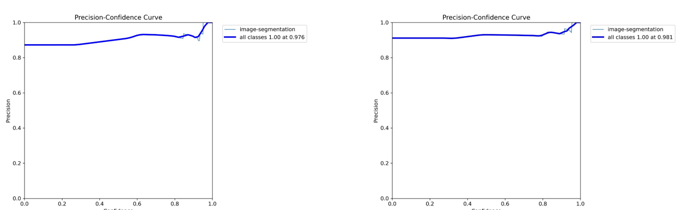
Table II summarizes the performance of object detection and segmentation. While both approaches yielded high accuracy, image segmentation demonstrated better precision and overall mAP values, highlighting its effectiveness in handling occlusions and intricate details of lime fruits. Conversely, object detection exhibited slightly higher recall, suggesting its robustness in identifying fruit presence even in ambiguous scenarios.

TABLE II: Performance Comparison of YOLOv8 Object Detection and Segmentation

Method	Mean Precision	Mean Recall	mAP
YOLOv8 Detection	0.9111	0.9546	0.9567
YOLOv8 Segmentation	0.9317	0.9523	0.9334

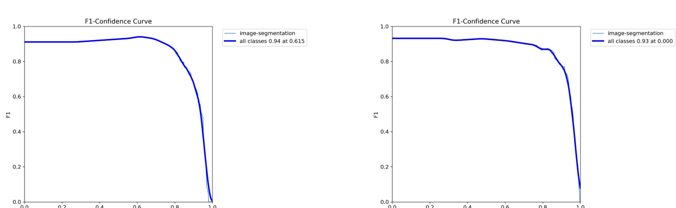
B. Visual Analysis of Performance

Figures 6, 7, 8, and 9 provide a side-by-side visual comparison of precision-confidence, F1-confidence, recall-confidence, and loss analysis curves for both object detection and image segmentation methods. These visualizations illustrate the differences in performance metrics and training behaviors between the two approaches.



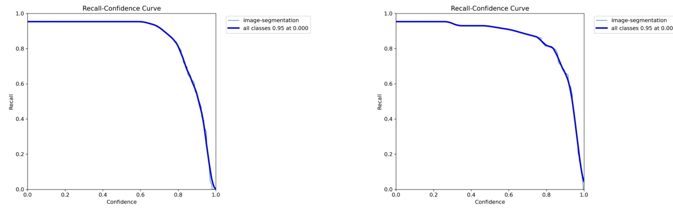
(a) Object Detection (b) Image Segmentation

Fig. 6: Precision-Confidence Curves for Object Detection (Left) and Image Segmentation (Right)



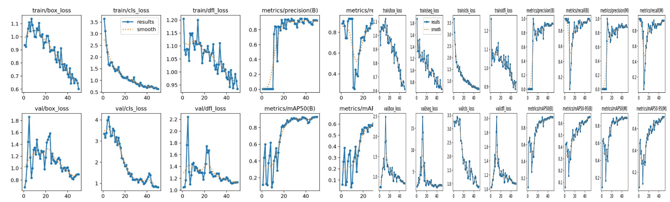
(a) Object Detection (b) Image Segmentation

Fig. 7: Precision-Confidence Curves for Object Detection (Left) and Image Segmentation (Right)



(a) Object Detection (b) Image Segmentation

Fig. 8: Recall-Confidence Curves for Object Detection (Left) and Image Segmentation (Right)

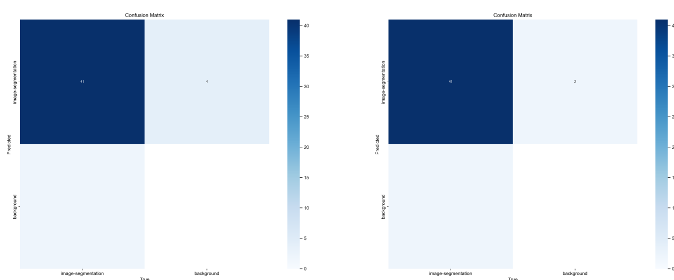


(a) Object Detection (b) Image Segmentation

Fig. 9: Loss Analysis for Object Detection (Left) and Image Segmentation (Right)

C. Confusion Matrix Comparison

Figures 10 display the confusion matrices for object detection and image segmentation.



(a) Object Detection (b) Image Segmentation

Fig. 10: Confusion Matrices for Object Detection (Left) and Image Segmentation (Right)

D. Discussion and Insights

The results from our study demonstrate that image segmentation is highly effective for tasks requiring accurate localization and boundary delineation, such as detecting camouflaged lime fruits in dense foliage. On the other hand, object detection provides faster processing and is better suited for resource-constrained environments. Therefore, the choice between detection and segmentation depends on the specific application, balancing accuracy and computational efficiency.

When compared to the work by [9], which proposed a Mask R-CNN-based model enhanced with a Gaussian non-local attention mechanism for apple segmentation, our approach offers several unique advantages:

1. **Challenge of Camouflaging Environments:** Unlike [9], which focuses on apples against a monochromatic background, our method addresses the detection and segmentation of lime fruits in a camouflaged environment where the fruits blend with dense green foliage. This significantly increases the complexity of the task and highlights the robustness of our model.

2. **Diverse Dataset:** The DLP353 dataset used in our work captures real-world conditions, including diverse lighting scenarios (morning, midday, and late afternoon) and natural occlusions. This makes our dataset more representative of practical agricultural settings compared to the controlled conditions considered in [9].

3. **Computational Efficiency:** While [9] utilizes a computationally intensive attention mechanism integrated with Mask R-CNN, our YOLOv8-based model provides a more lightweight solution that balances efficiency and accuracy. This makes our approach suitable for deployment on edge devices or drones used in agricultural monitoring.

4. **Performance Metrics:** Although the method in [9] achieves AP box and AP mask metrics of 85.6% and 86.2%, respectively, our segmentation approach achieves a competitive mAP of 93.34% while maintaining faster inference speeds and operational flexibility, making it well-suited for real-time applications.

5. **Generalizability:** Our framework is designed with scalability in mind, allowing it to extend beyond lime fruit detection to other fruit types and environmental conditions, including extreme lighting or weather changes. This generalizability provides a broader scope of applications compared to the specific apple segmentation task in [9].

VIII. Conclusion

This study addresses the challenge of detecting objects in camouflaging environments, specifically focusing on the detection of green fruits, such as limes, against dense green foliage using the YOLOv8 framework. The criticalness of this problem is rooted in addressing very similar colored and textured objects from their backgrounds, especially in the context of precision agriculture, and many more fields. Using object detection and image segmentation, performance evaluations were conducted with respect to YOLOv8 over the DLP353 dataset, underlining both merits and limitations of each approach. Results evidence that while object detection is computationally efficient, image segmentation would be a better option when there is a need of accurate boundary definition and occlusions handling.

Our findings reveal that image segmentation excels at revealing camouflaged subjects, suitable for precision works in agriculture and elsewhere. The more generalized data for different fruit types, extreme climates, and different environmental conditions would also add robustness to detection models. The methods and results presented in this study contribute to the ongoing research in object detection and segmentation under visually challenging conditions. They pave the way for innovative solutions in applications such as precision agriculture, wildlife monitoring, and automated systems operating in natural and dynamic environments.

References

[1] Chen, C. H. (2000). Information Processing for Remote Sensing. River Edge, N.J.: World Scientific Publishing.

[2] Cheng, X., et al. (2004). A novel integrated PCA and FLD method on hyperspectral image feature extraction for cucumber chilling damage inspection. *Transactions of ASAE*, 47(4), 1313–1320.

[3] Wachs, J. P., et al. (2010). Low and high-level visual feature-based apple detection from multi-modal images. *Precision Agriculture*, 11(6), 717–735.

[4] Singh, D., et al. (2020). PlantDoc: A Dataset for Visual Plant Disease Detection. *CoDS COMAD 2020*, 5–7.

[5] Liu, G., et al. (2023). Lightweight object detection algorithm for robots with improved YOLOv5. *Engineering Applications of Artificial Intelligence*, 123, 106217.

[6] Silva, D., et al. (2023). Automatic Counting and Georeferencing of Fruits from Video. *NOVA University Lisbon, Master's Thesis*, September 2023.

[7] Bansal, R., Lee, W. S., & Satish, S. (2013). Green citrus detection using fast Fourier transform (FFT) leakage. *Precision Agriculture*, 14(1), 59–70.

[8] Chebrolu, N., et al. (2017). Agricultural robot dataset for plant classification, localization, and mapping on sugar beet fields. *The International Journal of Robotics Research*, 1–8.

[9] Jia, W., et al. (2022). Improving apple fruit detection using a Mask R-CNN model under occlusion. *Journal of Agricultural Engineering*, 50(4), 118-125.

[10] Liu, M., et al. (2022). A modified FCOS method for green fruit object detection in complex orchard environments. *Computers and Electronics in Agriculture*, 197, 106984.

[11] S. Jana, S. Basak, and R. Parekh, "Automatic fruit recognition from natural images using color and texture features," in *2017 Devices for Integrated Circuit (DevIC)*, Kalyani, India, 2017, pp. 620-624, doi: 10.1109/DEVIC.2017.8074025.