Design and Development of Vegetable Detection and Recognition Model Using Deep Learning In Market Environment

Mitkumar Patel
Department of Computer Science and Engineering, School of Technology, Pandit Deendayal Energy University
Gujarat, India, 382007
Email: mit.pce21@sot.pdpu.ac.in

Mohammad Soban Shaikh
Department of Computer Science and Engineering, School of Technology, Pandit Deendayal Energy University
Gujarat, India, 382007
Email: soban1103@gmail.com

Hiren Kumar Thakkar*
Department of Computer Science and Engineering, School of Technology, Pandit Deendayal Energy University
Gujarat, India, 382007
Email: hiren.pdeu@gmail.com

Aditya Shastri
Department of Computer Science and Engineering, School of Technology, Pandit Deendayal Energy University
Gujarat, India, 382007
Email: aditya.shastri111@gmail.com

***Abstract*—Identifying and recognizing visually similar vegetables in a cluttered environment, such as a market, is a challenging task. We propose an efficient vegetable detection and recognition system that can effectively detect vegetables, specifically Gourd species and cucumber, by drawing bounding boxes around them and successfully recognizing the vegetable species. This system uses YOLO (You Only Look Once) for both vegetable detection and identification. The system is trained on a curated dataset containing images of these vegetables, with annotated bounding boxes for effective detection and classification even in a cluttered environment where multiple vegetables are present. We include an interactive user interface using Streamlit, through which users can upload images to detect and recognize the vegetables instantly. This paper introduces a deep learning approach capable of accurately distinguishing between similar-looking vegetables, adding value to automation processes in the agriculture and retail sectors.**

***Index Terms*—Vegetable Detection, Vegetable Recognition, Deep Learning, Vegetable Classification, YOLO.**

I. Introduction

In today's world, artificial intelligence takes over the majority of work in the different industries by solving problems that are not efficiently solved by the traditional approaches. It also improves efficiency through automation, prediction, and decision-making. Among those vital sections of life one cannot even think of the existence of human beings without agriculture. Computer vision is one of the subfields of AI, which is utilized in performing very complex tasks like weed detection, crop disease identification, and crop monitoring. In those areas of work where even the most basic labor work is too costly or highly prone to human error, these AI-driven systems can analyze crops with high accuracy and significantly reduce waste, and enhance productivity as a whole.

In this paper, we discuss the development of a system aimed at addressing the challenge of recognizing visually similar vegetables, specifically focusing on distinguishing gourd species like bitter gourd, bottle gourd, ridge gourd, sponge gourd, and cucumber. Additionally, the project is designed to operate effectively in cluttered environments such as supermarkets and markets, ensuring the model can accurately detect and classify vegetables amidst multiple other items.

Traditionally, the kinds of gourd species are detected through the respective physical features of the gourd, like shape, size, texture, and color. Experienced farmers and our parents too can easily identify the differences between some visually similar vegetables like cucumbers and different kinds of gourd species. In the retail environment, detection and recognition of vegetables through computer intervention is more effective than the traditional method based on manual identification of these kinds of vegetables, especially in open markets where barcodes are not practica and mistakes in recognizing similar vegetables result in error of pricing, thus slowing down the process of checkout and reducing the customer's satisfaction level. Hence this system is designed particularly to help in the case of situations when identification is a problem during vegetable purchasing for newbies and people like us. This would be achieved by integrating artificial intelligence and computer vision techniques with the details mentioned below.

We propose a real-time vegetable recognition system for identification and detection of some special kinds of vegetables, as mentioned earlier, for cluttered market environments.

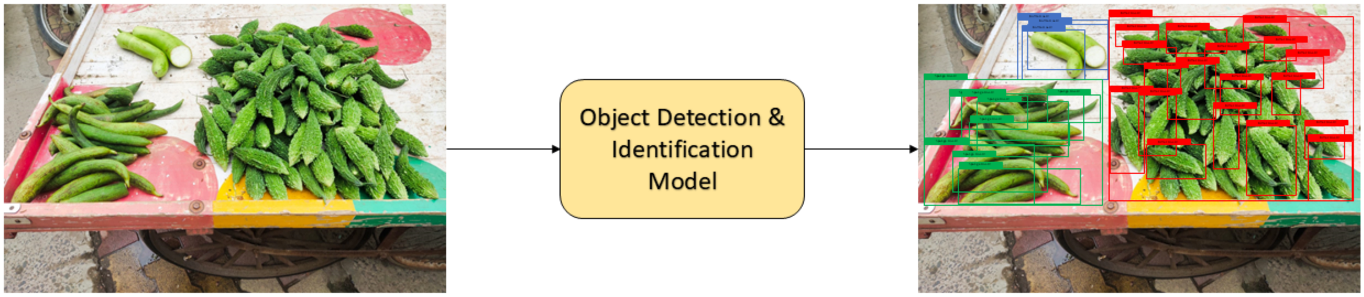


Fig. 1: Input and Output of the Proposed System

The input and output of the proposed system is shown in Fig. 1. The method used draws a reliance on object detection models like Yolov8, which is based on advanced deep learning techniques, which will be pre-trained on a huge dataset in order to object detection and identification thus it will be helpful in the case of situations when identification is a problem during vegetable purchasing for newbies and people like us. In fact, this process of identification will automatically be made by this system, thus being of great help even to novice buyers as well as professionals in making appropriate distinctions.

Additionally, to make it easily accessible for users, we integrated an interactive user interface using Streamlit. With this, users can upload images to detect and classify vegetables instantly. This proposed system shows the application of AI in agriculture and retail; it improves efficiency in detecting and classifying vegetables, contributing to the modernization of market operations.

II. RELATED WORKS

The previous research on the automation of fruit and vegetable detection and localization in unconstrained environments presents FRUVEG67: a novel dataset containing 67 classes of images captured in real-world scenarios using limited manual annotations. Also, proposed automatic data annotation algorithm SSDA(semi-supervised data annotation algorithm) and the Fruit and Vegetable Detection Network, FVDNet, that enhances detection accuracy with an ensemble approach using YOLOv7 [1]. System finds it difficult to capture finer details of smaller objects by its handling of occlusions and cluttered backgrounds.

The paper "Colombian Fruit and Vegetables Recognition Using Convolutional Neural Networks and Transfer Learning" presents an automatic classification system to recognize the columbian produce in supermarkets. This system uses convolutional neural networks with transfer learning. Model handles complexities like color, texture, lighting and produce inside transparent bag and outside of bag with around 4980 images across 22 categories, captured with samsung WB50F camera at a resolution of 4608 x 3456 pixels [2].

Projjal Sarkar, Paramita Sarkar and their colleagues proposed a classification model for vegetable recognition in the "Indian Vegetable Image Classification Using Convolutional Neural Network" to improve efficiency in vegetable production and distribution. Model achieved 97.5% accuracy in classifying 15 vegetables [3]. As the future scope of this research we can expand the dataset with more variety of vegetables, incorporating object tracking and motion detection.

To streamline the checkout process in supermarkets, image processing-based solution proposed in the "Automatic Vegetable Recognition system". Hridkamol Biswas, Faisal Hossain applied this model in MATLAB, by considering the various features like weight, color, shape, size, texture. The dataset includes color histograms and other visual features of various vegetables. This system utilizes basic image processing techniques, so it misclassifies the vegetables due to similar color profiles and varying imaging conditions [4]. For improvement of this system, required further refinement in feature recognition.

The review paper "Object Detection and Recognition Techniques Based on Digital Image Processing and Traditional Machine Learning for Fruit and Vegetable Harvesting Robots: An Overview and Review" is provide a systematic summary and analysis of object detection and recognition techniques for fruit and vegetable harvesting robots, focusing on digital image processing and traditional machine learning. The review serves as a reference for future research in improving harvesting robot technologies. Feng Xiao, Haibin Wang and their colleagues discuss various techniques based on digital image processing-color, shape, texture features, multi-feature fusion and machine learning methods-K means clustering, SVM, KNN, AdaBoost, Decision Tree and Bayesian algorithms. According to research result KNN clustering produced best detection and recognition results. Additionally, SVM algorithm achieved better accuracy compared to other classifiers [5] However, such studies are rather rare reporting absolute Precision of each method, and comparison of performances amongst the methods, in the same environment.

The paper 'Fruit and Vegetable Identification Using Machine Learning for Retail Applications' presents a Machine learning system for identifying fruits and vegetables in retail environments. Raspberry pi utilizes convolutional neural networks like MobileNet and inception, for improving usability and minimizing human-computer interaction. Dataset consists

Brief Title	Main Idea	Dataset	Key Findings	Drawbacks
Fruit and vegetable detection in unconstrained environment [1]	Enhance Fruit and vegetable detection in unconstrained environment with novel dataset.	Curated dataset named "FRU-VEG67" Total 67 classes.	Introduced SSDA annotation, FVDNet model - enhances the detection and localization of fruits, vegetables.	Fixed grid size in various configurations may lead to reduced precision.
Fruit and vegetable recognition using CNN and Transfer learning [2]	Developed a fruits and vegetables classification system for using a CNN, regardless of type, color, quantity, and texture.	Created a dataset with 4980 images, categorized into 22 classes. Dataset contains images - vegetables/fruits inside the bag.	Proposed classification system achieved high accuracy in identifying various fruits and vegetables, showcasing effectiveness of the model.	Dataset's potential lack of diversity.
Indian Vegetable classification using CNN [3]	Develop a model which can correctly recognize Indian vegetables.	Vegetable image dataset: 15 classes like bean, tomato, Bitter Gourd, Bottle Gourd, Brinjal, Broccoli, Cabbage, Capsicum, Carrot, Cauliflower, Pumpkin, Radish, Potato.	The CNN model achieved a recognition rate of 97.50%.	Model occasionally misidentifying background items as vegetables, indicating potential issues with accuracy in complex images.
Vegetable Vision: Automatic vegetable recognition system [4]	Introduces the automatic recognition system that uses image processing to classify and recognize vegetables, accelerating the checkout process in supermarkets.	The dataset used in this research consists of images of various vegetables, including cabbage, apple, zucchini, and broccoli.	Achieved a high accuracy of 96.55% in analyzing the features such as shape, color, size and texture.	The paper does not explicitly mention specific critiques or limitations of the system developed, which might lead to an incomplete understanding of potential weaknesses.
Fruit and Vegetable Harvesting Robots: An overview and review [5]	To systematically summarize and analyze object detection and recognition techniques for fruit and vegetable harvesting robots based on digital image processing and traditional machine learning.	Analyzed various models which are using different datasets: Fruit-360, Fruit-A, Fruit-B, Lemon quality control dataset, Apple, Cauliflower.	Highlight the effectiveness of multi-feature fusion and SVM algorithms.	Detection and recognition are affected by the factors like lightning condition and feature correlations among data.
In retail applications: Identify fruits and vegetables [6]	Enhancing self-service systems in retail by using computer vision to automate fruit-vegetable identification system based on visual features like shape, color, texture.	Created a dataset by combining ImageNet and self-collected images of fruits and vegetables.	Around 86% of users prefer this system over the current traditional system.	Inability to differentiate between varieties of the same product. Limited usability testing identifying only 50-70% of issues.
Vegetable Classification Using You Only Look Once Algorithm [7]	Develop a YOLO Model for classification of vegetable and fruit species which includes Green Capsicum, Cucumber and Green Apple.	A total of 300 images were taken from the internet, i.e 100 for the three classes including cucumber, green apple, and green capsicum from the internet , of which 180 images were used for training and 120 images were used for testing.	For images, more than 50% of the test images were detected properly and for videos more than 70% of the time the vegetables were correctly classified. The model was tested in challenging conditions and complex backgrounds.	The False positive percentage for cucumber was 40%, while that for Green Apple was 50% and that for Green Capsicums was 60%
A Vegetable Category Recognition System Using Deep Neural Network [8]	To develop a category recognition system based on Convolutional Neural Networks (CNN) using the framework Caffe for 8 different species of vegetables	The dataset used consists of 160 images for training with 20 images for each of the eight vegetable categories and 40 images for testing, that is 5 images for each category.	The system was trained on CNN based models using the framework Caffe, and on training on 3 million learning iterations, 99% performance was obtained which significantly dropped when 10 million iterations were done.	For the case of 10 million learning iterations, the system used the background and not the actual vegetables to detect the class of the given vegetable. This was due to the small size of the dataset.
A Deep CNN approach to detect and classify local fruits through a web interface [9]	Create a web based system to detect the locally available fruits in the Bangladesh market using Resnet-50, VGG-19, Inception-V3.	The dataset consists of 3240 high-quality images from eight distinct classes including Carambola, Bilimbi, Elephant Apple,Emblica, Burmese Grape, Sapodilla, Tamarind, and Wood Apple.	Using MobileNet the best f1 score of 0.985 was achieved which was best among the among the other models including VGG-19, Inception-V3 and Resnet-50.	The paper does not explicitly mention specific critiques or limitations of the system developed, which might lead to an incomplete understanding of potential weaknesses.
Detection of various categories of fruits and vegetables using machine learning techniques [10]	Detection of 20 categories of fruits and vegetables using the machine learning techniques like KNN and C4.5 models.	The dataset of fruits and vegetables, consists 2915 total images and 20 different categories.	The C4.5 model is more effective and yields much better performance than the KNN model. Otsu thresholding was done for preprocessing and 96.63% accuracy was achieved.	No drawback

TABLE I: Summary of Literature Review

of a mix of images from ImageNet and self-collected samples, divided into the 10 fruits and vegetables. Moreover, reliance on existing classifiers limits the system's adaptability [6].

The paper 'Vegetable Classification Using You Only Look Once Algorithm' presents a Yolo model for the classification of fruits and vegetable species, including green capsicum, cucumber, and green apple. The model detects with an accuracy of 70% when the threshold is set to 0.15. The dataset contains 300 images, of which 100 images belong to each class, and 180 images were used for training and 120 for testing. Moreover, the false positive rates for cucumbers were 40%, capsicums were 60%, and green apples were 50% [7].

The paper 'A Vegetable Category Recognition System Using Deep Neural Network' presents a category recognition system built on the Caffe framework that follows CNN architecture. The system was trained on 3 million iterations, and 99% performance was obtained. The dataset contains 160 images for training with 20 images for each of the categories of vegetables, while 40 images were used for testing with 5 images for each category of vegetables [8].

The paper 'A Deep CNN approach to detect and classify local fruits through a web interface' presents a web-based system built on the top of a combination of ResNet-50, VGG-19, Inception-V3, and MobileNet to detect the locally available vegetables and fruits in the Bangladeshi market. 3240 images from 8 different categories were used, and the best F1 score achieved was 0.985 using MobileNet [9].

The paper 'Detection of various categories of fruits and vegetables using machine learning techniques' presents a system with classical machine learning methods like K Nearest Neighbours and C4.5 models, trained after performing Otsu thresholding on the dataset containing 2915 images of 20 different kinds of vegetables and fruits. The results suggested that the C4.5 model was much more effective, resulting in 96.64% accuracy, much better than the KNN model [10].

The summary of the literature survey is depicted in the Table I given below.

III. METHODOLOGY

A. Dataset

The dataset used in this paper is a varied dataset containing different images of vegetables specificaly gourd species and cucumber in different illuminations, angles, distances. The dataset is developed on the foundation of five diverse species of gourds: bitter gourd, bottle gourd, sponge gourd, ridge gourd, and cucumber. The images are collected from various sources due to high variation in illumination, angles, and backgrounds. A large majority of the dataset is composed of images captured in the real market. Supplementary images are also sourced from other datasets of vegetables and public image sources such as Google.

All classes consist of 1,200 images each and are divided in an organized manner towards training, validation, and testing. In distribution, we assigned the training to 1,000 images per class, 100 to validation, and 100 for the purpose of testing.

(a) Bitter Gourd

(b) Bottle Gourd

(c) Ridge Gourd

(d) Sponge Gourd

(e) Cucumber

Fig. 2: Sample images from our dataset

After gathering all the images, the dataset was annotated on the Roboflow platform, where bounding boxes were drawn around every vegetable to allow the task of object detection. This annotation process enabled labeling the YOLOv8 model precisely to identify and differentiate amidst the different gourd species accurately.

B. Workflow of the project

The project workflow, as shown in Fig. 3 includes the following key steps:

*1) **Start**:* The workflow begins with defining the project scope and goals in specific terms, which are towards the identification and classification of five species of gourds, namely bitter gourd, bottle gourd, sponge gourd, ridge gourd,

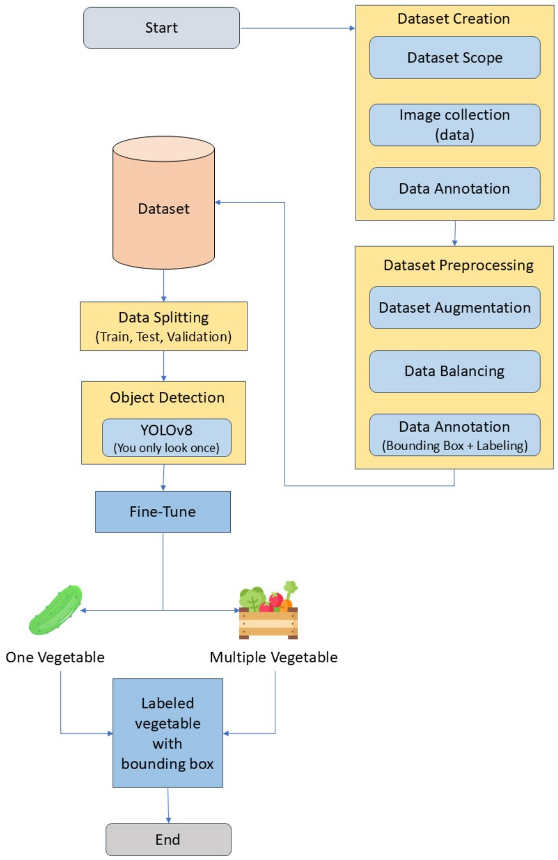
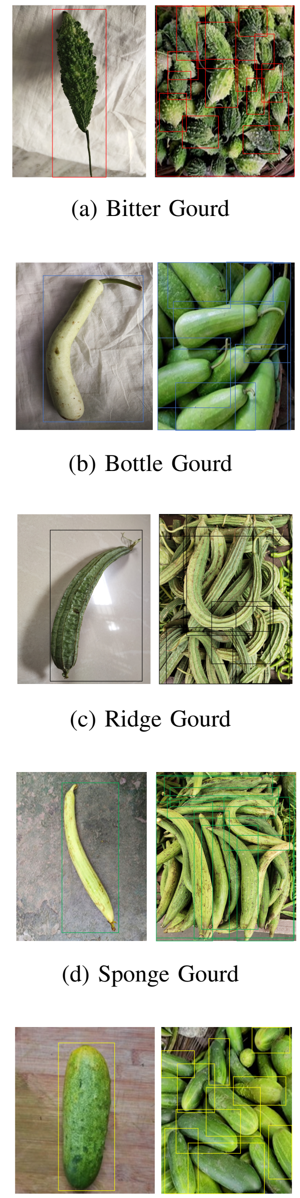


Fig. 3: The proposed methodology



(a) Bitter Gourd

(b) Bottle Gourd

(c) Ridge Gourd

(d) Sponge Gourd

(e) Cucumber

Fig. 4: Sample images with annotations from our dataset

and cucumber. Such an approach initially allows a structured, systematic methodology of vegetable recognition based on images.

*2) **Dataset Creation**:* Image Collection and Annotation. The dataset is acquired through the collection of images from various sources: real market environments and online image databases with different variations of lighting, angles, and backgrounds. This enhances the quality of the dataset and allows the model to generalize well in real-world applications. The images are annotated on Roboflow, where bounding boxes for each vegetable have been made so that the labeling is as accurate as possible. This enables the model YOLOv8 to learn what are the particular visual features of each gourd species.

*3) **Data Preprocessing**:* Augmentation and Data Balancing: It applies augmentation techniques such as rotation, flipping, and adjustment of brightness for adding variability to the dataset and to prevent overfitting during the preprocessing stage. Also, it performs data balancing with an aim at equal representation of all species to deter any biasing opportunity that the model may take. Then, this dataset is split into training, validation, and test for robust model training and proper estimation of performance. After the pre-processing step, the dataset appears as shown in Fig. 2.

*4) **Data Annotation**:* For object detection annotated dataset required for model training like training images have to label first. Have to create the bounding box around the vegetables(objects) is necessary to localize the object within the image. With well annotated data, clear signals get caught and upgrade the model's ability to detect and classify objects. Good annotation in images allows a model to distinguish objects from one another especially in a multi-object scene, which is very important especially for complex environments such as markets.

Sample images of annotated dataset shown in Fig. 4. For dataset annotation we utilized Roboflow. Roboflow is a platform that helps developers and data scientists to build,

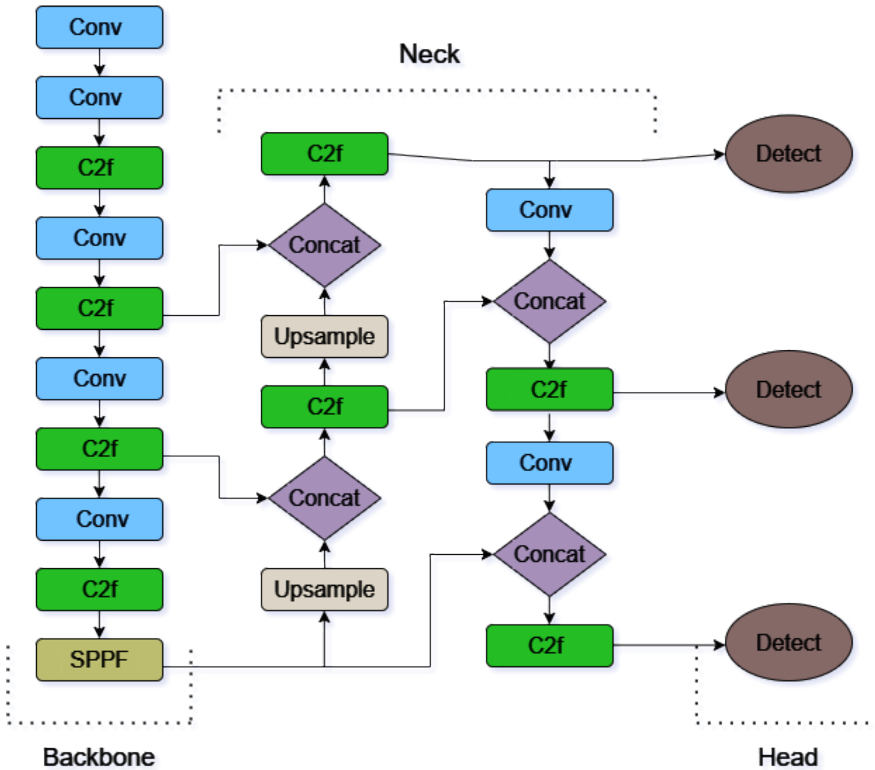


Fig. 5: Architecture of YOLO v8 [?]

train and deploy computer vision models. This platform provides one function for autolabel, where we have to upload the all images and have to set threshold then according to that threshold those images would be annotate. But we face challenge with this feature is that it is done with very lower confidence/threshold then it detect plastic bucket and other object like a vegetable. So, we manually delete such boxes and placed boxes around the vegetable.

*5) **Vegetable Detection with YOLOv8**:* Model Training and Fine-Tuning:the project basis uses the YOLOv8 model as the basis for the detection purpose. The YOLOv8 architecture "the Backbone, Neck, and Head" allows real-time detection through the feature extraction, multi-scale information fusion, and location output of bounding boxes with class labels. The model will iterate in its fine-tuning to change model parameters so that an appropriate balance between detection accuracy and speed may be achieved over an annotated dataset.

*6) **Output and Application**:* Real-Time Object Detection and Labeling. The learned model can identify individual and group of vegetables in an image with a bounding box surrounding the detected objects and the name of their species. This is very suitable for real-world applications, including automated vegetable recognition in market environments, in which adaptability and efficiency are desirable qualities of the model.

C. Architecture of YOLOv8

There are three main components to the architecture of YOLOv8: namely the Backbone, Neck, and Head, each uniquely contributing to the tasks of object detection and classification that allow YOLOv8 to be possible for highly accurate and highly efficient real-time performance. Fig. 5 illustrates the architecture of YOLO v8.

*1) **Backbone**:* The Backbone is the initial portion of the network, dedicated to extracting the essential features from input images. It consists of several pairs of convolutional layers (Conv) and C2f blocks that together respond to low-level features like edges, colors, and textures, used in the object classification. Hierarchical features captured by the convolutional layers can occur at any scale and occur gradually along the way from more simple to more complex patterns.

The backbone ends with an SPPF (Spatial Pyramid Pooling - Fast) block. The latter pools contextual information at different scales through spatial pooling that focuses the network's attention toward the informative regions and enhances its performance on object-size variations. This is one of the reasons why the backbone captures fine-grained information as well as global context in the input images, which is paramount in classifying and localizing the gourd species.

*2) **Neck**:* The Neck module, to say merely, connects the Backbone to the Head; therefore it aids in better object detection concerning various scales. It takes features from several layers and combines them into a more informative feature map that actually enhances the object localization. It consists of a wide range of operations: Concat, Upsample, and C2f blocks.

Concatenate operations pool feature maps from different layers to combine information from lower-resolution, deeper layers with higher-resolution, shallower layers. This fusion of information improves the model's ability to identify objects since it is able to register objects at various scales.

Upsample operations increase the resolution of feature maps. In this case, it allows the model to grab finer details regarding smaller objects. C2f further details the features because they improve the ability of the model to capture image relevant patterns and increase the accuracy in localisation.

Neck's structure comprised of alternating blocks of Concat, Upsample, and C2f allow YOLOv8 to effectively aggregate multi-scale features and further improve its ability to detect objects of all sizes and orientations.

3) ***Head:*** The final architecture component is the head and produces the final object detection outputs. Three layers of detection are used, independent of each other and all parallel, and correspond to features of different scales.

By way of this, each detection layer generates bounding boxes and class probabilities for objects in the image. Of course, every detection layer pays attention to different scales. Such a multi-scale structure supports Heads in the detection of small, medium, and large objects, especially the complexity of images with varied object sizes in real-world market images with different gourd species.

Outputs from the three layers are combined by YOLOv8, which allows it to accurately localize and classify objects with robust detection across all scenarios. The locations of each detected object in the image along with their respective classifications are the bounding boxes and class probabilities found in final boxes, thus completing the entire detection and classification of objects.

The model was fine-tuned on our annotated dataset for better recognition of the specific visual characteristics of each vegetable species. It was trained iteratively with alterations of parameters toward an optimum balance between accuracy and detection speed.

IV. EXPERIMENTAL RESULTS

As part of the research, we are able to create a dataset with 1200 images in each class and annotated images so it is compatible with YOLO. Also, this section highlights the performance of the YOLO v8 on the created dataset, focusing on both controlled and open-market environments. Also, it includes the class-wise analysis, key takeaways, and limitations in open market environments.

A. Dataset Overview

As shown in Table II, dataset consists of 5 classes: Each class contains 1000 training images, 100 images for validation and 100 images for testing, ensuring equal representation across all categories.

Vegetable Species Dataset			
Veg. Species	**Train**	**Validation**	**Test**
Bitter Gourd	1000	100	100
Bottle Gourd	1000	100	100
Ridge Gourd	1000	100	100
Sponge Gourd	1000	100	100
Cucumber	1000	100	100

TABLE II: Detailed description of the dataset

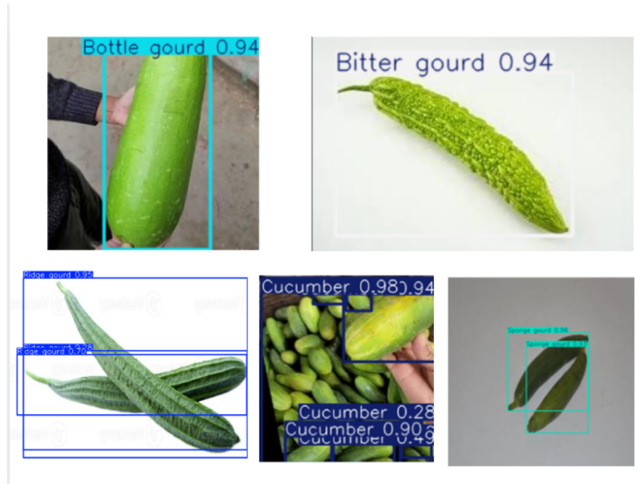


Fig. 6: Results of Test Images

B. Overview of Results

As depicted in Table III, promising results were obtained by training and testing the gourd detection system with YOLOv8 for the recognition of particular species in controlled environments. As the table shows, model evaluation metrics vary in performance between classes.

Fig. 6 shows the training results of the vegetable detection using YOLO v8. The mean Average Precision over all classes given IoU threshold 0.5 was 0.872, meaning that the location of all objects was mostly correct. The mean Average Precision over all classes given IoU threshold 0.50 to 0.95 was lower at 0.735, reflecting challenges in accurately localizing bounding boxes for smaller or overlapping objects.

C. Class-wise Accuracy

1) ***Cucumber:*** Reached precision of 0.845 and recall of 0.862, which means balanced performance.

2) ***Bottle Gourd:*** Has achieved the maximum values on precisions and recalls, i.e., it is the class for which differentiation by the model was perfect.

3) ***Bitter Gourd:*** Despite moderate precision (0.746), the recall dropped to 0.572, suggesting challenges in detecting all instances of this class.

4) ***Sponge Gourd and Ridge Gourd:*** Achieved high precision but slightly lower mAP@50-95 values, indicating variability in bounding box localization across different IoU thresholds.

Class	Images	Instances	Box(P)	R	mAP50	mAP50-95
All	506	1770	0.851	0.823	0.872	0.735
Ridge Gourd	100	211	0.885	0.848	0.926	0.735
Bottle Gourd	100	255	0.851	0.851	0.871	0.729
Bitter Gourd	100	339	0.746	0.572	0.656	0.439
Sponge Gourd	99	127	0.927	0.984	0.987	0.939
Cucumber	100	838	0.845	0.862	0.922	0.832

TABLE III: Results of Training

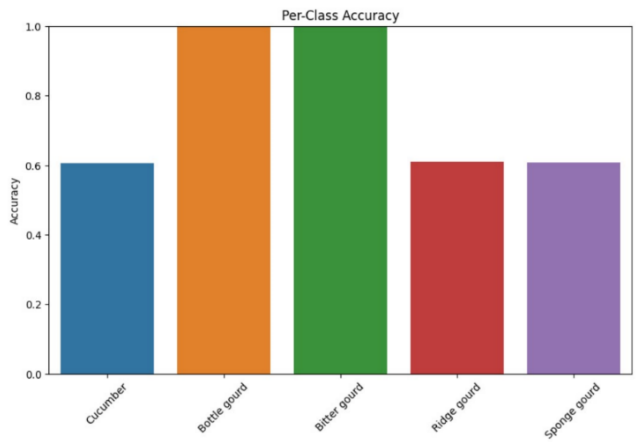


Fig. 7: Class-wise Accuracy

D. Visualization of Class-wise Accuracy

The given bar chart in Fig. 7, illustrates the per-class accuracy distribution. As observed: Bottle Gourd and Bitter Gourd is nearly perfect, with controlled datasets. Accuracy of Cucumber is low as there would be some overlap in features of other vegetables or environmental noise with images.Ridge and Sponge Gourds have intermediate performance and always work perfectly well in controlled settings.

E. Limitations in Open Market Environment

Despite satisfactory performance in controlled environments, the model seems to perform badly in open-market environments. Contributing factors include:

*1) **Unstructured and complex backgrounds**:* Multiple, non-homogenous objects, colors, and crowding confounded the model.

*2) **Lighting conditions**:* Inadequate and inconsistent lighting severely impacted the model's ability to identify and classify gourds correctly, thereby decreasing its reliability.

*3) **Object Occlusion**:* The model could not always see the gourds in their entirety due to market items or other vegetables, which obscured parts of them. It was challenging for the model to spot and accurately draw boxes around them sometimes.

V. CONCLUSION

In this project, we developed a gourd detection system to identify and localize different species of gourds including cucumber, bitter gourd, bottle gourd, ridge gourd, and sponge gourd. Initially, the objective of this project was very clear with the goal of developing a robust dataset and detection system. The data was created by applying subsequent pre-processing and annotation. This involved image acquisition, resizing, normalization, and labeling to represent the target objects accurately.

After preparing the dataset, YOLOv8, object detection model was used for detection and classification. The model performed remarkably well with an overall mAP@50 of 87.2%, thus demonstrating its strength in controlled environments. However, the mAP@50-95 of 73.5% indicated issues concerning localization accuracy, especially when dealing with smaller or overlapping objects. Class-wise performance analysis showed an excellent result in Bottle Gourd and Sponge Gourd with low accuracy in Bitter Gourd and Cucumber due to their complex shape and similarities with other items.

During the real experiment in an open market setting, it was found to be more challenging due to environmental complexities such as brightness, cluttered background and occlusions of objects in the foreground. These limitations underscore the need for training on a diverse and representative data set to improve robustness in uncontrolled environments. The proposed system, notwithstanding these challenges, shows remarkable promise as a scalable approach to automated vegetable classification. Future work will be invested in optimizing the model with real-world conditions, fine-tuning localization capabilities and deploying it on edge computing devices such as Raspberry Pi for real-time applications. This work becomes the basis for progress of autonomous systems in agricultural and retail surroundings.

REFERENCES

[1] S. Khanna, C. Chattopadhyay, and S. Kundu, "Enhancing fruit and vegetable detection in unconstrained environment with a novel dataset," *Scientia Horticulturae*, vol. 338, p. 113580, 2024.

[2] O. Alvarez-Canchila, D. Arroyo-Pérez, A. Patiňo-Saucedo, H. R. González, and A. Patino-Vanegas, "Colombian fruit and vegetables recognition using convolutional neural networks and transfer learning," in *Journal of Physics: Conference Series*, vol. 1547, p. 012020, IOP Publishing, 2020.

[3] P. Sarkar, P. Sarkar, D. Saha, C. Dutta, and A. Maiti, "Indian vegetable image classification using convolutional neural network," in *PREPARE@ u®— FOSET Conferences*, 2022.

[4] H. Biswas and F. Hossain, "Automatic vegetable recognition system,"

[5] F. Xiao, H. Wang, Y. Li, Y. Cao, X. Lv, and G. Xu, "Object detection and recognition techniques based on digital image processing and traditional machine learning for fruit and vegetable harvesting robots: an overview and review," *Agronomy*, vol. 13, no. 3, p. 639, 2023.

[6] F. Femling, A. Olsson, and F. Alonso-Fernandez, "Fruit and vegetable identification using machine learning for retail applications," in *2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pp. 9–15, IEEE, 2018.

[7] C. Sachin, N. Manasa, V. Sharma, and N. K. AA, "Vegetable classification using you only look once algorithm," in *2019 International Conference on Cutting-edge Technologies in Engineering (ICon-CuTE)*, pp. 101–107, IEEE, 2019.

[8] Y. Sakai, T. Oda, M. Ikeda, and L. Barolli, "A vegetable category recognition system using deep neural network," in *2016 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*, pp. 189–192, IEEE, 2016.

[9] M. M. Rahman, M. A. Basar, T. S. Shinti, M. S. I. Khan, H. M. H. Babu, and K. M. M. Uddin, "A deep cnn approach to detect and classify local fruits through a web interface," *Smart Agricultural Technology*, vol. 5, p. 100321, 2023.

[10] M. K. Tripathi and D. D. Maktedar, "Detection of various categories of fruits and vegetables through various descriptors using machine learning techniques," *International Journal of Computational Intelligence Studies*, vol. 10, no. 1, pp. 36–73, 2021.