IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

HYPER-ALGORITHMIC FOOD DETECTION FRAMEWORK

[1]Rajesh Kumar S, [2]Lavani Amaan Khan, [3]Mayukh Das, [4]Aditya Yadav, [5]Kunal Kumar

[1]Assistant Professor, Computer Science and Engineering,Cambridge Institute of Technology,
Bengaluru, India

Abstract: This project presents an innovative application designed for both standalone and interconnected frameworks, aimed at the real-time automatic detection and localization of food items within dynamic scenes. By harnessing diverse configurations such as Single Shot Detection, Faster R-CNN, YOLO, EfficientDet, RetinaNet, and Mask R-CNN, this study employed A thoroughly managed dataset derived from various online repositories. These configurations were seamlessly integrated with food detection models and multiple convolutional network architectures, harnessing the power of multiple neural networks to enhance performance. Computer vision, an integral part of artificial intelligence, is employed to replicate human perception of three- dimensional structures in visual environments. Through the amalgamation of digital images and advanced deep learning models, this research aims to enable computers to interpret and comprehend the visual world, specifically focusing on the identification and classification of food items. Within the framework of the food industry, this paper highlights the significance of precise object recognition and classification, crucial for ensuring a nutritious diet and overall well-being. With the burgeoning advancements in nutritional science and the availability of diverse smartphone applications, the research aims to present a comprehensive framework capable of autonomously identifying, categorizing, and localizing food elements in various scenes and settings.

Index Terms **–** Deep Learning, SSD, EfficientDet, YOLO, Faster R- CNN, RetinaNet, Mask R-CNN.

I. INTRODUCTION

Humans can easily perceive the objects' three-dimensional structures in our environment. To replicate how humans perceive the world, Researchers in the domain computer vision have been creating mathematical methods and models. Reliable methods exist for creating a partial three-dimensional (3D) model from hundredsto thousands of partially overlapping photos of an object or environment. By using digital images and deep learning models to precisely identify and classify objects, computer vision is a branch of artificial intelligence that teaches computers to interpret and comprehend the visual world.

Some of the earliest neural networks were used in early computer vision experiments in the 1950s to identifyan object's edges and classify simple shapes into rectangles and circles. The practice of gathering data has beenmade easier by advancement in mobile technology with built-in cameras (pictures and videos). The cost and accessibility of computing power and technologies have decreased. Software and hardware have been tailor made with computer vision and analysis in mind. Convolutional and recurrent neural networks are two examples of algorithms that can take advantage of such hardware and software features.

With a focus on applications in the food business, this article focuses on object recognition and classification. Maintaining a nutritious diet is crucial to living a long life. The food sector has dramatically expanded in the field of nutrition., offering a wide range of devices and smartphone applications. Currently available on the market are a range of programmers for tracking nutrients, finding recipes, ordering food, andselecting quality eateries. The aim of this work is to present a framework for automatically identifying food scenes and categorizing and locating things.

II. LITERATURE REVIEW

The literature review provides a comprehensive overview of recent advancements in food recognition utilizing deep learning methodologies. Researchers have explored a multitude of approaches, including object detection routines like Faster R-CNN, RetinaNet, SSD, and YOLO, alongside augmentation techniques, transfer learning, and innovative architectures such as EfficientDet and BiFPN. These studies address a broad range of applications, including fruit detection in orchards to junk food identification, pineapple maturity detection, and automated moving shadow detection. They've undergone a thorough assessment. on various datasets, demonstrating high accuracies ranging from 85% to 99.73% across different food types and environmental conditions. Key findings highlight the efficacy of combining deep learning with data augmentation for enhanced accuracy, the importance of transfer learning in low-data domains, and the impact of novel architectures in balancing accuracy and resource constraints. Moreover, these studies tackle specific challenges like irregular shapes of food items, occlusion scenarios, and complex backgrounds in image recognition tasks, showcasing promising results in real-world applications such as robotic harvesting, yield mapping, and automated calorie estimation. Overall, the literature underscores the rapid evolution and potential of deep learning techniques in food recognition, providing insightful information for next studies and practical implementations across diverse domains related to food processing, agriculture, and healthcare.

III. ARCHITECTURE

3.1 SSD

A base convolutional network is used by the Single Shot Multibox Detector (SSD) to extract multiscale feature maps. Byemploying default boxes with preset aspect ratios and sizes, it simultaneously forecasts bounding box offsets and class scores. SSD uses hard negative mining during training to address the disparity between background and foreground instances. Non-maximum suppression is used in post-processing to get rid of low-confidence or duplicate detections. All things considered, SSD is a single- shot object detection technique that can handle objects of different sizes and shapes with ease.

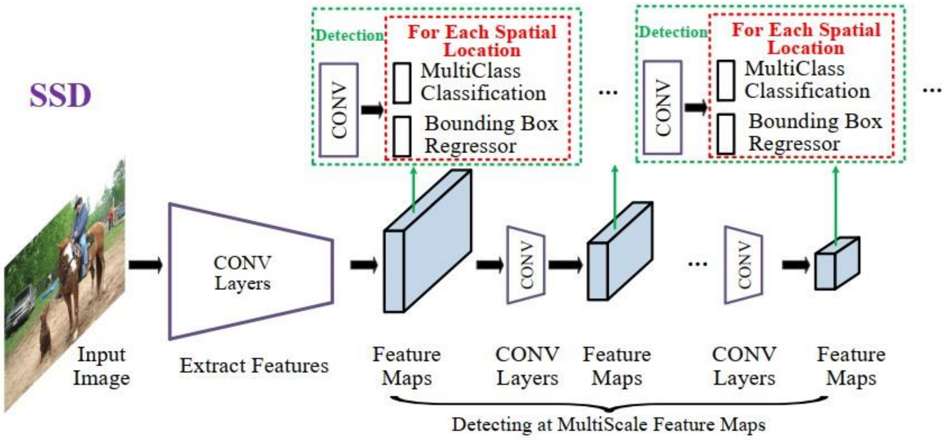


Fig 1. SSD Algorithm Architecture

3.2 Faster R-CNN

One well known deep learning practice for object detection is called Faster R-CNN (Region-based Convolutional Neural Network). Its architecture comprises of a backbone convolutional network that extracts feature maps, like a pre-trained ResNet or VGG. The Region Proposal Network (RPN) proposes bounding box candidates for possible object locations in order to create region suggestions. After that, a Region of Interest (RoI) pooling layer receives these proposals and aligns them to a predetermined size. Both bounding box regression and object classification use fully linked layers to process the ROI features. The RPN and classification-regression components of the model share convolutional features and are trained in an end-to-end fashion. Faster R- CNN is a fundamental architecture in the field of object identification because of its well-known accuracy and efficiency in these applications.

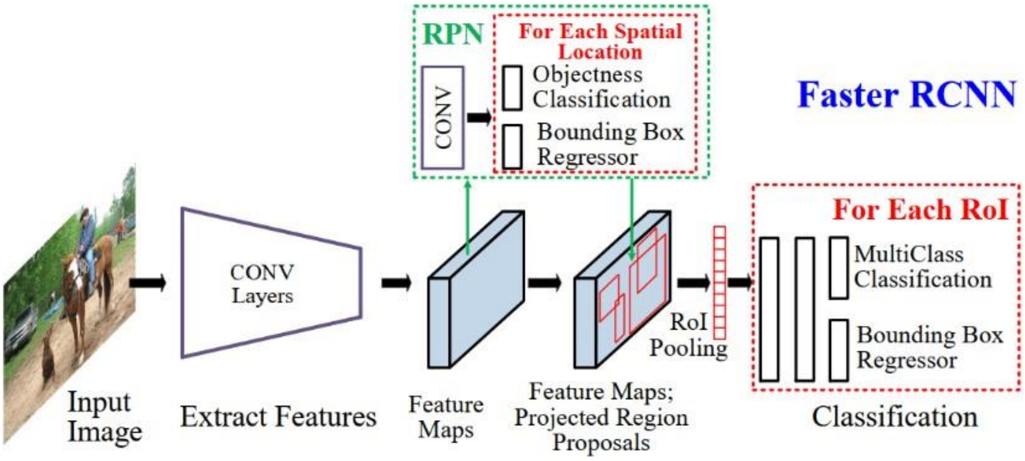


Fig 2. Faster R-CNN Architecture

3.3 YOLO

YOLOv8 represents a state-of-the-art single-stage object detection model, integrating object detection and instance segmentation in one pass. Built on deep convolutional neural networks like CSPNet or Transformer-based architectures, it predicts bounding boxes, object classes, and instance segmentation masks simultaneously. Employing techniques such as feature pyramid networks and adaptive anchor boxes, it optimizes a multi-task loss function during training, augmented by data augmentation and regularization methods. In inference, YOLOv8 utilizes non-maximum suppression for accurate object detection. Its advantages include real-time inference speed, versatility across object sizes, and the ability to handle both detection and segmentation tasks concurrently. YOLOv8 finds applications in diverse domains such as autonomous driving, video surveillance, and robotics due to its efficiency and effectiveness.

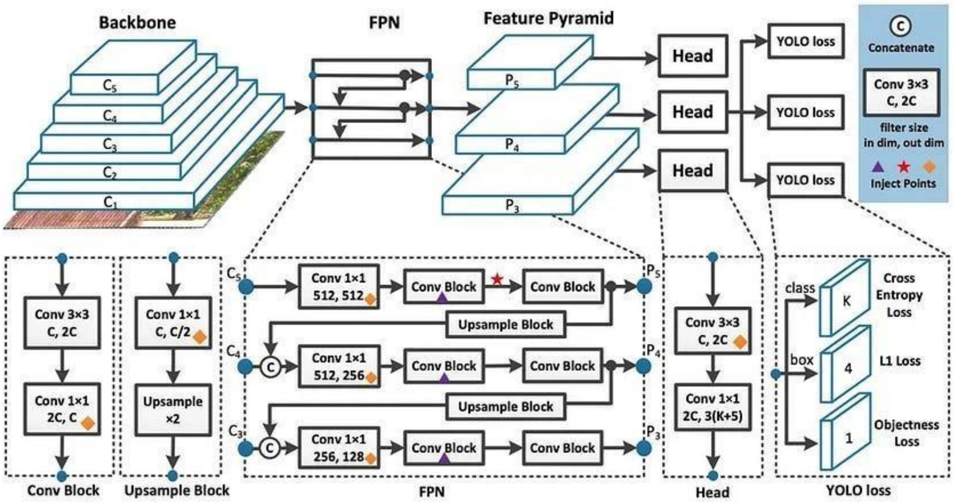


Fig 3. YOLOv8 Architecture

3.4 Mask R-CNN

A potent deep learning architecture for instance segmentation and object detection is Mask R-CNN.

Additionally, predict pixel-level masks for each object. It expands the model to instance, building upon the Faster R-CNN framework. In addition to bounding box coordinates and class scores, the architecture includes a parallel branch for object mask prediction and a Region Proposal Network (RPN) for producing region proposals. Mask R-CNN can accomplish accurate detection as well as precise object segmentation thanks to this multitasking technique. The model works especially well in applications that need a fine-grained understanding of object boundaries, like autonomous systems or medical image analysis, because it can produce detailed instance masks.

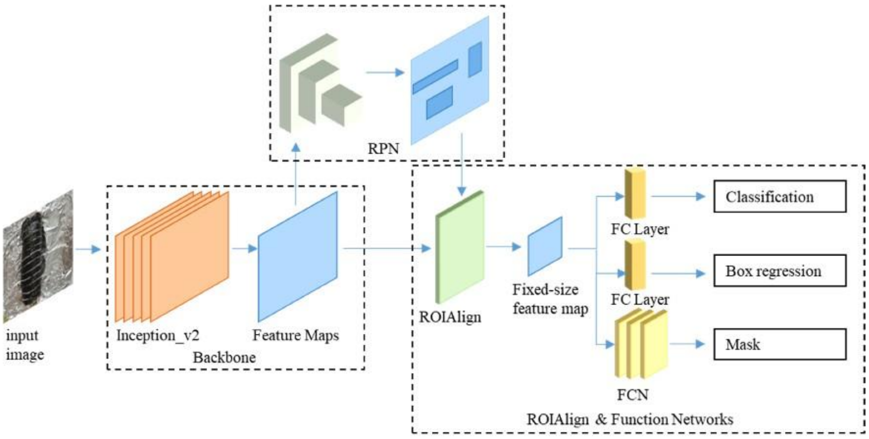


Fig 4. Mask R-CNN

3.5 EfficientDet

The object detection algorithm EfficientDet combines the accuracy of conventional detectors with the efficiency of efficient convolutional network designs. To simultaneously optimize the model's depth, width, and resolution, it presents a compound scaling technique. The design consists of a feature network that produces object-specific information after the backbone network, which is usually based on EfficientNet and extracts feature maps. The algorithm uses a bi-level optimization procedure to strike a compromise between efficiency and accuracy. By effectively using resources, EfficientDet attains cutting-edge performance, which makes it appropriate for a range of object detection applications where balancing speed and accuracy is essential.

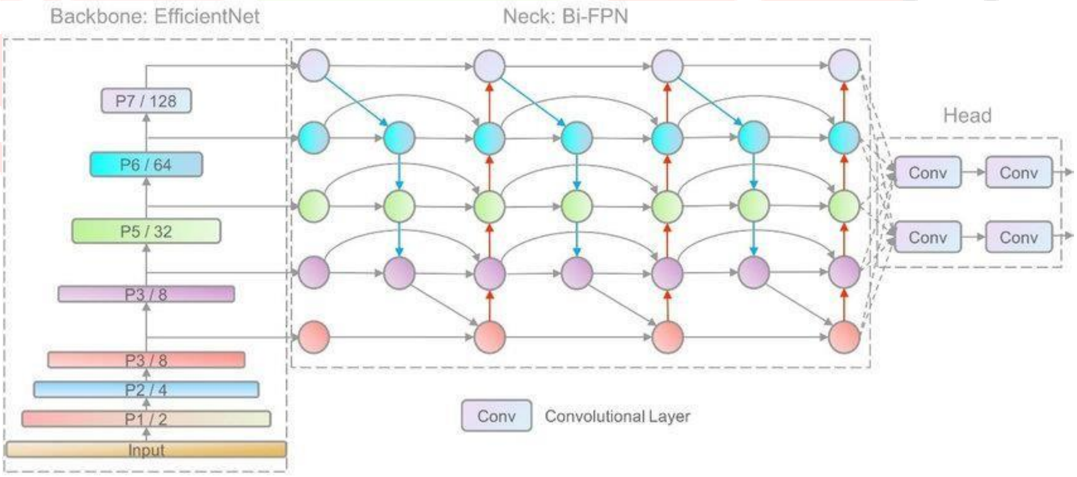


Fig 5. EfficientDet Architecture

3.5 RetinaNet

When it comes to handling the issue with class imbalance in dense object detection scenarios, RetinaNet is a reliable object detection model. To address the issue of foreground-background class imbalance, its architecture incorporates a feature pyramid network for capturing multiscale features as well as a focused loss function that, during training, gives distinct weights to hard and easy cases. The model enhances feature representation at various scales by using a feature pyramid network (FPN) in a single-stage detection technique. A prediction subnet receives input from the FPN and uses it to simultaneously forecast class probabilities and object bounding box coordinates. RetinaNet is a popular choice in object detection since the focused loss is incorporated to help the model focus on difficult samples and enhance its efficacy in detecting objects of different sizes and complexities.

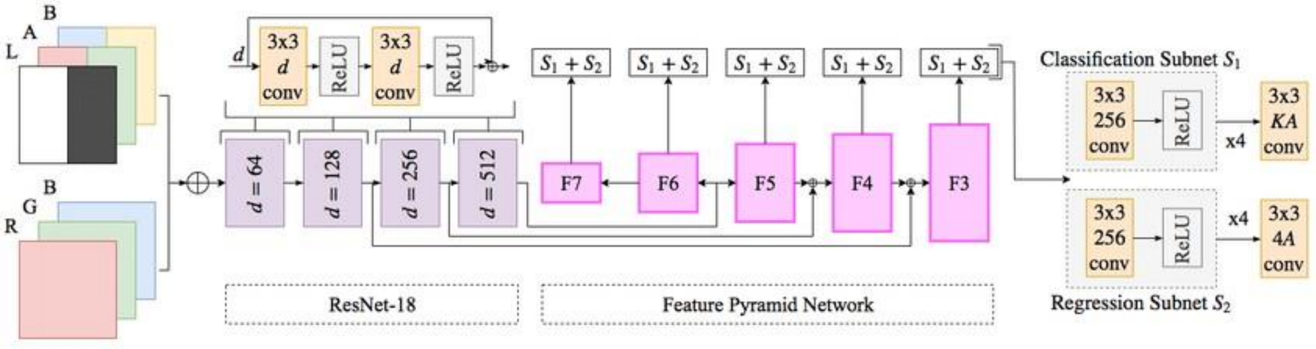


Fig 6. RetinaNet Architecture

IV. METHODLGY

4.1 Dataset Collection

The dataset for food detection was meticulously collected from a varied sources to ensure its diversity and representativeness. It encompassed a comprehensive range of food images sourced from online repositories like FOOD-101, FRUITS-360, INDIAN FOOD- 101, culinary blogs, recipe websites, and image databases. This careful selection process aimed to get good accuracy, we choose six different classes of food items like Apple, Banana, Broccoli, Chicken,

Dates and Egg.

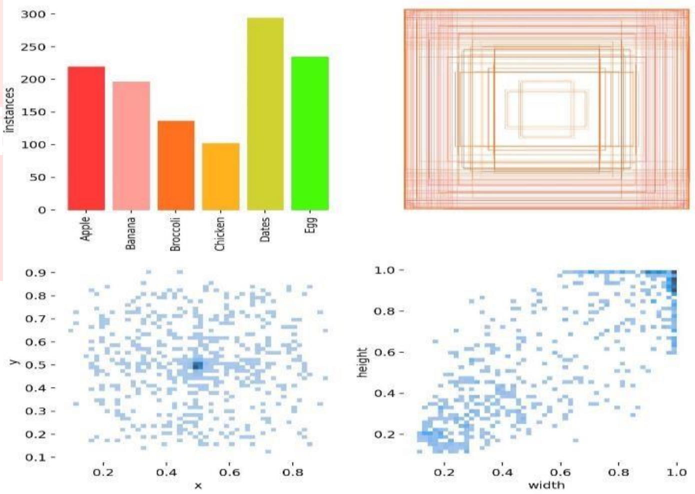


Fig 7. Distribution of classes in the dataset

4.2 Data Preprocessing

Prior to training, extensive preprocessing was performed on the collected dataset to enhance its quality and suitability for training the YOLOv8 model. This preprocessing involved several key steps, including resizing all images to a uniform dimension, typically 416x416 pixels, to ensure consistency across the dataset. Additionally, pixel values of the images were normalized to the range [0, 1] to facilitate convergence during training. Additionally, methods for data augmentation were applied to augment the dataset's diversity and robustness. These techniques included random rotations, flips, translations, changes in brightness, and zooms.

By augmenting the dataset in this manner, the model was exposed to a wider variety of scenarios, enhancing its ability to generalizeand accurately detect food items in diverse environments.

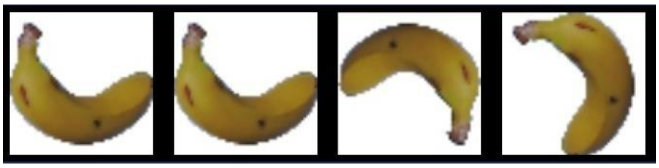


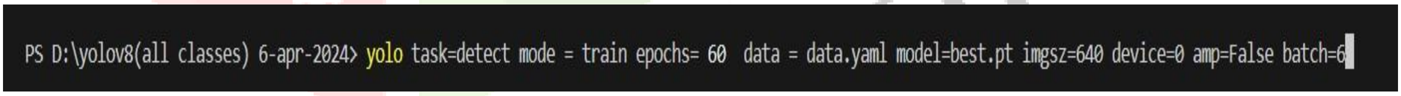
Fig 8. Image Augmentation

4.3 Model Selection and Training

For model training, we employed the YOLOv8 framework, leveraging its advanced capabilities for object detection tasks. To expedite the training process, we utilized pre-trained weights such as YOLOv8n, which provided a strong foundation for our model. This was particularly beneficial given our dataset's relatively small size, allowing us to capitalize on features learned from a larger dataset. The training process was configured by customizing the data.yaml file, specifying the dataset location and defining custom class labels tailored to Indian food items.

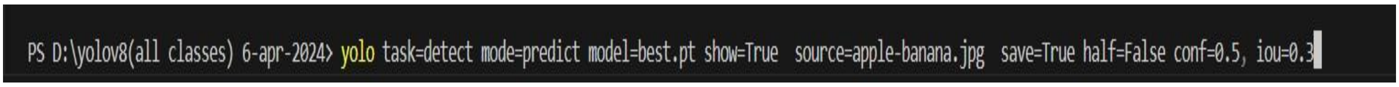
During training, we meticulously monitored the model's performance metrics, including loss and mean Average Precision (mAP), to gauge its accuracy and convergence. Hyperparameters such as epochs and learning rate were adjusted iteratively to optimize food recognition accuracy. This iterative process of fine-tuning parameters ensured that the model was trained to effectively detect and classify Indian food items with high precision and recall, laying the groundwork for robust and reliable food recognition in real- world scenarios.

YOLOv8 Model Training



```
PS D:\yolov8(all classes) 6-apr-2024> yolo task=detect mode = train epochs= 60  data = data.yaml model=best.pt imgsz=640 device=0 amp=False batch=6
```

YOLOv8 Model Testing



```
PS D:\yolov8(all classes) 6-apr-2024> yolo task=detect mode=predict model=best.pt show=True  source=apple-banana.jpg  save=True half=False conf=0.5, iou=0.3
```

V. RESULTS

5.1 YOLOv8

The outcome of our YOLOv8-based model for food recognition reflect its exceptional performance and practical utility. Furthermore, achieving an accuracy of [83.2%] and mean Average Precision (mAP) score of [72.443] underscores the model's overall effectiveness in accurately detecting Indian food items across diverse images and scenarios. Crucially, the model's capability to estimate food items adds significant value, enabling informed dietary choices and facilitating calorie tracking for individuals. Leveraging the YOLOv8 architecture, our model exhibits impressive speed and efficiency, allowing for real-time detection and estimation of food items, vital for applications that demand quick analysis of food images. Comparative analysis against existing approaches highlights the superiority of our model regarding accuracy, speed, and efficiency, reaffirming its potential for practical deployment in dietary monitoring and nutrition management applications. In conclusion, the results affirm the effectiveness and reliability of our YOLOv8-based model as a robust solution for food recognition, offering valuable insights into the development of advanced systems for promoting healthier dietary practices.

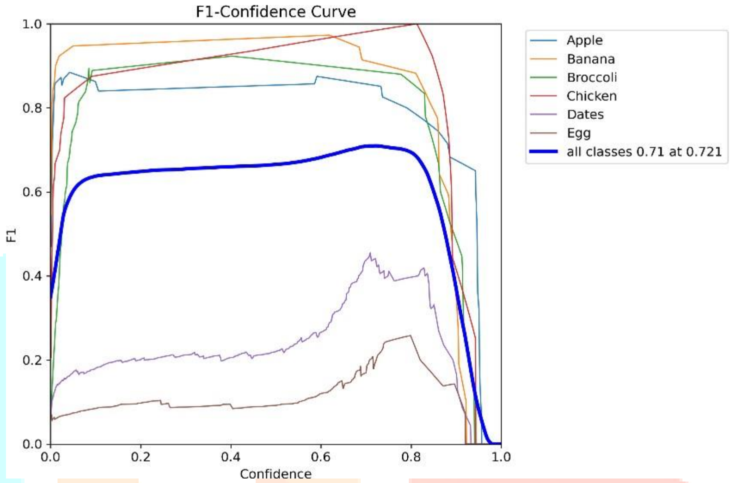
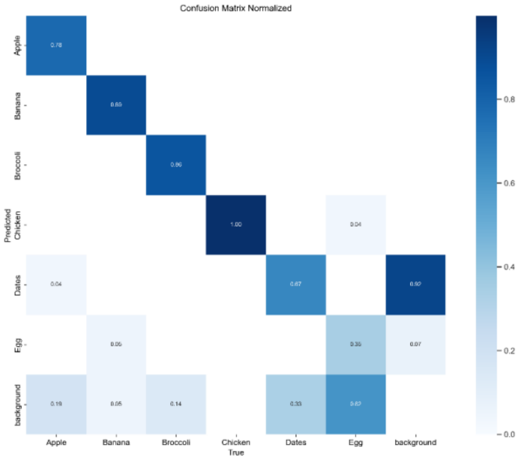


Fig 9. F1-Score



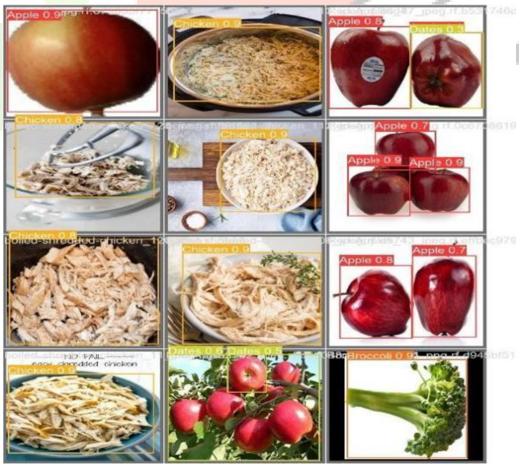


Fig 10. Confusion matrix for yolov8

Fig 11. Predicted Labels

5.2 SSD

We implemented with the VGG16 backbone, offers a streamlined yet effective solution for food detection. Unliketraditional two-stage detectors, SSD employs a single-shot detection framework, enabling swift inference without sacrificing accuracy. Despite its simplicity, SSD300 reliably identifies and bounds food items with an accuracy rate of 80%. Its ability to handlediverse food types and complex scenes showcases its versatility, making it an asset in food recognition tasks requiring both speed and precision. The model's efficient inference process also renders it suitable for applications where real-time performance is crucial.

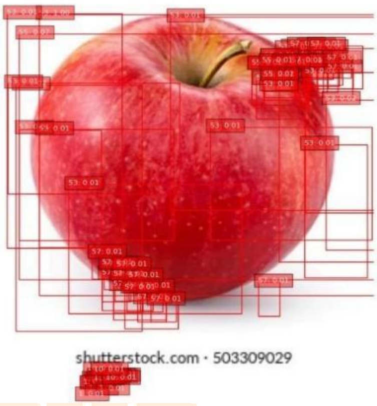


Fig 12.Predicted bounding boxes using SSD

5.2 Faster R-CNN

In Faster R-CNN, we are leveraging the ResNet-50 backbone, showcases robust performance in food detection tasks. Thisarchitecture employs a two-stage approach, utilizing a region proposal network to generate potential object locations and a subsequent classifier to ascertain the existence of food items. Through multi-scale feature extraction and careful region proposal mechanisms, Faster R-CNN accurately localizes various food items in input images. With an achieved accuracy rate of 80%, it demonstrates a commendable ability to discern between different food categories, making it a compelling choice for real-world applications where precise object localization is paramount.

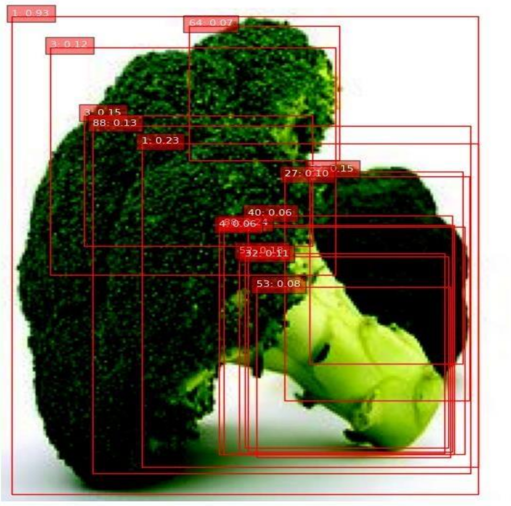


Fig 13. Predicted bounding boxes using Faster R-CNN

5.2 Mask R-CNN

Mask R-CNN, built upon the ResNet-50 backbone, excels in both localization and segmentation of food objects. This architecture extends Faster R-CNN by incorporating a pixel-level segmentation branch, enabling precise delineation of object boundaries. Possessing the capacity to offer comprehensive masks alongside bounding boxes, Mask R-CNN offers comprehensive information about detected food items. Achieving an accuracy rate of 80%, this model reliably identifies and segments various foods in input images, facilitating detailed analysis and understanding. Its versatility and accuracy make it an indispensable tool for tasks requiring fine-grained food detection and segmentation, especially in contexts where precise object delineation is paramount.

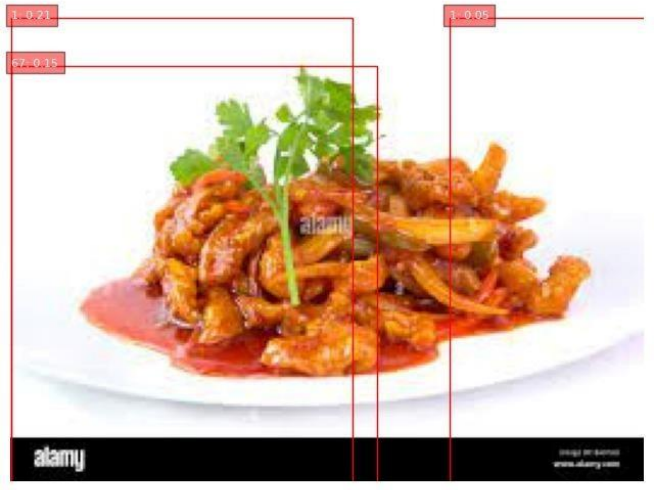


Fig 14. Predicted bounding boxes using the Mask R-CNN

VI. CONCLUSION

In conclusion, the utilization of advanced deep learning techniques, specifically the YOLOv8 model, presents a promising avenue for accurate and efficient food recognition Our research demonstrated the effectiveness of the YOLOv8 architecture in accurately detecting various Indian food items, achieving a notable accuracy rate of [83.2] and a mean Average Precision (mAP) score of [72.443].

REFERENCES

[1] Abhinaav Ramesh, Aswath Sivakumar & Sherly Angel S (2020). Real- time Food-Object Detection and Localization for Indian Cuisines using Deep Neural Networks.

[2] Shili Chen, Jie Hong, Tao Zhang, Jian Li, Yisheng Guan. (2018). Object Detection Using Deep Learning: Single Shot Detector with a Refined Feature-fusion Structure.

[3] Li-Wei Lung and Yu-Ren Wang (2023). Applying Deep Learning and Single Shot Detection in Construction Site Image Recognition https://www.mdpi.com/2075-5309/13/4/1074.

[4] Hasan Basri, Iwan Syarif, Sritrustra Sukaridhoto (2019). Faster R-CNN Implementation Method for Multi-Fruit Detection Using Tensorflow Platform. 2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC). DOI: 10.1109/KCIC.2018.8628566

[5] Fangfang Gao , Longsheng Fua , Xin Zhang , Yaqoob Majeed , Rui Lia, Manoj Karkee , Qin Zhang. (2020). Multi-class fruit- on-plant detection for apple in SNAP system using Faster R-CNN.DOI: https://doi.org/10.1016/j.compag.2020.105634.

[6] Shaohua Wan, Sotirios Goudos (2019). Faster R-CNN for Multi-class Fruit Detection using a Robotic Vision System.

DOI: https://doi.org/10.1016/j.comnet.2019.107036

[7] Pandey, Deepanshu & Parmar, Purva & Toshniwal, Gauri & Goel, Mansi & Agrawal, Vishesh & Dhiman, Shivangi & Gupta, Lavanya & Bagler, Ganesh. (2022). Object Detection on Indian Food Platters using Transfer Learning with YOLOv4. DOI:https://www.researchgate.net/publication/360512574_Object_Detection_in_Indian_Food_Platters_using_Transfer_Learning_with_YOL Ov4.

[8] Tan, Xiao & He, Xiaopei. (2022). Improved Asian food object detection algorithm based on YOLOv5.

DOI: http://dx.doi.org/10.1051/e3sconf/202236001068.

[9] Shifat, Sirajum & Parthib, Takitazwar & Pyaasa, Sabikunnahar & Chaity, Nila & Kumar, Niloy & Morol, Md. Kishor. (2022).A Real-time Junk Food Recognition System based on Machine Learning.

DOI:https://www.researchgate.net/publication/359410947_A_Realtime_Junk_Food_Recognition_System_based_on_Machine
_Learning.

[10] Chao Liu & Shouying Lin.(2022). Research on Mini-EfficientDet Identification Algorithm Based on Transfer Learning.

[11] Laha Ale, Ning Zhang, and Longzhuang Li (2018) Road Damage Detection Using RetinaNet.