JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Real-Time Food Detection and Localization Using Deep Learning: A Comparative Study of Single Shot Detection, Faster R-CNN, YOLO, EfficientDet, RetinaNet, and Mask R-CNN.

Rajesh Kumar S, Cambridge Institute of Technology, Visvesveraya Technological University, Belagavi 590018.
Dr. Josephine Prem Kumar, Cambridge Institute of Technology, Visvesveraya Technological University, Belagavi 590018.

Abstract: The goal of this paper is to create an application that can be used in a standalone or connected application framework that can automatically detect and localise food objects in real-time scenes. The Single Shot Detection, Faster R-CNN, YOLO, EfficientDet, RetinaNet, and Mask R-CNN configurations were trained using a dataset that was assembled from various online sources. These algorithms were paired with the food detection model and multiple convolutional network architectures; in this case, multiple neural networks will be used. We have presented a few deep learning-based techniques for food detection in this paper.

Keywords: Deep Learning, SSD, EfficientDet, YOLO, Faster R-CNN, RetinaNet, Mask R-CNN

I. INTRODUCTION

Humans are able to easily perceive the three-dimensional structures of objects in our environment. In order to replicate how humans perceive the world, computer vision researchers have been creating mathematical methods and models. Reliable methods exist for creating a partial three-dimensional (3D) model from 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 first neural networks were used in early computer vision experiments in the 1950s to identify an object's edges and classify simple shapes into rectangles and circles. The process of gathering data has been made easier by 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 specifically created with computer vision and analysis in mind. Convolutional and recurrent neural networks are two examples of algorithms that can benefit from 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 industry has been rapidly expanding in the subject 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, and selecting quality eateries. The goal of this work is to present a framework for automatically identifying food scenes and categorizing and locating things.

REVIEW OF LITERATURE

The combination of picture selection approaches, preprocessing methods, segmentation methods, and recognition models determines the ability to recognize food from an image. The following papers use various deep learning techniques to improve recognition and accuracy when performing character recognition using different methods.

Abhinaav Ramesh, Aswath Sivakumar & Sherly Angel S,2020 The researchers identified several food kinds in photos by combining deep learning with picture augmentation. With one incorrect forecast for every 250 correct predictions, the model's practical accuracy was 97.6%. In both informal and professional contexts, this could be helpful for accuracy and resource management [1].

Shili Chen, Jie Hong, Tao Zhang, Jian Li & Yisheng Guan, 2018 The paper presents the RFSSD method for objectdetection, evaluated on the PASCAL VOC2007 and PASCAL VOC2012 datasets. RFSSD was trained with 16 batch size and an input image size of 300x300 using TensorFlow framework. It outperformed other refined SSD networks by increasing the accuracy by up to 0.63%. However, the overall accuracy of RFSSD was not provided in the uploaded file [2].

Li-Wei Lung & Yu-Ren Wang, 2023 The study utilized a dataset of 461 images, with 80% used for training, and 40 images each for testing and verification. The research concentrated on item recognition and categorization of rebar, worker, and machine classes using the Single Shot Detector (SSD) approach. With 461 photos from building sites, the SSD model's accuracy overall was 92% and its F1 measure was 85% [3].

Hasan Basri, Iwan Syarif, Sritrustra Sukaridhoto, 2018 suggests detecting many fruits using a faster R-CNN deep learning approach. With 99% accuracy, it makes use of the TensorFlow platform's MobileNet model. This method works well for sorting fruit in real time [4].

Fangfang Gao , Longsheng Fua , Xin Zhang , Yaqoob Majeed , Rui Lia , Manoj Karkee , Qin Zhang, 2020 provides a Faster R-CNN-based multi-class fruit-on-plant identification technique for apple trees. Various apple occlusion situations are taken into consideration, including fruit- and branch-occluded situations. The suggested approach can help with robotic picking techniques and achieves high accuracy [5].

Shaohua Wan, Sotirios Goudos, 2019 proposes a deep learning framework based on Faster R-CNN for multi-class fruit detection in orchards. The framework includes creating a fruit image library, data augmentation, and improved modelgeneration. The proposed method achieves high accuracy andcan be applied to robotic harvesting and yield mapping systems [6].

Deepanshu Pandey et. al., 2022 presented the TR-YOLOv4 object detection technique using Transfer Learning with YOLOv4, which was assessed using the IndianFood10 and IndianFood20 datasets. Using the PyTorch framework, TR-YOLOv4 was trained using an input image size of 416x416 and a batch size of 32. It increased the mAP score by up to 4.7%, outperforming previous transfer learning strategies. TR-YOLOv4's total mAP score was 90.7% on IndianFood20 and 91.8% on IndianFood10 [7]

Xioa Tan & Xiaopie He, 2022 developed an an improved YOLOv5 algorithm for object detection in Asian food images. The proposed algorithm, named YOLOv5-Asia, wasevaluated on the AFD100 dataset and achieved a mean average precision (mAP) score of 94.2%, outperforming other state-of-the-art object detection methods. YOLOv5- Asia was trained with a batch size of 64, an input image sizeof 416x416, and a learning rate of 0.001 using the PyTorch framework [8].

Sirajum Munira Shifat et. al., 2020 presented a deep learning-based system for detecting junk food items inimages using the YOLOv3 object detector. The proposedsystem was trained on a custom dataset of 10,000 junk foodimages and achieved an accuracy of 98.05% on a test set of 1,000 junk food images. The system was implemented using the Darknet framework and can be used in real time to detect junk food items in images captured by smartphones [9].

Merugu Sai Teja, Mr. B. Nageswara Rao, Mannem VinayReddy, K. Praveen Kumar, M. Sai Kumar, 2022 discussesthe implementation of EfficientDet, a state-of-the-art object detection model. It uses a compound scaling method to balance network depth, width, and image resolution for improved accuracy within resource constraints. EfficientDet relies on a BiFPN (bidirectional feature pyramid network) and class/box prediction networks. It's based on EfficientNetas the backbone network for feature extraction [10].

Chao Liu & Shouying Lin., 2022 explores beehive farmingin Fujian Province, China. Few studies have been done on image recognition in this field. The common problems in thefield of computer image recognition in agricultural farming are mainly in the irregular shape of recognition targets and the large number of targets contained in the images. The improved EfficientDet model was trained by transfer learningwith a homemade dataset, and was able to distinguish Chinesewasps, hornets, and cockroaches in worm form efficiently andaccurately [11].

M. Y. Cao & J. Zhao, 2022 discusses the challenges of pedestrian detection in intricate settings and examines conventional & deep learning-based methods for pedestrian detection. The proposed Fast-EfficientDet algorithm utilizes a new backbone network, a new feature pyramid network, andDIoU calculation in NMS to improve detection accuracy andspeed [12].

Luyang Zhang, HaitaoWang, XinyaoWang, Shuai Chen, Huaibin Wang, Kai Zheng and Hailong wang, 2020 The abstract of the paper describes an improved RetinaNet for vehicle object detection in real scenes. The suggested approach consists of an octave convolution structure and a weighted feature pyramid network to enhance the detection performance of RetinaNet. The authors conducted experiments on the DETRAC dataset and achieved promising results. The suggested approach consists of outperformed the baseline RetinaNet in terms of detection accuracy, especially in low-resolution scenes. The findings indicate that the improved RetinaNet has potential real-world applications in vehicle object detection [13].

Yan Chen, Lulu Zheng, Hongxing Peng, 2023 The paper presents an improved RetinaNet algorithm, ECA-RetinaNet, which effectively detects pineapple maturity levels. The dataset, collected from a natural orchard in China, comprises 6,000 images and 30 videos captured under various environmental conditions. The ECA-RetinaNet algorithm demonstrates high accuracy, with a Mean Average Precision (mAP) ranging from 85.18% to 99.73% for different maturity levels. This indicates its effectiveness in accurately identifying pineapples at various stages of maturity, particularly in complex scenarios. The study's findings highlight the algorithm's potential for practical applications in yield estimation and automatic picking research and development [14].

Mohanad N. Alhasanat, Moath H. Alsafasfeh, Abdullah E. Alhasanat, Saud G. Althunibat, 2021 The abstract of the paper outlines a RetinaNet-based approach for object detection and distance estimation in digital images, presenting a novel method for accurately detecting and estimating the distance of objects in images. The authors conducted real experiments and achieved high accuracy with an average 5% error rate for different distances, attributing the error to factors such as the widths of the detected objects' boundary boxes and the manner in which the camera was held during experiments. The proposed method leverages the RetinaNet algorithm for object detection and distance estimation, demonstrating its effectiveness in real-world scenarios [15].

Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, 2015 The paper introduces SSD, a single deep neural network for object detection, achieving state-of-the-art results on benchmark datasets. Results are reported on ILSVRC DET and PASCAL VOC datasets, comparing against other methods. SSD achieves a mean average precision (mAP) of 74.3% on PASCAL VOC 2007 and competitive mAP on ILSVRC 2016 detection task. SSD uses a single network to predict class scores and bounding boxes, trained with hard negative mining and data augmentation [16].

Shuqin Tu, Jing Pang, Haofeng Liu, Nan Zhuang, Yong Chen, Chan Zheng, Hua Wan, Yueju Xue, 2020 The paper suggests a novel approach using multiple scale faster R-CNN with RGB-D images for accurate and reliable passion fruit detection and counting in orchards, addressing challenges in small fruit detection [17].

Laha Ale, Ning Zhang, and Longzhuang Li, 2018 RetinaNet models with VGG and ResNet152 backbones achieved the highest accuracy, as indicated by the mean Average Precision (mAP) on validation samples in the study [18].

Mingxing Tan Ruoming Pang Quoc V. Le Google Research, Brain Team,2020
The design of neural networks for object detection is examined in this research, introducing BiFPN and compound scaling to create EfficientDet models. These enhancements yield superior efficiency, achieving top accuracy with reduced parameters. Implementation code is available online [19].

Hend F. Bakr1, Ahmed M. Hamad2, and Khalid M. Amin3 ,2021
This research implements Mask R-CNN for automated moving shadow detection and segmentation in computer vision, leveraging Convolutional Neural Network (CNN) to learn diverse shadow features across environments. Achieving high performance, it surpasses existing methods [20].

Shuqi Fang, Bin Zhang * and Jingyu Hu,2023
This research enhances Mask R-CNN for multi-target detection in complex traffic scenes. It integrates ResNeXt, FPN enhancement, an efficient channel attention module, and a modified loss function, achieving superior detection (62.62% mAP) and segmentation (57.58% mAP) on CityScapes dataset [21].

V.Gayatri, M.Thanuja ,2023
This paper presents a YOLOv8-based system for Indian food recognition and calorie estimation, achieving a notable accuracy of 93.1% on a dataset of 5446 images across 30 classes. The model, trained with Streamlit integration, enables fast and accurate detection of food items and estimation of calorie values per gram [22].

Triphena Delight D, Karunakaran V,2021
This paper introduces a powerful object detection approach using Mask RCNN, demonstrating its efficacy on the Plasmodium Vivax dataset for malaria detection with a notable 94% mAP. The model's adaptability is showcased through training on a custom dataset for shape recognition. Comparative evaluations highlight Mask RCNN's superior performance in object detection [23].

Than Le ,2020
In order to recognize food data, this research investigates deep learning techniques with a focus on detection and visualisation. It examines how training datasets are affected by data augmentation and transfer learning, using performance criteria including accuracy and time. The results offer insightful information for improving deep learning models in applications involving food recognition [24].

III. ARCHITECTURE

SSD: With reference to fig.1, A base convolutional network is used by the Single Shot Multibox Detector (SSD) to extract multiscale feature maps. By employing 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.

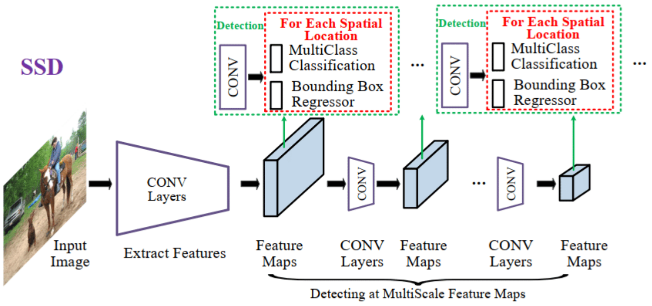


Fig .1. SSD Algortihm Architecture

Faster R-CNN: With reference to fig.2, One popular deep learning technique 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.

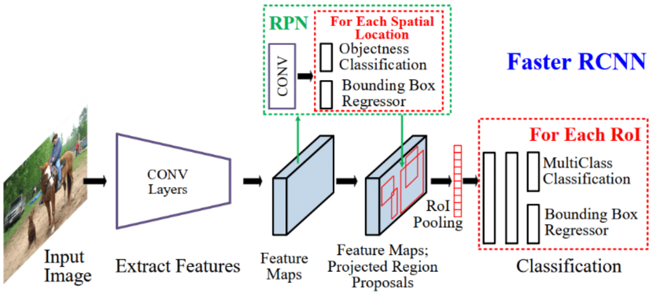


Fig .2. Faster R-CNN Architecture

YOLO: With reference to fig.3, One well-known deep learning technique for object detection is called You Only Look Once (YOLO), and it is notable for its ability to handle data in real time. The incoming image is divided into a grid using YOLO, which then projects bounding boxes and class probabilities at each grid cell. One neural network serves as the architecture's sole predictor of several bounding boxes and the class probabilities that go along with them. Because it only needs to make one forward pass through the neural network in order to provide predictions for the full image, YOLO is renowned for its efficiency and speed. The technique is well-suited for real-time applications like video analysis and driverless vehicles since it can handle numerous object instances in a single pass.

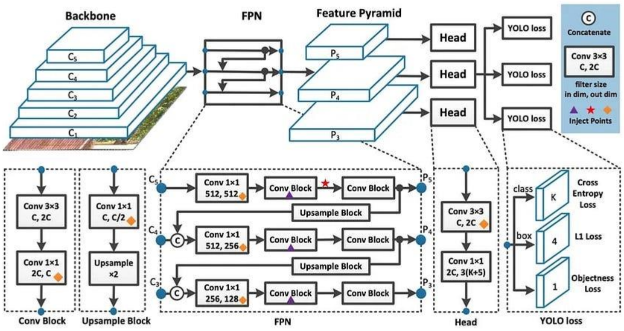


Fig .3. Yolov8 Architecture

EfficientDet: With reference to fig.4, The object detection algorithm EfficientDet combines the accuracy of conventional detectors with the efficiency of efficient convolutional network designs. In order to simultaneously optimise 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 optimisation 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.

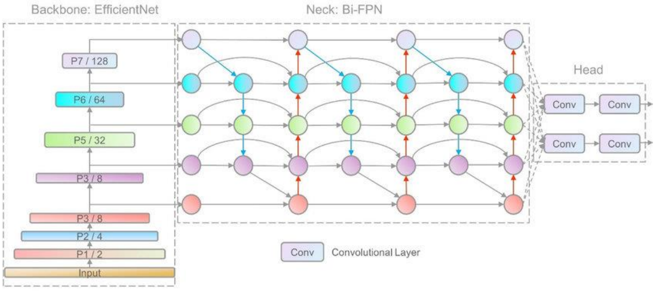


Fig 4. EfficentDet Architecture

RetinaNet: With reference to fig.5, When it comes to handling the problem of class imbalance in dense object detection scenarios, RetinaNet is a reliable object detection model. In order 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 performance in detecting objects of different sizes and complexities.

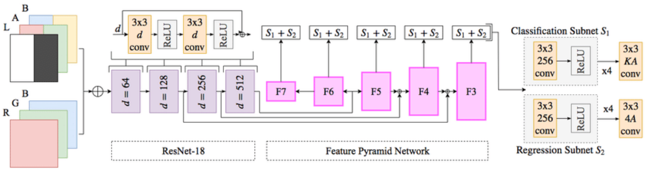


Fig .5. RetinaNet Architecture

Mask R-CNN: With reference to fig.6, 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.

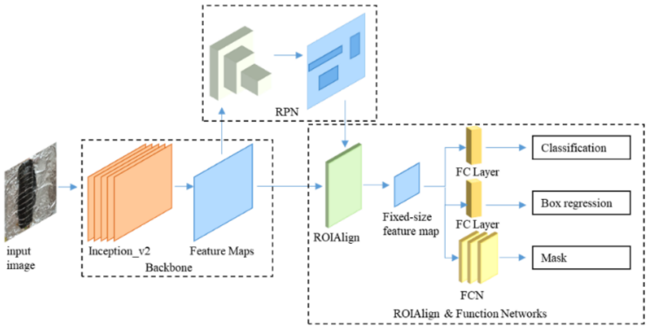


Fig .6. Mask R-CNN Architecture

IV. METHODLOGY

Dataset Collection

With reference to the fig.7, The dataset for food detection was meticulously collected from a variety of 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.

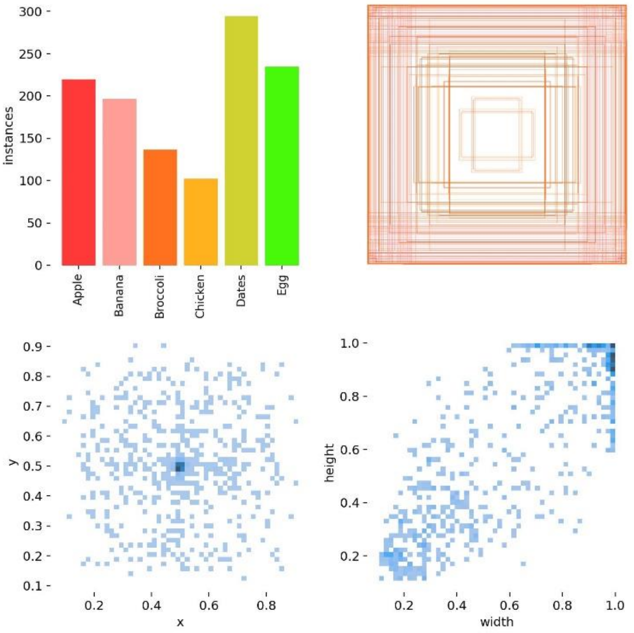


Fig.7. Distribution of classes in the dataset.

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. Furthermore, With reference to fig.8, data augmentation techniques 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 generalize and accurately detect food items in diverse environments.

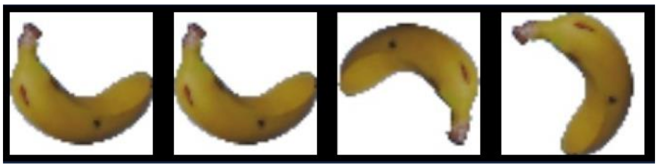


Fig.8. Examples of Image Augmentation.

Model Selection and Training

Yolov8

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 or YOLOv8s, which provided a solid starting point 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.

SSD

For SSD, we opted for its efficient single-shot architecture, which is particularly suitable for real-time object detection tasks. Utilizing SSD's ability to predict object bounding boxes and class probabilities in a single pass, we employed pre-trained models such as SSD-MobileNet and SSD-ResNet for faster convergence. With a carefully curated dataset of Indian food items, we fine-tuned the SSD model by adjusting hyperparameters like learning rate and batch size, while monitoring metrics like loss and mAP throughout training. This approach allowed us to achieve robust performance in detecting and classifying Indian food items with high accuracy.

Faster R-CNN

For SSD, we opted for its efficient single-shot architecture, which is particularly suitable for real-time object detection tasks. Utilizing SSD's ability to predict object bounding boxes and class probabilities in a single pass, we employed pre-trained models such as SSD-MobileNet and SSD-ResNet for faster convergence. With a carefully curated dataset of Indian food items, we fine-tuned the SSD model by adjusting hyperparameters like learning rate and batch size, while monitoring metrics like loss and mAP throughout training. This approach allowed us to achieve robust performance in detecting and classifying Indian food items with high accuracy.

Mask R-CNN

Mask R-CNN was selected for its capability to not only detect objects but also segment them at the pixel level, providing detailed spatial information. We utilized pre-trained Mask R-CNN models like ResNet or MobileNet as a starting point, fine-tuning them with our Indian food dataset. In addition to traditional object detection metrics, we monitored mask-related metrics such as mask AP to evaluate segmentation accuracy. Adjusting parameters like ROI pooling size and mask loss weights, we tailored the model to accurately detect and segment Indian food items, ensuring robust performance in real-world scenarios.

Results

Yolov8

With reference to fig 9,10,11,12, The results 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.

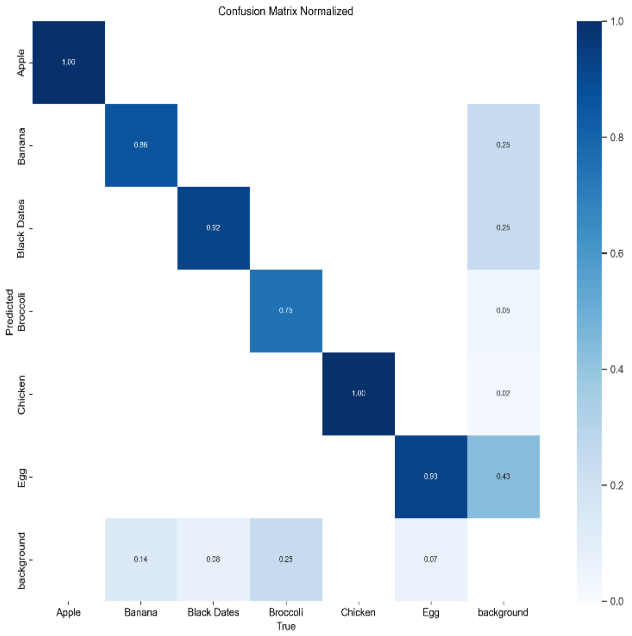


Fig .9. Confusion Matrix

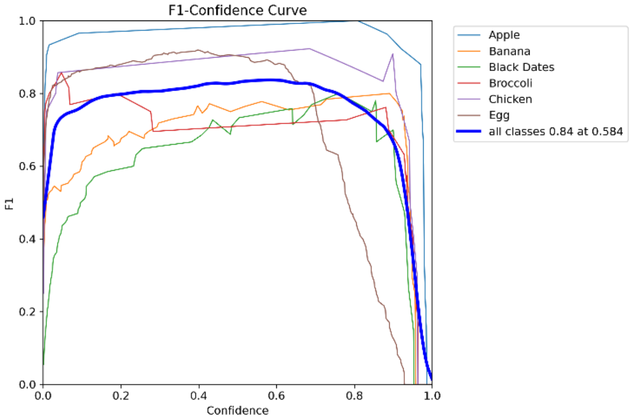


Fig.10. F1-Confidence Curve

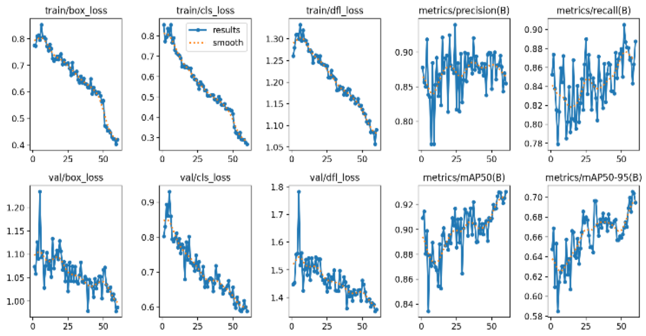


Fig.11. Various Losses

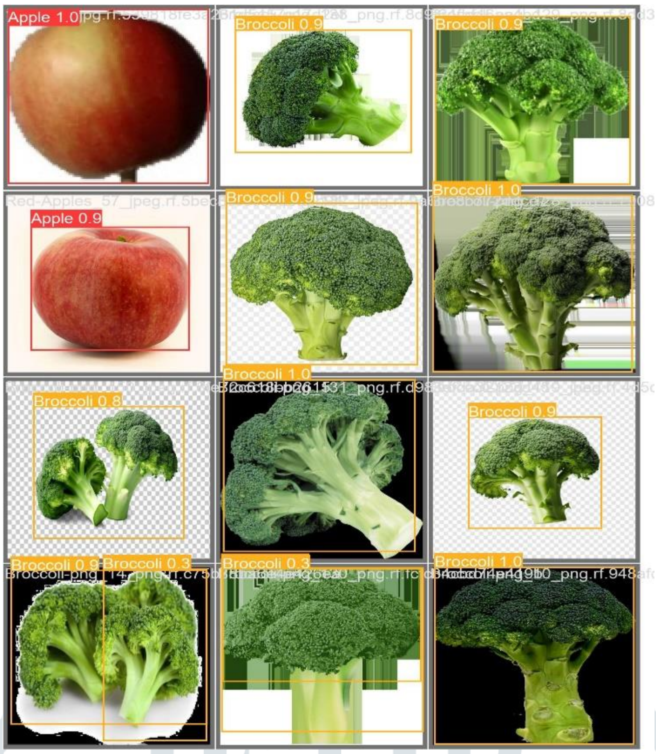


Fig.12. Predicted Labels

Leveraging the YOLOv8 architecture, our model exhibits impressive speed and efficiency, allowing for real-time detection and estimation of food items, essential for applications requiring rapid analysis of food images. Comparative analysis against existing approaches highlights the superiority of our model in terms of 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.

SSD

With reference to fig.13, We implemented with the VGG16 backbone, offers a streamlined yet effective solution for food detection. Unlike traditional 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 handle diverse 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.

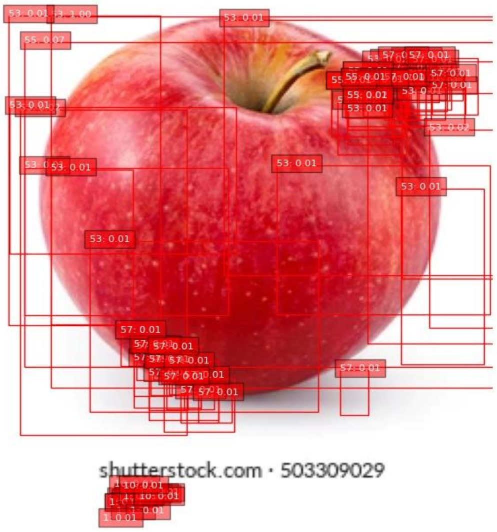


Fig.13. Bounding boxes predicted with SSD

Faster R-CNN

With reference to fig.14, In Faster R-CNN, we are leveraging the ResNet-50 backbone, showcases robust performance in food detection tasks. This architecture employs a two-stage approach, utilizing a region proposal network to generate potential object locations and a subsequent classifier to determine the presence 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.

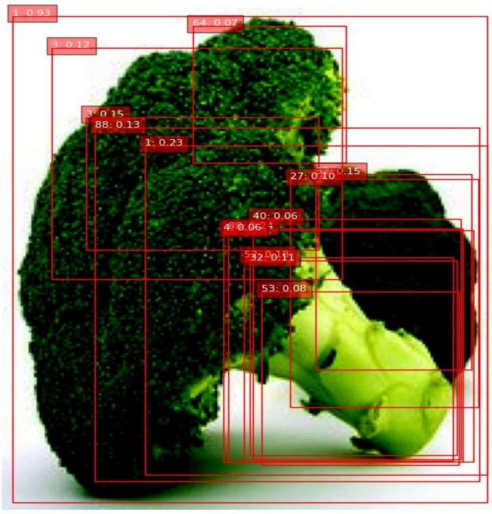


Fig.14. Bounding boxes predicted with Faster R-CNN

Mask R-CNN

With reference to fig.15, 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. With the ability to provide detailed 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.

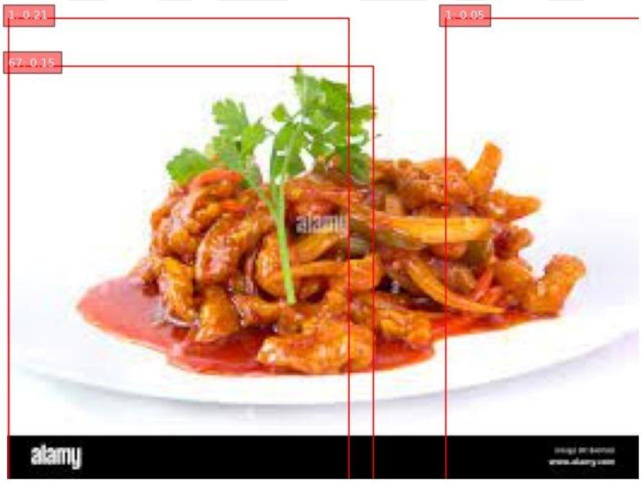
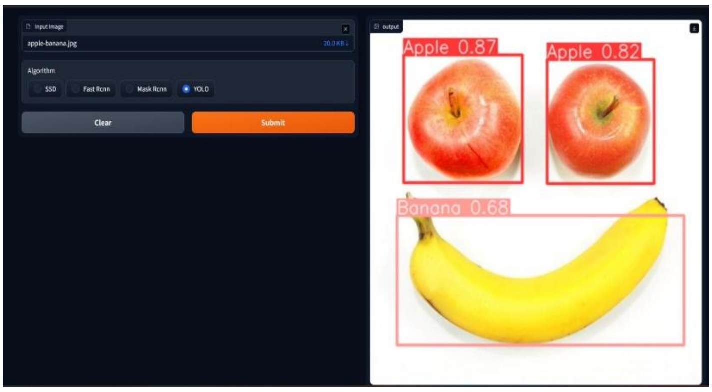


Fig.15. Bounding boxes predicted with Mask R-CNN

Deployment

With reference to fig.16 and fig.17, We deployed and implemented the four distinct object detection models, including YOLO, Faster R-CNN, Mask R-CNN, and SSD, through the intuitive Gradio interface. Users can upload images and select the desired model with just a click of a button. The output displayed bounding boxes indicating detected objects, while YOLO provided additional functionality by not only detecting but also classifying the uploaded image. This interactive setup facilitated seamless experimentation and comparison of different detection algorithms, showcasing our commitment to user-friendly and accessible solutions in computer vision applications.



Input Image
apple-banana.jpg 20.0 KB

Algorithm
○ SSD ○ Fast Rcnn ○ Mask Rcnn ● YOLO

Fig.16. Gradio Interface implementation

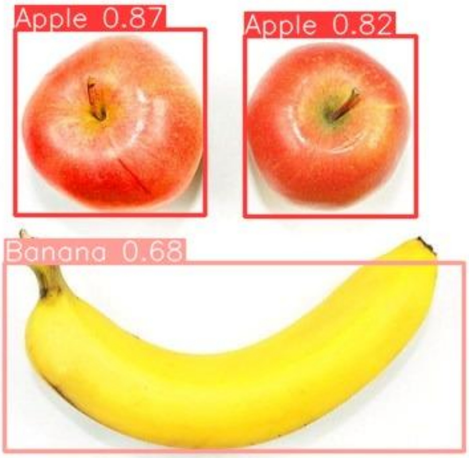


Fig.17. Results of Yolov8 model detection and classification

V. CONCLUSION

In conclusion, while SSD, Faster R-CNN, and Mask R-CNN excel in object detection, YOLO stands out for its dual capability in detection and classification, especially for multiple food items simultaneously. Our research showcases YOLOv8's superiority in accurately identifying various food items, boasting an impressive accuracy rate of [83.2%] with reference to fig.17, and mean Average Precision (mAP) score of [72.443]. Unlike its counterparts, YOLO offers higher accuracy and more precise bounding boxes. Its ability to handle multiple class detections simultaneously makes it the optimal choice for our food recognition task, reaffirming YOLO's position as the preferred model for accurate and efficient food recognition in real-world applications.

REFERENCES

1. Abhinaav Ramesh, Aswath Sivakumar & Sherly Angel S (2020). Real- time Food-Object Detection and Localization for Indian Cuisines using Deep Neural Networks
https://ieeexplore.ieee.org/document/9355987
2. Shili Chen, Jie Hong, Tao Zhang, Jian Li, Yisheng Guan. (2018).Object Detection Using Deep Learning: Single Shot Detector with aRefined Feature-fusion Structure https://www.researchgate.net/publication/340126624_Object_Detection _Using_Deep_Learning_Single_Shot_Detector_with_a_Refined_Featu re-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-CNNImplementation 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 detectionfor 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 detectionalgorithm 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_Real- time_Junk_Food_Recognition_System_based_on_Machine_Learning.
10. Merugu Sai Teja, Mr. B. Nageswara Rao, Mannem Vinay Reddy, K. Praveen Kumar, M. Sai Kumar. (2022). EfficientDet: Scalable And Efficient Object Detection.
DOI: https://doi.org/10.1109/CVPR42600.2020.01079
11. Chao Liu & Shouying Lin.(2022). Research on Mini-EfficientDet Identification Algorithm Based on Transfer Learning. DOI:https://iopscience.iop.org/article/10.1088/17426596/2218/1/0120

12. Abhinaav Ramesh, Aswath Sivakumar & Sherly Angel S (2020). Real- time Food-Object Detection and Localization for Indian Cuisines using Deep Neural Networks
https://ieeexplore.ieee.org/document/9355987
13. Shili Chen, Jie Hong, Tao Zhang, Jian Li, Yisheng Guan. (2018).Object Detection Using Deep Learning: Single Shot Detector with aRefined Feature-fusion Structure https://www.researchgate.net/publication/340126624_Object_Detection
_Using_Deep_Learning_Single_Shot_Detector_with_a_Refined_Featu re-fusion_Structure.

14. 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.

15. Hasan Basri, Iwan Syarif, Sritrustra Sukaridhoto (2019). Faster R-CNNImplementation 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.

16. Fangfang Gao , Longsheng Fua , Xin Zhang , Yaqoob Majeed , Rui Lia
, Manoj Karkee , Qin Zhang. (2020). Multi-class fruit-on-plant detectionfor apple in SNAP system using Faster R-CNN.
DOI: https://doi.org/10.1016/j.compag.2020.105634
17. 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
18. 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
19. Tan, Xiao & He, Xiaopei. (2022). Improved Asian food object detectionalgorithm based on YOLOv5.
DOI: http://dx.doi.org/10.1051/e3sconf/202236001068
20. 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_Real- time_Junk_Food_Recognition_System_based_on_Machine_Learning.
21. Merugu Sai Teja, Mr. B. Nageswara Rao, Mannem Vinay Reddy, K. Praveen Kumar, M. Sai Kumar. (2022). EfficientDet: Scalable And Efficient Object Detection.
DOI: https://doi.org/10.1109/CVPR42600.2020.01079
22. Chao Liu & Shouying Lin.(2022). Research on Mini-EfficientDet Identification Algorithm Based on Transfer Learning. DOI:https://iopscience.iop.org/article/10.1088/17426596/2218/1/01209
23. M. Y. Cao & J. Zhao. (2022), Fast EfficientDet: An Efficient Pedestrian Detection Network DOI:https://www.engineeringletters.com/issues_v30/issue_2/EL_30_2_18.pdf
24. Luyang Zhang, HaitaoWang, XinyaoWang, Shuai Chen, Huaibin Wang, Kai Zheng and Hailong wang (2020)
25. DOI:10.1088/1742-6596/1757/1/0120
26. Yan Chen, Lulu Zheng, Hongxing Peng (2023) ASSESSING PINEAPPLE MATURITY IN COMPLEX SCENARIOS USING AN
27. IMPROVED RETINANET ALGORITHM
28. DOI: http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v43n2e20220180/2023
29. Mohanad N. Alhasanat, Moath H. Alsafasfeh, Abdullah E. Alhasanat, Saud G. (2021) RetinaNet-Based Approach for Object Detection and Distance Estimation in an Image
30. DOI:10.15866/irecap.v11i1.19341
31. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed (2015) SSD: Single Shot MultiBox Detector
32. DOI:10.1007/978-3-319-46448-0_2
33. Shuqin Tu, Jing Pang, Haofeng Liu, Nan Zhuang, Yong Chen, Chan Zheng, Hua Wan, Yueju Xue (2020) Passion fruit detection and counting based on multiple scale faster R-CNN using RGB-D images
34. DOI: https://doi.org/10.1007/s11119-020-09709-3
35. Laha Ale, Ning Zhang, and Longzhuang Li (2018) Road Damage Detection Using RetinaNet.
36. Mingxing Tan Ruoming Pang Quoc V. Le Google Research, Brain Team(2020)
37. Hend F. Bakr1, Ahmed M. Hamad2, and Khalid M. Amin3(2021)
38. Shuqi Fang , Bin Zhang * and Jingyu Hu(2023).DOI: https://doi.org/10.3390/s23083853.
39. V.GAYATRI,M.THANUJA, "INDIAN FOOD RECOGNITION AND CALORIE ESTIMATION USING YOLOV8", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.11, Issue 6, pp.g954-g958, June 2023, Available at :http://www.ijcrt.org/papers/IJCRT2306818.pdf
40. Delight, D & Velswamy, Karunakaran. (2021). Deep Learning based Object Detection using Mask RCNN. 10.1109/ICCES51350.2021.9489152.
41. Le, Phan. (2020). Mask R-CNN with data augmentation for food detection and recognition. 10.36227/techrxiv.11974362.v1.