A Computer Vision-based Model to Implement Automation in Retail Checkouts Using YOLOv5

Anmol Jain[a], Kshiteesh Kumar[b], Nandita Yadav[c], Kumari Bhavya Chaubey[d]

[a]KIET Group of Institutions, Delhi-NCR Meerut Road, Ghaziabad, 201206, India
[b]KIET Group of Institutions, Delhi-NCR Meerut Road, Ghaziabad, 201206, India
[c]KIET Group of Institutions, Delhi-NCR Meerut Road, Ghaziabad, 201206, India
[d]KIET Group of Institutions, Delhi-NCR Meerut Road, Ghaziabad, 201206, India

Abstract: With the advent of technology, the retail sector has witnessed significant changes that once used to be believed too unfathomable. Self-service checkout systems have been established in a variety of stores in which customers have to scan all the bought items with a barcode reader and then pay for the bought items. However, these services could result in thefts in these stores and reduce customer experience and impersonal communication between employees and customers. Technologies such as Amazon Go, released by Amazon in 2017, which is based on "Just Walk Out Technology", eased the self-checkout process as the customers do not have to wait in queues to pay for the purchased items. After installing the Amazon Go app, customers scan their iPhones as they enter the store. The track of their purchases is maintained. They can leave the store without physically checking out as their accounts are charged automatically. It makes use of deep learning, computer vision, and sensors. It is built on technology like self-driving cars. However, this technology can hinder customer experience as customers unfamiliar with the app would find it very difficult to shop. Besides the technology used in apps like Amazon Go is expensive. Therefore, to solve these problems and produce an outcome that could favor both the customer and the retailer is described in this paper. Paper focuses on enhancing customer experience by creating a system that can scan the bought items, detect these items with the help of a pre-trained model, and generate a bill containing the total cost that needs to be paid.

Keywords: Computer Vision, Deep Learning, Object Detection, tkinter, YOLOv5

1. Introduction

Nowadays automation techniques have proved to be very beneficial in the retail sector. They play a crucial role in the retail sector by providing better efficiency, reducing costs, enhancing customer experience as well as giving a competitive edge in today's fast-paced market. Many applications make use of Computer Vision techniques to simulate the behavior of human eyesight. The study of Computer Vision allows systems to extract useful data from digital photos, videos, and other forms of media.

One of the steps involved in Computer Vision is Object Detection. Its main objective is to find and localize the objects present in an image or a video. It entails two tasks: Image Classification and Object Localization [1]. While object localization involves locating objects in an image and using a bounding box to denote their locations, image classification identifies the kind or class of an object.

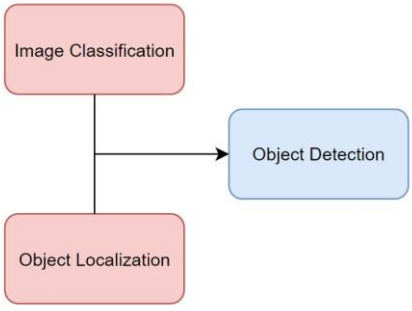

Fig. 1 - Object Detection Tasks

In the case of detecting multiple objects, multiple object detection algorithms are used.

Convolutional Neural Networks (CNN) are popularly used in object detection. Over the past few years, CNN has become famous for all the tasks involved in computer vision as well as object detection [2].

1.1. Types of object detection models:

Object detection methods that are based on CNN, are grouped into two models: One-Stage Models and Two-Stage Models [3]. They are explained as:

A. One-Stage Models:

These models require one pass only through the neural network and predict the bounding boxes in one go. They prioritize inference speed. YOLO series, SSD, and RetinaNet are the representative networks [4].

You Only Look Once (YOLO) is a widely used object detection algorithm, capable of detecting objects in real-time. It has many versions such as YOLOv3, YOLOv5, YOLOv6, etc. SSD (Single Shot Detector) pertains to multi-scale target feature extraction helping in predicting multiple classes. RetinaNet introduces a new function known as the Focal Loss Function that deals with class imbalance during training.

B. Two-Stage Models:

These models consist of two steps: creating multiple proposal bounding boxes in the first phase, and using these bounding boxes to perform item categorization and location information prediction in the second phase [5]. These models prioritize detection accuracy over inference speed. R-CNN,

Fast R-CNN, SPPNet, Faster R-CNN, FPN, and Mask-RCNN are the representative models of Two-Stage Models [6].

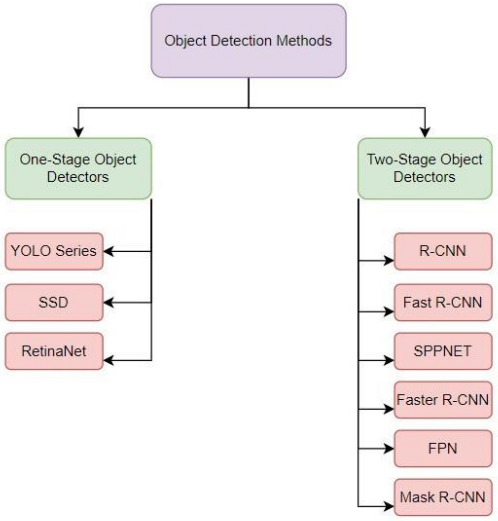

Fig. 2 - Classification of Object Detection Tasks

R-CNN uses selective search approach to propose regions that likely contain objects [7]. Instead of using a selective search algorithm, Fast R-CNN establishes a pooling layer known as Region of Interest (RoI) that extracts fixed-size feature maps from CNN feature maps whereas Faster R-CNN uses an integrated Region Proposal Network (RPN) that shares convolutional features with the detection network [8]. Mask R-CNN extends Faster R-CNN by including a branch for predicting segmentation masks, along with bounding boxes. Spatial Pyramid Pooling Network (SPPNet) is used for arbitrary images without requiring resizing.

Markov Clustering Networks (MCN) and the Viola-Jones technique are used to detect multiple objects and the presence of faces and text within an image. However, these approaches do not account for things like picture size, gradient, orientation, and characters [9].

Our system is based on the YOLOv5 technique of object detection. It is a model in the YOLO series of Computer Vision models and is used to detect multiple objects in an image with fast speed and better accuracy [10]. In this paper, we have discussed the implementation of an automated retail checkout system which generates the bill of the bought products by taking images or videos of the bought products. The main objective of this system is to ease the process of checkout at the retail stores. It can help to reduce the long queues at the checkout counters and can therefore, save customers' time. It aims to enhance the customer-employee relationship and helps to reduce the cashiers' burden to manually scan every item one by one at these stores. It can also be used in maintain the stocks at the retail stores.

The system checks whether the items present in the image/video are present in the dataset or not. For pre-processing the dataset, Roboflow is used. Roboflow has an inbuilt feature that can be used to give annotations. Using it, we put labels on each of the photos. Along with this, Roboflow increases the dataset by rotating every image with different angles. After processing the dataset from Roboflow, we import data in YOLOv5 for further detection. Object Detection and Object Recognition are performed through YOLOv5 which is trained on the COCO Dataset. YOLOv5 is used to achieve better accuracy in object detection. It is trained on custom datasets. The system can then generate an automated bill that includes the information like name of the product, its quantity, and its price. The final amount of all the bought items is also written at the end of the report.

2. Related Work

Many researchers have investigated the applications of Computer Vision in managing retail stores. They have successfully carried out studies and explained the methodologies that could help the retailers manage these stores. A brief literature review is explained in this section.

In 2019, D. A. Mora Hernandez, O. Nalbach, and D. Werth introduced a conceptual tracking system that could generate movement tracks over time for individual customers [11]. In 2020, N. Shekokar, A. Kasat, S. Jain, P. Naringrekar, and M. Shah, prepared an innovative model 'Shop and Go' which makes use of deep learning and sensor fusion [12]. Sensors were used to detect whether the weight had been changed if the item was picked or not. They also stated some real-time applications along with their pros and cons. Figure 3 points to the real-time applications of Self-Checkout Systems.

Fig. 3 - Applications of Self-Checkout Systems [12]

Besides there was also research conducted as to how computer vision can be used to manage product stock at offline retail stores. In 2020, M. A. Majdi, B. Sena Bayu Dewantara, and M. M. Bachtiar proposed a system to find out which goods were nearly empty and misplaced [13]. A camera was used to capture all the displayed products and the products were recognized using YOLOv3, which resulted in achieving the accuracy of 97.61% and 76.67% for misplaced detection. In 2020, C. G. Melek, Elena Battini Sonmez, and Songul Albayrak compared various object detection algorithms and concluded that YOLOv2 is a far better object detection algorithm in terms of both performance and speed for object detection in shelf images [14].

S. K. Yedla, V. M. Manikandan, and V. Panchami in 2020, proposed a novel approach for real-time scene change detection with object detection for automated stock verification [15]. They used a scene change detection technique based on a Structural Similarity Index (SSIM) with the goal of optimizing the processing of video frames.

In the work proposed by M. Sugadev, K. Sucharitha, I. R. Sheeba, and B. Velan, the design and implementation of a hardware-based automated billing system aimed for fruit shops and achieving the Granny Smith Accuracy of 98.896%, was described [16]. It proposed more efficient and accurate billing system than traditional billing systems. They used neural network to classify the fruit and load cell to find the weight of the fruit. In 2020, H. Y. Putra proposed a fraud detection system to detect fraud at self-

checkout stores using Data Mining. He predicted fraud using classification techniques and visualized the results to obtain new insight. J48 model had the best performance with F-measure 0.921 [17].

Along with these, there have been research on various object detection algorithms and how do they differ from each other in terms of various factors. The application requirements as well as several other factors, such as speed, accuracy, handling small objects, training complexity, and real-time object detection capacity, influence the choice of object detection method. YOLO excels in real-time object detection as compared to other algorithms. SSD provides a good compromise between speed and accuracy and can handle small objects efficiently.

In Table 1, one-stage object detection methods - SSD and YOLO, and two-stage object detection methods - R-CNN, Fast R-CNN, and Faster R-CNN, are compared based on various parameters like speed, real-time object detection, accuracy in detecting small objects, and training complexity [18].

Table 1 - Comparison of some object detection methods [18]

Point of Difference	R-CNN	Fast R-CNN	Faster R-CNN	SSD	YOLO
Speed	Slow	Medium	Medium	Fast	Fast
Real Time Object Detection	Low	Better than R-CNN	Better than R-CNN and Fast R-CNN	Best	Best
Accuracy in Detecting Small Objects	Low	Better than R-CNN	Better than YOLO	High	Low
Training Complexity	High	Easier than R-CNN	Easier than R-CNN and Fast R-CNN	Easier than two-stage detection methods	Low

3. Proposed Methodology

The proposed system takes video or image as input and generates a bill containing the price and quantity of the bought items. The whole methodology is explained in this section.

The dataset contains 50-50 images of each of the products in various categories. These products include Apples, Oranges, Bananas, Coconut, and Eggs.

We have used Robo Flow, which is a development framework for Computer Vision, used for enhancing data collection to pre-process the data and train the model. It is used to meet the following two objectives:

1) Pre-processing

There is an inbuilt feature in Roboflow for giving annotations. We had put labels on each of the photos. Apart from this, Roboflow also rotated the image from different angles. It was done for every image and hence 2000 approx. images were achieved in comparison to the earlier 720 images.

2) Data Production

In this process, Roboflow divided our dataset into three folders, that is, Training, Testing, and Validation. In each of these folders, there were 2 subfolders - images and labels. Data.yml files extracted all the labels that we gave and then allotted respective labels on respective items. After processing the dataset from Roboflow we received 1 API key and this API key was used to import data in Yolo for further detection.

YOLO is a well-recognized object detection method used widely because of its accuracy and quickness. It is a one-stage approach to object detection. We have used YOLOv5 which is a model trained on the COCO dataset which uses a transfer learning technique [18]. We have tuned this model on our dataset to get better accuracy for object detection. YOLOv5 has five models: yolov5n, yolov5s, yolov5m, yolov5l and yolov5x [19]. In our model, we have used yolov5s which has 7.2 M parameters.

YOLOv5 uses Efficient Det, which is a more complex architecture unlike the previous versions of YOLO. This helps to getter better accuracy and improved generalization of object categories. Apart from this, YOLOv5 is trained on 600 object categories which is a more diverse dataset than other versions. It also utilizes Spatial Pyramid Pooling (SPP) which helps to improve the performance of the model in detecting the smaller objects.

We have used tkinter, which is a standard GUI library for Python, to provide the user interface. The users can upload the images or videos using this interface and the results get displayed.

In Figure 4, we have explained the whole process of checkout at the retail stores with the help of the workflow diagram.

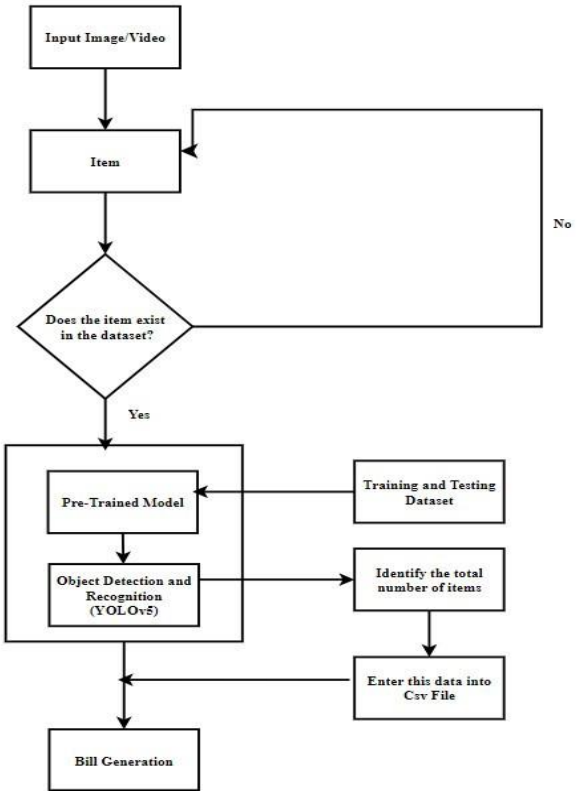

Fig. 4 - Work Flow Diagram of Bill Generation

4. Results and Discussions

When an input image/video containing the items is passed to the system, the system detects whether the data exists in the dataset or not. If yes, then the pre-trained model detects and recognizes the object. Then the detect method, which is made by Tkinter, has a list that contains the per unit cost of each item. The bill is then created in the data frame and the price of each item is calculated. These price values are then added to the data frame and output is displayed.

Figure 5 shows that the model works well in prediction on test samples. The egg is predicted with 70% confidence.

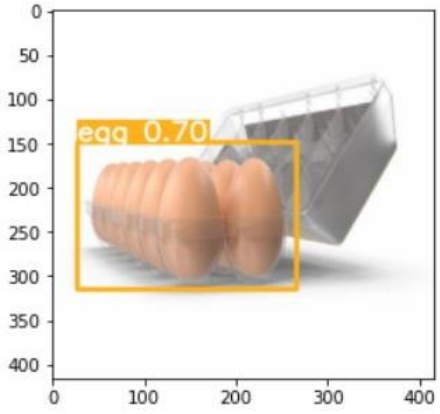

Fig. 5 - Confidence Value for Egg Prediction

When the user clicks on open a file, the interface allows the user to select the image (either in .jpeg or .png format) or video and upload it.

In Figure 6, the input image contains 2 bananas, 2 apples, and 2 oranges. The price of each quantity as mentioned in the list is [‘banana’: 10, ‘apple’:50, ‘orange’:40]. Hence the bill generated is calculated by multiplying the quantity with price and then adding it to the final value.

Fig. 6 - Output (Bill Generation)

5. Conclusion

This system provides an innovative solution that can help the masses to smoothen their checkout experience. It can help the retailers and the customers to effectively checkout using a simple approach. It can generate the bill of the items that contain the information regarding the bought products like name of the time, quantity, weight, price as well as the total bill. The system uses YOLOv5, a very precise and efficient object detection model, to efficiently detect the objects in the video or the image. Since this model uses fewer resources and is cost-effective in comparison to the technologies existing in the market, it can provide an easier approach to checkout services. Retailers as well as customers can benefit from it. It provides a solution to the long queues the customers must stand in to pay the bill.

This system shows great promise but there are certain limitations and areas for further improvement. It is crucial for the retailers to maintain up-to-date datasets for object detection. Apart from this, there might be some errors in accurately identifying and classifying the objects in situations like poor lighting in photos or low-resolution pictures. The future work can be done regarding the improved and novel models for object detection tasks and enhancing the training strategies like transfer learning or domain learning.

Acknowledgement

The authors sincerely thank the editorial board and reviewers for their valuable time and insightful comments.

References

D. Triphena Delight and V. Karunakaran. 2021. “A Comprehensive Analysis of Methodologies used for Object Detection and Localization,” in 2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021, Institute of Electrical and Electronics Engineers Inc. pp. 448–453

Y. Hu, X. Li, N. Zhou, L. Yang, L. Peng, and S. Xiao. (2019). “A Sample Update-Based Convolutional Neural Network Framework for Object Detection in Large-Area Remote Sensing Images,” IEEE Geosci. Remote Sens. Lett., vol. 16, no. 6, pp. 947–951

“Object Detection: Models, Architectures & Tutorial [2023].” Accessed: Oct. 31, 2023. [Online]. Available: https://www.v7labs.com/blog/object-detection-guide

“One-stage object detection.” Accessed: Oct. 31, 2023. [Online]. Available: https://machinethink.net/blog/object-detection/

Anhui da xue et al.,2020, Proceedings of the 32nd Chinese Control and Decision Conference (CCDC 2020), Hefei, China.

“A guide to Two-stage Object Detection: R-CNN, FPN, Mask R-CNN | by Sieun Park | CodeX | Medium.” Accessed: Oct. 31, 2023. [Online]. Available: https://medium.com/codex/a-guide-to-two-stage-object-detection-r-cnn-fpn-mask-r-cnn-and-more-54c2e168438c

N. Yadav and U. Binay. (2017). “Comparative Study of Object Detection Algorithms,” Int. Res. J. Eng. Technol. [Online]. Available: www.irjet.net

A. John and D. Meva. (2020). “A Comparative Study of Various Object Detection Algorithms and Performance Analysis,” Int. J. Comput. Sci. Eng. Open Access Res. Pap., vol. 8, no. 10

Shandong da xue. (2019). Asia-Pacific Signal and Information Processing Association, and Institute of Electrical and Electronics Engineers, 2019 2nd IEEE International Conference on Information Communication and Signal Processing (ICICSP 2019). Weihai, China.

J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. (2015). “You Only Look Once: Unified, Real-Time Object Detection.”, [Online]. Available: http://arxiv.org/abs/1506.02640

D. A. Mora Hernandez, O. Nalbach, and D. Werth. (2019). “How computer vision provides physical retail with a better view on customers,” in

Proceedings - 21st IEEE Conference on Business Informatics, CBI 2019, Institute of Electrical and Electronics Engineers Inc., pp. 462–471.

N. Shekokar, A. Kasat, S. Jain, P. Naringrekar, and M. Shah. (2020). “Shop and go: An innovative approach towards shopping using deep learning and computer vision,” in Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020, Institute of Electrical and Electronics Engineers Inc., pp. 1201–1206.

M. A. Majdi, B. Sena Bayu Dewantara, and M. M. Bachtiar. (2020). “Product stock management using computer vision,” in IES 2020 - International Electronics Symposium: The Role of Autonomous and Intelligent Systems for Human Life and Comfort, Institute of Electrical and Electronics Engineers Inc.

B. Dumnić et al.(2019). IEEE EUROCON 2019 : 18th International Conference on Smart Technologies, Novi Sad, Serbia.

S. K. Yedla, V. M. Manikandan, and V. Panchami. (2020). “Real-Time Scene Change Detection with Object Detection for Automated Stock Verification,” in ICDCS 2020 - 2020 5th International Conference on Devices, Circuits and Systems, Institute of Electrical and Electronics Engineers Inc.

M. Sugadev, K. Sucharitha, I. R. Sheeba, and B. Velan. (2020). “Computer vision based automated billing system for fruit stores,” in Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020, Institute of Electrical and Electronics Engineers Inc.

H. Y. Putra. (2020). “Fraud detection at self checkout retail using data mining,” in 2020 International Conference on Information Technology Systems and Innovation, ICITSI 2020 - Proceedings, Institute of Electrical and Electronics Engineers Inc.

“What is YOLOv5? A Guide for Beginners.” Accessed: Oct. 31, 2023. [Online]. Available: https://blog.roboflow.com/yolov5-improvements-and-evaluation/

“YOLOv5 - Fine Tuning & Custom Object Detection Training.” Accessed: Oct. 31, 2023. [Online]. Available: https://learnopencv.com/custom-object-detection-training-using-yolov5/