IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 12, December 2024

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.524

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 12, December 2024

|DOI: 10.15680/IJIRSET.2024.1312118|

Real-Time Image Recognition for Retail Checkout

Ishaan Mukherjee

Student, Delhi Public School East, Bengaluru, India

ABSTRACT: A typical shopping cart in a retail store consists of multiple products. During checkout, each product is individually picked up, its barcode is scanned, and the item gets added for billing. The billing process is thus tedious and time-consuming. This paper proposes the use of Artificial Intelligence (AI) and Image Recognition to attack the checkout problem, speed up the billing process and thereby improve customer satisfaction. The idea here is to use a deep neural network model for object detection in a typical shopping cart, bag, or basket. A Deep Learning model is trained with images of products from a retail store. During checkout, given the image of a whole basket of products, the model can detect or infer the items in the basket. This promises to enable quick, hassle-free billing and checkout at supermarkets.

KEYWORDS: Deep Learning, AI, Object Detection, Retail Analytics, Automation.

I. INTRODUCTION

Economic growth, social transformation, and the associated changes in the consumption pattern of consumers have necessitated the effective use of emerging technologies for improving and revolutionizing the shopping experience. Customers are often dissatisfied with the waiting time in billing counters of shops and malls. People are forced to wait patiently in long queues for billing, painfully watching the shop employee or checkout operator scanning the barcode of each product and finally clearing customers from the queue one by one. The entire mechanism is tedious, time-consuming and prone to human error.

To overcome this widespread inconvenience, smarter automated billing systems in stores have been researched quite extensively in recent times [1]. Such automated billing systems can reduce and eliminate the waiting time of buyers in retail stores and also achieve good efficiency in the overall output by eliminating the need of excessive human attention. These systems indeed improve the overall customer satisfaction. However, most of these systems are based on the use of custom electronic devices such as smart cards, barcode readers and RFID scanners [2,6]. The tags pertaining to their respective products are scanned against a reader as the customer drops the item into the cart or the till checkout. The prices are available on a centralized server, often in the cloud, and this is used to generate the final bill.

In this paper, we suggest an alternative billing approach that utilizes state of the art image recognition technology [3] against photos or images of the market basket or shopping cart. Object detection [4] algorithms have been effectively used to detect individual products in the cart and add them for billing. The novelty of the idea lies in the following areas:

- Simplicity of the infrastructure: Cameras can be used as opposed to an elaborate and expensive system of custom sensors. It is to be noted that cameras are ubiquitous in the stores already for existing security concerns like thefts etc.
- Eliminates tags and stickers: There is no need to print barcode stickers or write RFID tags and affix them manually. Tags and stickers involve heavy manual work to generate and affix and errors are not uncommon. Additionally, tags and stickers are often damaged or lost due to the vagaries of transportation, weather conditions and mishandling.
- Eliminates machine readable codes: The additional layer of machine-readable codes becomes unnecessary when the photos of the products themselves serve as the detectable items to be billed.
- Bulk item scanning: Traditional checkout methods require individual scanning of each item in a market basket or shopping cart, a process that is both time-consuming and frustrating for customers and retailers alike. The proposed approach enables the simultaneous scanning of entire shopping baskets, automatically identifying and capturing each billable product item.
- Amenable to self-service checkouts using mobile apps: It can be conceived that the proposed mechanism could be utilized to build a mobile app that customers can install on their own mobile phones and use it to scan the shopping

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 12, December 2024

|DOI: 10.15680/IJIRSET.2024.1312118|

cart using the inbuilt phone camera. As phone cameras have become powerful and commonplace, the cost savings can be significant.

II. DATA SET

A brief about the data used in this project is as follows:

A. Training data

A total of 1,595 images belonging to 6 retail products are used for training the deep learning model. For each product many images are captured from multiple angles and orientation for a comprehensive 360-degree view [5]. In our project, each product is assigned a simple one-digit unique identifier and associated with a simple label for convenience of reporting the results in a compact way. As an example, one of simplified labels is 'candy'. In a full-scale real-world solution, the identifier will be the product SKU or UPC code and the label will be the full product title, e.g. "Cadbury 5 Star Chocolate Bar 38 gm". "Fig. 1" shows the products selected. The training data set contains 250-300 images for each product.

B. Testing data

The test dataset consists of 200 sample market basket images. Each image has multiple product items and represents a basket of retail products which needs to be billed. Sample test images can be seen in "Fig. 2".

More details are provided in the Section V where the results are discussed.

Fig. 1: Sample images for the 6 product items, used for training the ML model

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 12, December 2024

|DOI: 10.15680/IJIRSET.2024.1312118|

Fig 2: Sample images of market baskets containing multiple product items, used for testing the ML model

III. METHODOLOGY

The data was processed as described in the following steps:

A. Data Preparation

The image annotations provided in the dataset are in json-based COCO format [7]. A YOLOv8s [8] model was selected for the project. Hence, the annotations were converted to YOLO format using the conversion tools provided by the YOLO libraries.

B. Model Training

A YOLOv8s model was selected for the project after due research. Although later versions of YOLO are available, v8 is widely used and well-proven in the real world and suffices for the purposes of this project.

The coding language used was Python. The model fine-tuning using the training dataset was done in Google Colab using GPU as the runtime. Based on the infrastructure constraints and the consequent restricted scope of this project, 10 epochs were used to train the ML algorithm. With the 1,595 training set images each epoch took approximately 10-15 minutes. Thus with 10 epochs we ran the training process for around 2-3 hours.

C. Model Testing

Multiple images of market baskets were analysed with the finetuned and fully trained ML model. Test data with different number of product items and at different clutter levels were used for testing the trained ML model. Sample images were also manually generated using image editors to augment the test dataset for further in-depth understanding of the ML model performance.

Non-max suppression (NMS) was used to make a final selection of the best object bounding boxes and eliminate redundant bounding boxes from those predicted by the model.

IV. BILLING

Depending on the items detected by our finetuned model, the total bill was calculated based on the product identifiers, unit prices and quantities detected. The sample product price list for the 6 products which have been considered in our project is provided in Table 1. This is used for calculating the total bill for each of the market baskets.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 12, December 2024

|DOI: 10.15680/IJIRSET.2024.1312118|

TABLE I PER UNIT COST OF PRODUCT ITEMS IN RUPEES

Product label	Cost/Item(Rs)
alcohol	160
candy	10
dried_fruit	85
personal_hygiene	75
tissue	50
instant_drink	60

V. RESULTS AND FINDINGS

Images have been tested against the trained ML model. For each image tested the following output have been reported in our Python code:

- Labels: The labels of product items detected in the basket.
- Boxes: List of bounding boxes detected for the various product items
- Scores: List of confidence scores from the model for the bounding boxes above
- Post_NMS_indexes: Indexes of final bounding boxes from the list of detected bounding boxes above, i.e. final product items selected after NMS
- Labels_for_billing: Labels for the product items finally detected. These will be used for checkout and billing.

Given below are the outputs for a representative sample of 5 images of market baskets from the test dataset of 200+ images that was analysed. We start with an image of a single product item in Case 1 below. The number of product items and the clutter level of the market basket progressively increases as we move to Case 5. Salient observations and findings along with a detailed output from the Python model inference code are reported next.

A. Case 1: Single item as shown in "Fig. 3a"

Fig. 3a: Single personal hygiene product

Output from model inference code	
Labels:	['personal_hygiene']
Boxes:	tensor ([[986.4863, 1101.3569, 1908.6455, 1548.1135]])
Scores:	tensor ([0.9978])
Post_NMS_indexes:	[0]
Labels_for__billing:	['personal_hygiene']

Bill Amount (Rs.)			
Product	Rate	Qty	Price
personal_hygiene	75	1	75
		Total:	75

Observation: The object in the image is exactly similar to the detected image and things work smoothly. Please note that this image was taken from the training dataset. Many other training dataset images were tried and all were detected correctly indicating good fit to the training dataset.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 12, December 2024

|DOI: 10.15680/IJIRSET.2024.1312118|

B. Case 2: Three items with low clutter as shown in “Fig. 3b”

Fig. 3b: 2 alcohol and 1 dried-fruit

Bill Amount (Rs.)			
Product	Rate	Qty	Price
dried_fruit	85	1	85
alcohol	160	2	320
		Total:	405

Output from model inference code	
Labels:	['dried_fruit', 'alcohol', 'alcohol', 'dried_fruit', 'instant_drink', 'alcohol', 'instant_drink']
Boxes:	tensor ([[156.7694, 58.9699, 413.1499, 417.5941], [506.9289, 41.6780, 635.9908, 450.0270], [764.7944, 141.7199, 895.5837, 549.3628], [170.9240, 317.8016, 402.1803, 409.8457], [504.2130, 96.3378, 645.0374, 463.3001], [790.2310, 142.0179, 864.2974, 291.5605], [763.5851, 183.7858, 905.0693, 564.5370]])
Scores:	tensor ([0.9961, 0.9924, 0.9867, 0.2038, 0.2016, 0.1358, 0.1236])
Post_NMS_indexes:	[0, 1, 2]
Labels_for_billing:	['dried_fruit', 'alcohol', 'alcohol']

Observation: The objects in the image have been detected correctly and accurately without any errors or omissions. Here, the objects are well-separated unlike in the subsequent ones described below which are more reflective of the real-world scenarios.

C. Case 3: Three items with some clutter as shown in “Fig. 3c”

Fig. 3c: 1 instant_drink, 1 tissue, 1 personal_hygiene

Bill Amount (Rs.)			
Product	Rate	Qty	Price
personal_hygiene	75	1	75
instant_drink	60	1	60
		Total:	135

Output from model inference code	
Labels:	['personal_hygiene', 'instant_drink', 'personal_hygiene', 'instant_drink', 'alcohol']
Boxes:	tensor ([[1062.8499, 808.6298, 1305.1072, 1388.0894], [862.7060, 853.8253, 1109.2147, 1391.3528], [864.0807, 771.4449, 1289.1107, 1408.7584], [868.7659, 781.9387, 1276.7812, 1411.6917], [859.5369, 818.2517, 1112.1440, 1346.2878]])
Scores:	tensor ([0.9864, 0.8227, 0.6777, 0.2924, 0.1251])
Post_NMS_indexes:	[0, 1]
Labels_for_billing:	['personal_hygiene', 'instant_drink']

Observation: The objects ‘personal_hygiene' and 'instant_drink' are accurately detected and with high scores. However, the model could not detect the image of ‘tissue'. In this and also other cases we have observed that smaller size objects are not detected well in presence of other bigger sized products.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 12, December 2024

|DOI: 10.15680/IJIRSET.2024.1312118|

D. Case 4: Six items with repetition and medium clutter as shown in “Fig. 3d”

Output from model inference code	
Labels:	['alcohol', 'alcohol', 'dried_fruit', 'alcohol', 'personal_hygiene', 'personal_hygiene', 'candy', 'alcohol', 'dried_fruit', 'tissue', 'candy']
Boxes:	tensor ([[618.3820,669.2455, 786.3887, 1291.6223], [886.0617, 459.0985, 1067.3031, 1063.2598], [1141.8190, 1001.0757, 1692.9375, 1611.7036], [873.4615, 472.1278, 1105.4487, 1514.4152], [876.9941, 829.8938, 1263.4655, 1521.5916], [1159.3469, 941.1329, 1324.7551, 1245.3639], [1164.5406, 950.2405, 1321.8884, 1233.1842], [865.6069, 817.4528, 1206.3895, 1534.4136], [866.8254, 884.5579, 1324.9884, 1518.1936], [1168.0889, 952.9962, 1318.7053, 1235.5690], [890.6416, 796.6884, 1183.1213, 1535.6150]])
Scores:	tensor ([0.9956, 0.9771, 0.9703, 0.9265, 0.6405, 0.4215, 0.3806, 0.2128, 0.1154, 0.0807, 0.0678])
Post_NMS_indexes:	[0, 1, 2, 4, 5]
Labels_for_billing:	['alcohol', 'alcohol', 'dried_fruit', 'personal_hygiene', 'personal_hygiene']

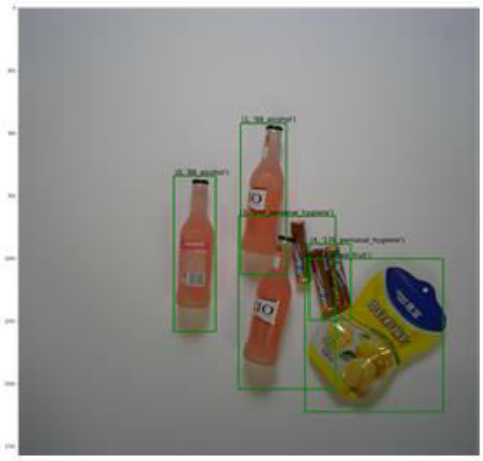

Fig. 3d: 3 alcohol, 1 dried_fruit, 3 candy

Bill Amount (Rs.)			
Product	Rate	Qty	Price
alcohol	160	2	320
dried_fruit	85	1	85
personal_hygiene	75	2	150
		Total:	555

Observation: The objects detected with high confidence are 'alcohol' and 'dried_fruit'. However, ‘personal_hygiene' is wrongly detected. The model could detect the image of 'candy’ but with low prediction score. Because of non-max suppression algorithm, the bounding boxes for candy got suppressed. Like in Case 3, the smaller size objects(candy) are not detected well in presence of other bigger sized products. It is to be noted that the ‘dried_fruit' is placed in an inclined position while the bounding boxes determined by the object detector are always axis-aligned. This can cause the larger object’s bounding box to suppress those of nearby smaller objects through undesirable overlap.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 12, December 2024

|DOI: 10.15680/IJIRSET.2024.1312118|

E. Use Case 5: Six items with repetition and higher clutter as shown in "Fig. 3e"

	Output from model inference code
Labels:	['dried_fruit', 'alcohol', 'candy', 'alcohol', 'candy', 'alcohol', 'dried_fruit', 'candy', 'personal_hygiene', 'dried_fruit', 'personal_hygiene', 'candy', 'personal_hygiene', 'personal_hygiene']
Boxes:	tensor ([[672.5255, 501.7234, 1362.2582, 1232.0538], [1138.0996, 424.1945, 1366.7843, 1217.8722], [809.1041, 1128.0518, 887.2255, 1413.2192], [934.8916, 1025.7693, 1107.1512, 1699.4375], [1179.5281, 978.8802, 1392.1659, 1259.6604], [671.5580, 598.8249, 892.4268, 1226.9850], [806.6865, 503.7394, 1189.5353, 1092.6766], [1072.8545, 1064.2828, 1226.1503, 1328.4275], [936.7919, 1071.5308, 1111.1021, 1688.5388], [1069.8129, 452.3283, 1358.3667, 1231.7231], [1119.3275, 443.6690, 1382.5175, 1229.6810], [1124.9666, 431.6636, 1378.2477, 1256.8555], [671.2133, 606.2527, 900.8597, 1199.6556], [809.8979, 1122.5916, 884.8666, 1410.6927]])
Scores:	tensor ([0.9868, 0.9330, 0.8359, 0.7751, 0.6281, 0.4143, 0.3683, 0.2079, 0.1168,0.0930, 0.0918, 0.0768, 0.0637, 0.0623])
Post_NMS_indexe:	[0, 2, 3, 4, 7]
Labels for billing:	['dried_fruit', 'candy', 'alcohol', 'candy', 'candy']

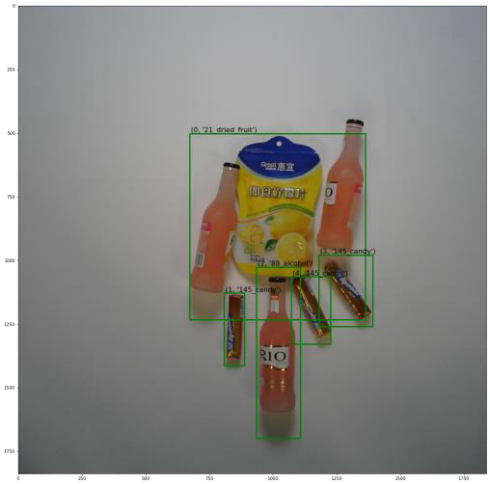

Fig. 3e: 3 alcohol, 3 candies, 1 dried fruit

Bill Amount (Rs.)			
Product	Rate	Qty	Price
dried_fruit	85	1	85
candy	10	3	30
alcohol	160	1	160
		Total:	275

Observation: It is seen that dried fruit and alcohol are clearly detected with high confidence in the image. Two of the alcohol bottles are originally detected but are removed when the best bounding boxes are getting selected. In this image, unlike the previous examples, all the candies are detected correctly. This is probably because the candies in this image are dispersed and away from each other and hence the bounding boxes are easily separable and detectable. It has been observed that increasing the value of iou_threshold (iou: Intersection over Union is a measure of overlap) in NMS (Non-Max Suppression) helps in retaining the other alcohol bottles in the selected objects list.

VI. CONCLUSIONS AND SCOPE FOR IMPROVEMENT

The problem was complex, and the computation resources used for this research was modest. However, this study helped us to understand the power of contemporary object detection models in simplifying, speeding up and enhancing retail checkout experience. Based on the learnings, some of the items that can be taken up for future work are as follows:

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 12, December 2024

|DOI: 10.15680/IJIRSET.2024.1312118|

- Improving on the compute infrastructure and training with more products in scope
- Training the model with more epochs and augmented datasets
- Better and more complex test images
- Optimization of parameters of the NMS algorithm
- Research better model architectures, and hyper-parameter tuning to come to the optimized model parameters

After the initial phase, we can integrate this system with a shopping store app which shall further ease the shopping for the customers and billing for the stores. The same use case can be taken one step further by introducing the concept of inventory management. With that, the inventory of different categories of products will be updated based on the sale of items as detected by this solution. With a bit of effort this can be useful in measuring the performance in the real-world by reconciling the on-shelf counts with the checked-out inventory.

REFERENCES

[1] Schögel, Marcus & Lienhard, Severin Dominic (2020) Cashierless Stores – the New Way to the Customer? Marketing Review St. Gallen

[2] F. Maulana, Nixon, R. P. Putra and N. Hanafiah, "Self-Checkout System Using RFID (Radio Frequency Identification) Technology: A Survey," 2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI), 2021, pp. 273-277, doi: 10.1109/ICCSAI53272.2021.9609762.

[3] Wei Y, Tran S, Xu S, Kang B, Springer M. Deep Learning for Retail Product Recognition: Challenges and Techniques. Comput Intell Neurosci. 2020 Nov 12;2020:8875910. doi: 10.1155/2020/8875910. PMID: 33273903; PMCID: PMC7676964

[4] Z. -Q. Zhao, P. Zheng, S. -T. Xu and X. Wu, "Object Detection With Deep Learning: A Review," in IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11, pp. 3212-3232, Nov. 2019, doi: 10.1109/TNNLS.2018.2876865.

[5] S. Wei, Q. Cui, L. Yang, P. Wang, and L. Liu, "RPC: a large-scale retail product checkout dataset," 2019, https:// arxiv.org/abs/1901.07249.

[6] N. K. Nagam and K. V. S. S. R. S. S. Sarma, "Smart Phone Self Checkout Payments in Super bazaar," 2019 Fifth International Conference on Image Information Processing (ICIIP), 2019, pp. 229-233, doi: 1109/ICIIP47207.2019.8985705.

[7] Lin, Tsung-Yi & Maire, Michael & Belongie, Serge & Hays, James & Perona, Pietro & Ramanan, Deva & Dollár, Piotr & Zitnick, C.. (2014). Microsoft COCO: Common Objects in Context.

[8] https://docs.ultralytics.com/models/yolov8

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details