SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Retail Product Object Detection using YOLOv5 for Automatic Checkout System in Smart Retail Environment

Muhammad Haikal Hamir[1], Mohd Hanapiah Abdullah[1], Mohd Nizar Zainol[1], Syahrul Afzal Che Abdullah[2], Zuraidi Saad[1], Irni Hamiza Hamzah[1], Zainal Hisham Che Soh[1,*]

[1] Electrical Engineering Studies, College of Engineering, Universiti Teknologi MARA (UiTM), 13500 Permatang Pauh, Pulau Pinang, Malaysia
[2] School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia

ABSTRACT

Most supermarket checkout systems use barcode scanners, but some also use QR codes to identify the items being bought. In practise, these methods take a lot of time, need some level of human supervision, and require people to wait in long lines. In this case, we propose a system that make the checkout process at retail store counters faster, more convenient, and less dependent on a human operator. The method uses a computer vision system called a Convolutional Neural Network, which scans objects that are put in front of a webcam to figure out what they are. A retail product classification model based on YOLOv5 object detection network is designed. First, a dataset was obtained by modifying the RPC dataset to reduce the size of the dataset. The dataset then being annotate, augment, and split into training and validation set. Secondly, a YOLOv5 was built and trained by using the datasets obtained. Third, the Machine Learning model based on the training showing the at F1-score of 1 at the confidence of 0.862. However, detection image result shows that model not performing well in detecting and recognizing retail products on a new and random image.

Keywords:
Automatic checkout; Self-checkout; Smart retail; multi-object detection; Deep learning; YOLOv5

1. Introduction

Long lines at the checkout section turned out to be the biggest problem with shopping in stores [1]. Retail shop employees often must organise inventory, restock shelves, and oversee store displays as part of their duties and it include handling customers on the checkout counter. The traditional way of checkout system involves the retail shops employee to move each of the retail products from the counter's conveyer to the barcode scanner and the employee also need to scan the barcode on those products. When there are more items, this procedure will take a long time, and the customers may have to wait a long time to pay the bill [2]. Most people who shop online do so because they don't want to wait in long checkout lines. Customers hate having to stand in long lines to pay for their purchases. Long lines can be annoying, and people often leave if they must wait for too long. Because

* *Corresponding author.*
E-mail address: zainal872@uitm.edu.my

https://doi.org/10.37934/araset.58.2.182195

of this, businesses lose money, customers are less happy, and the customer experience gets worse. To solve this problem, new ways are being thought up to improve shopping and make it easier. For example, by changing the way people shop by getting rid of the checkout counter all together.

The retail business necessitates a substantial quantity of human labour, with product recognition accounting for a major portion of the burden. Checkout has been enhanced throughout history by technology such as barcodes and QR codes since it is one of the most time-consuming experiences for consumers in retailers. With the recent advancements in computer vision, using image recognition technology to automate product detection has become more important and viable solution. Automatic Checkout (ACO), which tries to produce a shopping list from the images of the things to be purchased, is a main user-case of this trend [3].

The recent success of Amazon Go [4] has sparked significant interest in grocery store self-checkouts. Amazon Go shops have removed the need for traditional checkout counters and drastically decreased checkout times. The system uses computer vision and sensor fusion to identify which things have been bought, and then charges clients automatically through a mobile application when they leave the shop. This eliminates the inconvenience of lengthy lines, hence drastically cutting check-out time. Panasonic has designed a self-checkout system [5] based on radio frequency identification (RFID) tags that may be accessed through a walk-through. RFID-based item detection has been developed before [6], but the contribution made here is a cost-effective method for grocery store adoption.

The subject of visual object recognition and classification has been intensively researched by the research community, and work has been done on product detection and classification in grocery shops, particularly on product detection on shelves. Using CHoG descriptor, low-level characteristics are collected from the query picture and submitted to a data server for recognition in the research done by Sam S. Tsai and his fellow researchers [7]. The suggested method focuses primarily on application development and pays very little attention to the issue of recognition. [8] presents a dataset of 120 grocery goods (GroZi-120) using images of products collected from the internet (in-vitro) and extracted from a camcorder video of retail items recorded inside a supermarket (in-situ). On the dataset, colour histogram, SIFT, and boosted Haar-like features were used, and a comparison was made against product type and imaging circumstances.

The paper [9] employs augmented SURF descriptors taken from the GroZi-120 dataset to train a multiclass Naive-Bayes classifier. Shelf Scanner is a method proposed by the authors to assist visually challenged shoppers at grocery shops. This research is most like the work by Bing-Fei Wu, Wan-Ju Tseng, Yung-Shin Chen, Shih-Jhe Yao, and Po-Ju Chang [10], which offers an Intelligent Self-Checkout System (ISCOS) that uses a single camera to identify several goods in real-time. Web scraping is used to obtain product data from the websites of three marketplaces and photos from three image search engines for training a YOLO classifier. The laborious work of human annotating is reduced by applying backdrop subtraction to automatically bound product positions. The system recognises several items in the camera's field of view over a stationary platform until the number of objects exceeds three, at which point it becomes difficult to identify smaller things in the presence of bigger ones. In addition, the approach merely utilises coarse characteristics and does not go into the specifics of each item.

An Intelligent Automatic Checkout System based on computer vision and Deep Learning is offered in this study to solve non-barcode and camera reduction solutions at retail checkout counters. When a consumer places their chosen items on the checkout counter, an ideal ACO system should be able to detect each item and provide a full shopping list in a single look. A single camera situated above the counter adds up the goods overlooked by the system. Furthermore, product identification inhibits fraud by preventing customers from scanning the labels of less priced goods instead of the more expensive ones.

This paper presents a retail product detection for detecting products during automatic checkout in a retail store. This research work only covers the parts on analysing the best model for the ACO. The Deep Learning Technique used in this research is YOLOv5. The remainder of the paper is organized as follows. Section 2 describes the methodology of retail product detection. Next, in section 3 discussed the experimental result of retail product image augmentation and detection performances for YOLOv5 models. Finally, section 4 provides the concluding remarks and points out the ideas for future extension of this work.

2. Methodology

This research proposes an accurate and fast framework for retail product object detection on the checkout counter in retail stores. This is different from other works, which are either accurate but not up to date [11,12] or did not have a high level of accuracy in favour of speed [13]. Figure 1 shows the framework that has been suggested. First, the Retail Product Checkout (RPC) dataset downloaded was reduced from 200 classes object to 26 classes. Then, all images are ensured to be in the same size. Next, the images dataset is augmented using processes such as crop, rotated and noise addition. This makes the training process better by lowering the number of false positives (FPs). The techniques called "data augmentation" to add to dataset, which improves the training process and makes it possible for our model to work with images that have random changes in brightness, contrast, and blurriness [14]. Lastly, the You Only Look Once (YOLOv5) deep neural network is first introduced here for mask detection. This network was just released to speed up and improve the accuracy of legacy YOLO.

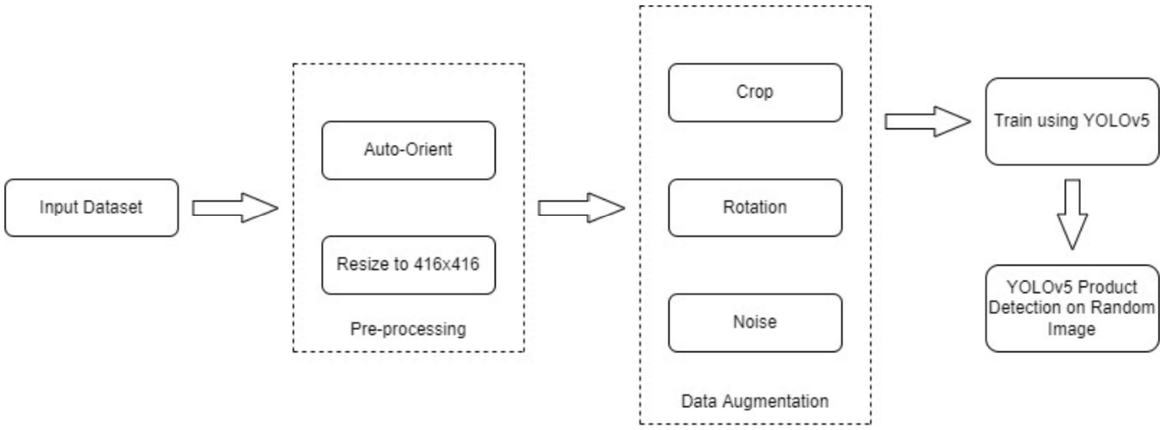

Fig. 1. Methodology framework

2.1 Input Dataset Collection

In the building of a Deep Learning model, a dataset is required to train the model. The training data is used to analyse the model's accuracy, while the testing data is used to train the algorithm created to identify patterns in a dataset. The dataset used to train an algorithm or model so that it can reliably anticipate the result is known as training data.

Research by Xiu-Shen Wei and his colleague propose to produce a Retail Price Checkout (RPC) dataset to overcome the fundamental difficulty in gathering training images that reflect genuine checkout scenarios owing to the goods' constant updating. Despite its practical and research, this issue has received little attention in the computer vision field, owing to a lack of a high-quality dataset. RPC dataset was proved based on research to be the largest dataset for ACO implementation. In terms of product categories (stock keeping units or SKUs) and product images, the RPC dataset is the biggest so far for retail ACO [4]. 200 SKUs were chosen for this dataset and purchased an average of four instances for each SKU, almost doubling the category size of the previous biggest dataset. It takes a total of 83,739 images, including 53,739 single product exemplar shots and 30,000 checkout shots.

Figure 2 shows the comparison of RPC dataset size compared to others such as SOIL-47 [15], Supermarket Produce Dataset [16], Grozi-120 [17], Grocery Products Dataset [18], Freiburg Groceries Dataset [19] and MVTec D2S [20]. Besides that, RPC dataset was collected and divided into two types of images. One type is the image of each product based on Figure 2, and other type is checkout image obtained at the checkout counter. The individual image was taken by camera in every angle while the checkout images were captured and collected based on the realistic checkout scenario which each image will have a different number or type of products in it.

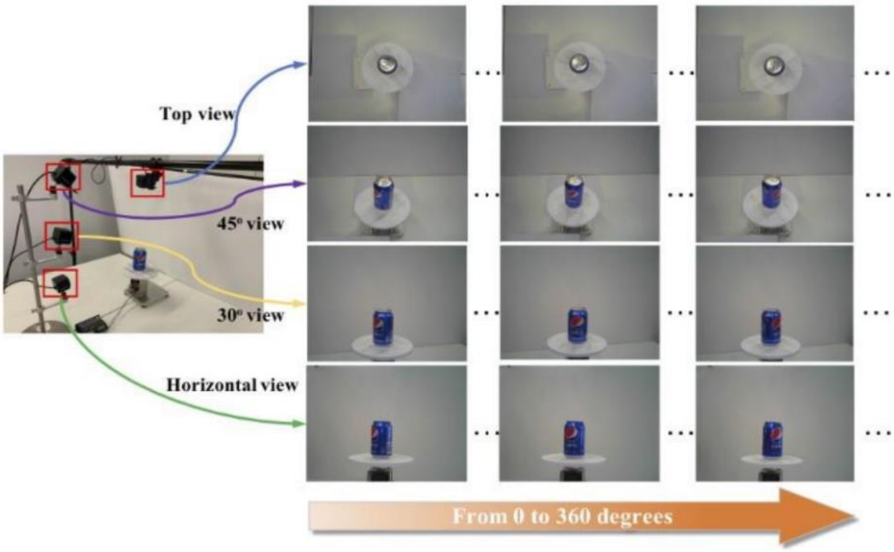

Fig. 2. Collection equipment for single product images

RPC was firstly downloaded in the Kaggle Website. The original RPC dataset contains 200 classes of retail products. To save the computational cost of the training process [21], the RPC dataset was reduced until it contains 26 classes. Those 26 classes of retail products are mostly chosen based on its availability in Malaysia. Each of the classes contain 160 images with different angle of view. The 26 classes in the modified RPC dataset and the sample images are shown in Figure 3.

Fig. 3. Sample images of RPC datasets

2.2 Data Pre-Processing

In the pre-processing phase, this research mainly performs two processes, First, the process is the image undergo auto-orient process. This process can speed up the encoding of the image at the time of capture so that cameras can get good data from their sensors without artefacts if it is implemented in a hardware [22]. Next, the input dataset was pre-processed by stretching the images size into 416x416. In computer vision, resizing images is an important part of the pre-processing step. In general, smaller images help our machine learning models learn faster. If we give the network an image that is twice as big, it must learn from four times as many pixels, which takes time. Also, many deep learning models' architectures need our images to be the same size, but the sizes of the raw images we collect may vary [23].

The images of the gesture are in the data set. For YOLOv5 to be able to train data, the dataset needs to have labels and a bounding box. The value of the coordinate of the annotating box should be between 0 and 1. Roboflow (www.roboflow.com) is an online site that is used to make the bounding box for each image. This site makes it easy to label and annotate data in the format desired [24]. Creating annotation & bounding box to identify the target easily [25].

2.3 Data Augmentation

The datasets obtained have a small amount of data to use to train a model. Data augmentation is a common and well-proven way to make the model more general and stop it from overfitting [26]. Data was added more to the images that are used to train the system. The Roboflow website gives access to training data that has been changed, and it also takes care of labelling the new, changed data. Table 1 shows how the data is increased and up to what values the data is increased. Each image has three versions that have been changed. This means that the number of test images can be up to three times the number of original images.

Table 1
The augmentation technique

Augmentation Technique	Crop	Rotation	Noise
Value	0% Minimum Zoom 30% Maximum Zoom	Between -20% and +20%	Up to 5% of pixel

2.4 Model Training of YOLOv5 for Retail Product Image Detection

There are 3 parts to the data set. About 87% of the data was for the training set, 8% was for validation, and 5% was for the test set. The model is trained on the training set, and, simultaneously, the model evaluation is performed on the validation set after every epoch. The major purpose of separating the dataset into a validation set is to avoid our model from overfitting, i.e., being very excellent at categorizing the samples in the training set but unable to generalize and make correct classifications on new data [27]. The test set will be used to make the model be tested on the data that are new to it. The strategy is to create small adjustments to the YOLOv5s model, which has already been trained on the COCO data set [24]. Transfer learning [28] is used to train on the new set of data, which is relatively small. Also, different processes were used to improve the image, such as HSV, colour spacing, mosaic, and image scaling. Here, this research uses the fine-tuned hyperparameters from the COCO data set, such as SGD optimizer, 0.01 learning rate, 0.0005 weight decay, 100 epoch, batch size 16, and so on [24]. During the training process of the YOLOv5 training, this research used cloud computing at Google-Colab. For the training process, a computer with a GPU (Tesla T4, 1.5 GB, 40 processors assigned by Google) is needed [29].

There are four different ways to find out if the product detected was right. As an example. if the image is mineral water and it is found to be mineral water, the result is true positive (TP). If it is found to be something other than mineral water, the result is false negative (FN). If the image is not mineral water and it is found to not be mineral water, the result is true negative (TN). If it is found to be fire, the result is false positive (FP). The common evaluation metrics for object detection include precision, recall, accuracy, and F1-score. F1-score is the metric that is used to characterize the balance degree between recall and precision, which can be computed as in Eq. (1), Eq. (2), Eq. (3) and Eq. (4) where P is the precision and R is recall.

$$\text{Precision} = \frac{TP}{(TP+FP)} \tag{1}$$

$$\text{Recall} = \frac{TP}{(TP+FN)} \tag{2}$$

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+FP+FP+TN)} \tag{3}$$

$$\text{F1 Score} = \frac{(Precision*Recall)}{(Precision+Recall)} \tag{4}$$

2.5 Detection of Random Retail Product

After the model being trained and being evaluated. The model was then will be test its performance based on the real situation. It is because of to apply the model in the market, the model must give the best results in every way possible. Therefore, an image was taken to test the model. The image contains the products that have the same appearance as in the modified RPC dataset. The image contains a bottle Coke, bottle Sprite, a red sunflower seed, Doublemint Chewing Gum and chocolate M&M. The model will see as performing well when it can detect all of the four products with precise bounding box around the retail products. The image taken is shown in the Figure 4.

Fig. 4. Random sample images of retail product

3. Results

This section discusses the results obtained from performance evaluation of object identification of a retail product using the YOLOv5. Firstly, he data augmentation of RPC dataset using Roboflow are discussed in the next sub section. Next, the performance metric of precision, recall and F1-Score of object detection model the YOLOv5 for retail product on RPC dataset is discussed. Then the model is tested and validated using three batches of validation test data batches. Finally, the random image also tested on the proposed model is also tested.

3.1 Model Data Augmentation

Figure 5 and Figure 7 shown was the original images in the dataset whereas Figure 6 and Figure 8 shows images that were successfully being augmented by using the Roboflow website. The original images shows that there is no adjustment in their images, while the variant images produce from the augmentation process shows that the images have been crop, rotated or noise presence in it.

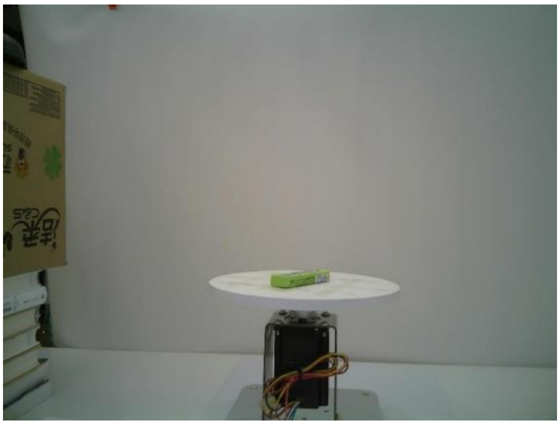

Fig. 5. Sample image of Doublemint chewing gum before augmentation

Fig. 6. Sample of augmented images of Doublemint chewing gum with rotation and noise technique

Fig. 7. Sample of mineral water image before augmentation

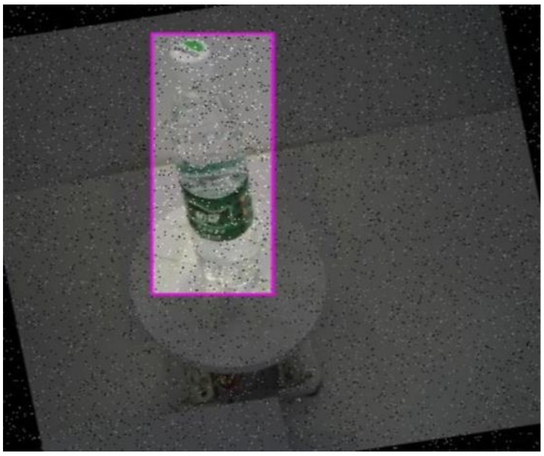

Fig. 8. Sample of augmented mineral water images with crop, rotation, and noise technique

3.2 Model Performance Metric Evaluation

Based on the Figure 9, the confidence value that optimizes the precision and recall is 0.862. In many cases a higher confidence value is desirable. In the case of this model, it may be optimal to select a confidence of 0.8 since the F1 value appears to be 1, which match the maximum value of 1. Observing the recall values in Figure 10 and precision values in Figure 11 at a confidence of 0.8 also confirms that this may be a suitable design point. Starting at about 0.8, the recall value begins to suffer, and the precision value is still roughly at the maximum value.

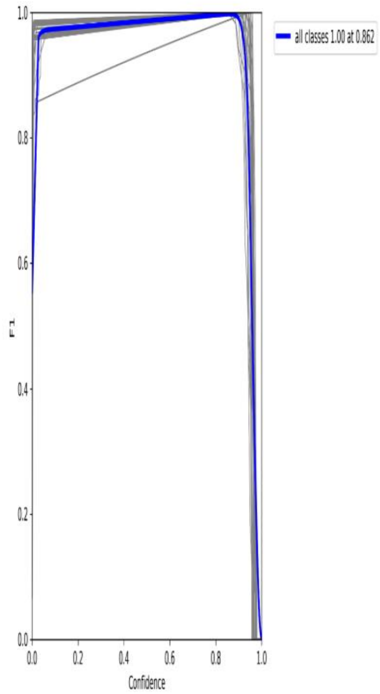

Fig. 9. F1 curve graph

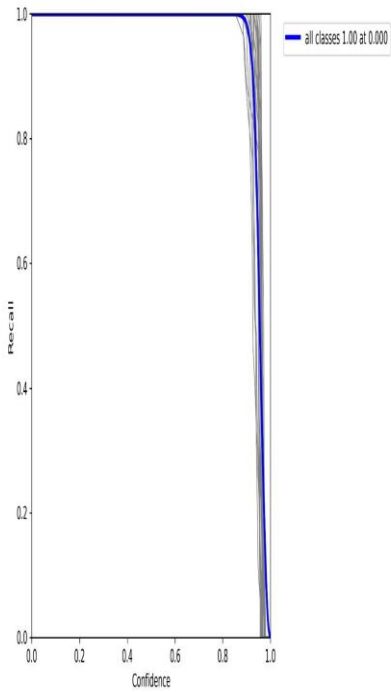

Fig. 10. Recall curve graph

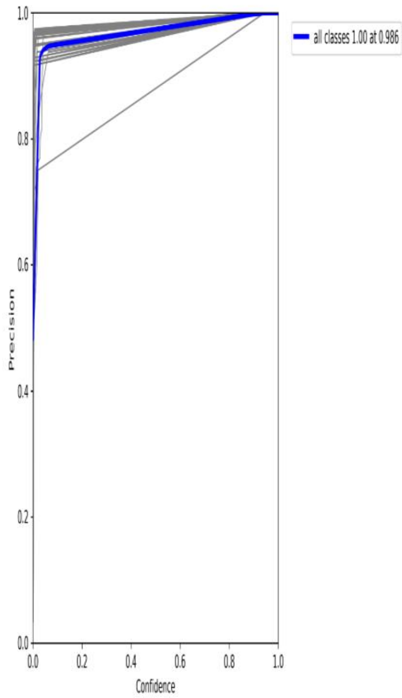

Fig. 11. Precision curve graph

3.3 Model Detection on Validation Images

Figure 12, Figure 14 and Figure 16 shows the batches of validation set images with labels on it while Figure 13, Figure 15 and Figure 17 shows the prediction of the model towards all the batches of validation set images. Most of the detection giving the confidence value of 1 which means that the model confident about its detection. The confidence score shows how likely it is that the box has an object of interest and how sure the classifier is that it does. The confidence score should be zero if there is nothing in that box. Higher confidence value and precise detection shows that the model perform well in detecting the object based on the validation images dataset.

Fig. 12. Labels on the first batch of validation images

Fig. 13. Predictions on the first batch of validation images

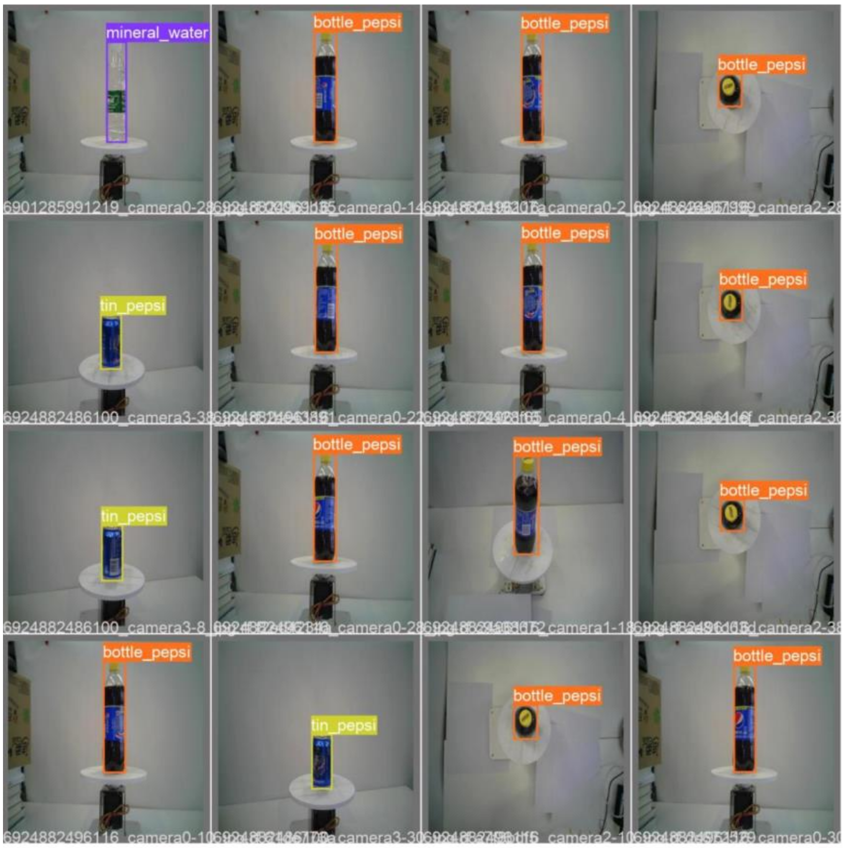

Fig. 14. Labels on the second batch of validation images

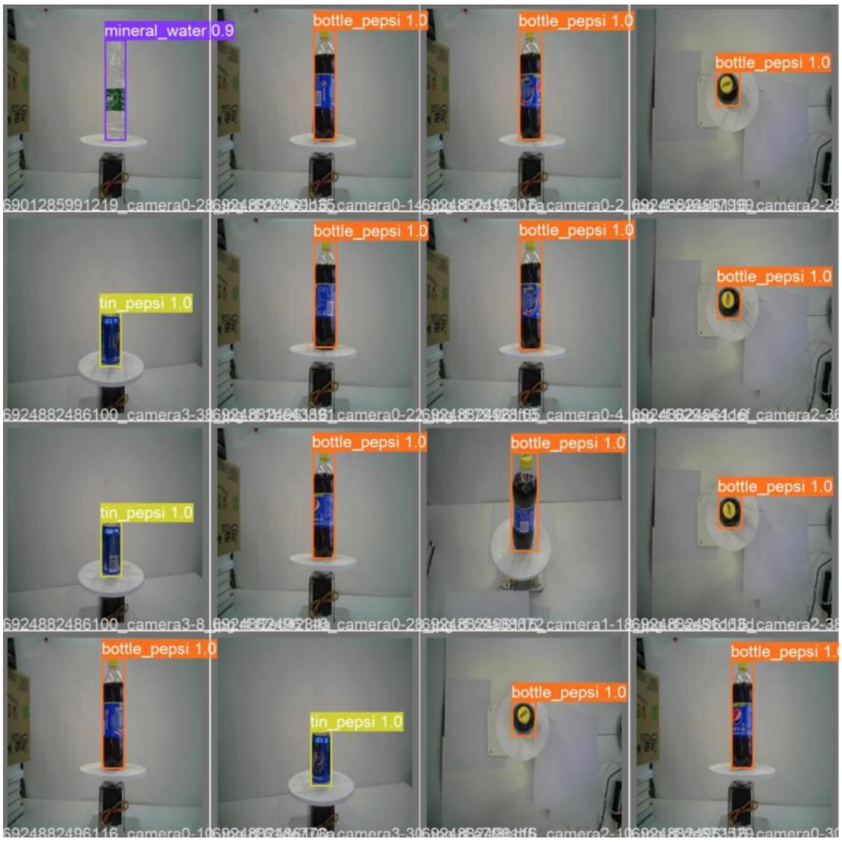

Fig. 15. Prediction on the second batch of the validation images

Fig. 16. Labels on the third batch of validation images

Fig. 17. Prediction on the third batch of the validation images

3.4 Model Detection on Random Images

Figure 18 the results from the model detection on a random image. The image taken consists of a bottle Coke, a red sunflower seed, a bottle sprite, a Doublemint Chewing Gum, and a chocolate M&M. The model was supposed to detect all the object in the image. However, as shown in the Figure 18, the model only detects the bottle coke. The model even did not detect the object of the other products.

Fig. 18. Detection of random images

It shows that the model not performing well in detecting and recognizing retail products on a new and random image. Even though the Figure 9 shows the model performing well in the training and validation process, the ability of the model did not show when detecting random images. In the validation process, the images used are same with the images used in the training process. These results showing that the model was overfitting. Overfitting is a term used in data science to describe when a statistical model fits the training data too well. When this happens, the algorithm cannot work well with data it hasn't seen, which defeats its purpose.

4. Conclusion

In this research, the project addressed the problems that have been faced in the retail industry, which may impact in few aspects involving recognizing product on the checkout counter. Moreover, this research will provide the retail industry with an effective solution to the checkout counter. As mentioned above, a machine learning model for recognizing retail products on the checkout counter successfully being created. The model created was also giving the good performance based on the evaluation metrics during the training and validation process. Therefore, this project has its relevance to the current checkout counter to enhance efficiency and productivity in addition to the hassle-free shopping experience. As for future works and improvement, the model can improve by increasing the input dataset to the training and the model also can be train with large number of epochs. Another improvement that can be done in this research is the user interface. This project only focuses on the retail products detection without having any interface that makes this research become even more impactful to the retail industry. The improvement can be included in using more advanced techniques such as prove a reliable product information by linking the product detected a cloud database, adding the bill generation and paying method based on electronic wallet. This will make this research become applicable and efficient to the retail industry.

Acknowledgement
The author would like to express sincere gratitude to Universiti Teknologi MARA, Cawangan Pulau Pinang for permitting us to use their facility to complete this research work. Special thanks for those who helped directly and indirectly.

References
[1] Gazzola, Patrizia, Daniele Grechi, Ilaria Martinelli, and Roberta Pezzetti. "The innovation of the cashierless store: A preliminary analysis in Italy." *Sustainability* 14, no. 4 (2022): 2034. https://doi.org/10.3390/su14042034
[2] Sugadev, M., Kudapa Sucharitha, I. Rexline Sheeba, and Balamurugan Velan. "Computer vision based automated billing system for fruit stores." In *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pp. 1337-1342. IEEE, 2020. https://doi.org/10.1109/ICSSIT48917.2020.9214101
[3] Wei, Xiu-Shen, Quan Cui, Lei Yang, Peng Wang, and Lingqiao Liu. "RPC: A large-scale retail product checkout dataset." *arXiv preprint arXiv:1901.07249* (2019).
[4] Polacco, Alex, and Kayla Backes. "The amazon go concept: Implications, applications, and sustainability." *Journal of Business and Management* 24, no. 1 (2018): 79-92.
[5] Panasonic. "RFID Based Walk-through Checkout Solution for Future Retail | Panasonic Newsroom Global." https://news.panasonic.com/global/topics/2018/55288.html
[6] Khan, Mohammad Zuber. "Generic RFID application framework." *M. Eng. thesis, Indian Institute of Technology, Kanpur, India* (2008).
[7] Tsai, Sam S., Huizhong Chen, David Chen, Ramakrishna Vedantham, Radek Grzeszczuk, and Bernd Girod. "Mobile visual search using image and text features." In *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pp. 845-849. IEEE, 2011. https://doi.org/10.1109/ACSSC.2011.6190127
[8] Merler, Michele, Carolina Galleguillos, and Serge Belongie. "Recognizing groceries in situ using in vitro training data." In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8. IEEE, 2007. https://doi.org/10.1109/CVPR.2007.383486
[9] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." *arXiv preprint arXiv:1804.02767* (2018).
[10] Wu, Bing-Fei, Wan-Ju Tseng, Yung-Shin Chen, Shih-Jhe Yao, and Po-Ju Chang. "An intelligent self-checkout system for smart retail." In *2016 International Conference on System Science and Engineering (ICSSE)*, pp. 1-4. IEEE, 2016. https://doi.org/10.1109/ICSSE.2016.7551621
[11] Militante, Sammy V., and Nanette V. Dionisio. "Deep learning implementation of facemask and physical distancing detection with alarm systems." In *2020 third international conference on vocational education and electrical engineering (ICVEE)*, pp. 1-5. IEEE, 2020. https://doi.org/10.1109/ICVEE50212.2020.9243183
[12] Zhang, Edward. "A real-time deep transfer learning model for facial mask detection." In *2021 Integrated Communications Navigation and Surveillance Conference (ICNS)*, pp. 1-7. IEEE, 2021. https://doi.org/10.1109/ICNS52807.2021.9441582
[13] Khamlae, Ponlawat, Kingkarn Sookhanaphibarn, and Worawat Choensawat. "An application of deep-learning techniques to face mask detection during the COVID-19 pandemic." In *2021 IEEE 3rd global conference on life sciences and technologies (LifeTech)*, pp. 298-299. IEEE, 2021. https://doi.org/10.1109/LifeTech52111.2021.9391922
[14] Youssry, Nouran, and Ahmed Khattab. "Accurate real-time face mask detection framework using YOLOv5." In *2022 IEEE International Conference on Design & Test of Integrated Micro & Nano-Systems (DTS)*, pp. 01-06. IEEE, 2022. https://doi.org/10.1109/DTS55284.2022.9809855
[15] Koubaroulis, D., J. Matas, J. Kittler, and CTU CMP. "Evaluating colour-based object recognition algorithms using the soil-47 database." In *Asian Conference on Computer Vision*, vol. 2, p. 2. 2002.
[16] Rocha, Anderson, Daniel C. Hauagge, Jacques Wainer, and Siome Goldenstein. "Automatic fruit and vegetable classification from images." *Computers and Electronics in Agriculture* 70, no. 1 (2010): 96-104. https://doi.org/10.1016/j.compag.2009.09.002
[17] Merler, Michele, Carolina Galleguillos, and Serge Belongie. "Recognizing groceries in situ using in vitro training data." In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8. IEEE, 2007. https://doi.org/10.1109/CVPR.2007.383486
[18] George, Marian, and Christian Floerkemeier. "Recognizing products: A per-exemplar multi-label image classification approach." In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part II 13*, pp. 440-455. Springer International Publishing, 2014. https://doi.org/10.1007/978-3-319-10605-2_29

[19] Jund, Philipp, Nichola Abdo, Andreas Eitel, and Wolfram Burgard. "The freiburg groceries dataset." *arXiv preprint arXiv:1611.05799* (2016).
[20] Follmann, Patrick, Tobias Bottger, Philipp Hartinger, Rebecca Konig, and Markus Ulrich. "MVTec D2S: densely segmented supermarket dataset." In *Proceedings of the European conference on computer vision (ECCV)*, pp. 569-585. 2018. https://doi.org/10.1007/978-3-030-01249-6_35
[21] Lopez-Tapia, Santiago, Rafael Molina, and Nicolás Pérez de la Blanca. "Deep CNNs for object detection using passive millimeter sensors." *IEEE Transactions on Circuits and Systems for Video Technology* 29, no. 9 (2017): 2580-2589. https://doi.org/10.1109/TCSVT.2017.2774927
[22] Roboflow. "When Should I Auto-Orient My Images?" https://blog.roboflow.com/exifauto-orientation
[23] Hashemi, Mahdi. "Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation." *Journal of Big Data* 6, no. 1 (2019): 1-13. https://doi.org/10.1186/s40537-019-0263-7
[24] Dima, Tasnim Ferdous, and Md Eleas Ahmed. "Using YOLOv5 algorithm to detect and recognize American sign language." In *2021 International Conference on information technology (ICIT)*, pp. 603-607. IEEE, 2021. https://doi.org/10.1109/ICIT52682.2021.9491672
[25] Jindal, Manik, Nikhil Raj, P. Saranya, and V. Sundarabalan. "Aircraft detection from remote sensing images using YOLOV5 architecture." In *2022 6th International Conference on Devices, Circuits and Systems (ICDCS)*, pp. 332-336. IEEE, 2022. https://doi.org/10.1109/ICDCS54290.2022.9780777
[26] Aquino, N. Romero, Matheus Gutoski, Leandro Takeshi Hattori, and H. Silvério Lopes. "The effect of data augmentation on the performance of convolutional neural networks." *Braz. Soc. Comput. Intell* (2017). https://doi.org/10.21528/CBIC2017-51
[27] Bai, Yu, Minshuo Chen, Pan Zhou, Tuo Zhao, Jason Lee, Sham Kakade, Huan Wang, and Caiming Xiong. "How important is the train-validation split in meta-learning?" In *International Conference on Machine Learning*, pp. 543-553. PMLR, 2021.
[28] Olivas, Emilio Soria, Jos David Mart Guerrero, Marcelino Martinez-Sober, Jose Rafael Magdalena-Benedito, and L. Serrano, eds. *Handbook of research on machine learning applications and trends: Algorithms, methods, and techniques: Algorithms, methods, and techniques*. IGI global, 2009. https://doi.org/10.4018/978-1-60566-766-9
[29] Magalhães, Sandro Augusto, Luís Castro, Germano Moreira, Filipe Neves Dos Santos, Mário Cunha, Jorge Dias, and António Paulo Moreira. "Evaluating the single-shot multibox detector and YOLO deep learning models for the detection of tomatoes in a greenhouse." *Sensors* 21, no. 10 (2021): 3569. https://doi.org/10.3390/s21103569