IOT BASED QUICK RETAIL CHECKOUT USING MACHINE LEARNING AND RFID TECHNOLOGY

[1]Reuben Pinto, [2]Aisha Fernandes

[1]PG student, [2]Associate Professor
[1,2]Department of Information Technology,
[1,2]Goa College of Engineering, Farmagudi, Goa, India.

Abstract : Businesses these days have moved to an online segment, but a few of them still offer to keep stores where the customer can have a more personal touch and judge for themselves what it is that they are spending their money on. Getting all the items in the shopping cart is all well and good but when it comes time to checkout the efficiency of the process is something that can make or break a business in terms of retaining a customer. The most cost effective method used is to use barcodes and scan them individually at checkout. To improve on efficiency an idea is proposed to use machine learning models for object recognition and backup the product count with radio frequency identification technology. The concept attempts to make use of system on chip devices to minimize size and cost of the system.

***IndexTerms* - Machine Learning, IoT, RFID, Retail Checkout, YOLOv3.**

I. INTRODUCTION

We live in a world where instant gratification is a must and where customers expect to have anything they want at the tip of their hands. With online shopping on the rise and the convenience of the 'buy now' button, shoppers have become increasingly impatient when it comes to retail services. The problem of optimising checkout in a retail setup is one that has plagued businesses for a long time. Many attempts are being made to incorporate emerging technologies to help improve customer satisfaction during checkout. Recently attempts were made by corporate giants Amazon to introduce a grab and go concept.The store was the first with the "Amazon Dash Cart." This lets shoppers filling up to two grocery bags forgo the store's cashiers via technology systems that discern what they put in the cart and bill a credit card on file once customers leave through a designated lane [1]. But such sophisticated setups is not something that everyone can shell out to have for their stores. Thus this project proposes an alternative concept to tackle the retail checkout problem.

The proposed solution is to train a machine learning model to recognize products in a retail setup and add it to the billing system to be totalled up. If a camera is placed in such a way that it overlooks a cart placed below at the checkout counter it can take still images or a video feed that can be an input to the product recognition system. Shopping carts usually have items stacked within them which obscures the items below from the line of sight of the camera. To overcome this problem Radio Frequency Identification(RFID) readers are to be placed surrounding the checkout station in close proximity of where the shopping cart is going to be placed or on the cart itself. The RFID reader is placed to capture the presence of those items that are obscured from the camera. This however does have some pitfalls which will be discussed in a later section.

II. RELATED WORK

Machine Learning and Artificial intelligence has been around for quite some time even though it has been in prime focus in recent times. Object detection algorithms using Convolutional Neural Networks(CNN) have seen vast improvements in prediction accuracy and faster detection intervals.

Earlier methods of object recognition were based on Scale Invariant Feature Transform(SIFT) and Histogram of Oriented Gradient(HOG) which are blockwise orientation histograms. The method evaluated well-normalized local histograms of image gradient orientations in a dense grid. Dalal et al. [2], used locally normalized histogram of gradient orientations features similar to SIFT descriptors in a dense overlapping grid to obtain good results for person detection.

Girshik el al. [3], developed an object detection algorithm that performed 30 percent better than other detection algorithms at the time. They applied a high capacity CNN to bottom-up region proposals to localize and segment objects.

Deep learning methods were used by Wang et al. [4] to perform video saliency based detection. The module used was split into two to handle static and dynamic saliency. Static saliency network used image inputs passed through a CNN to detect objects based on striking contrast differences. The dynamic saliency network processed video by splitting the frames and keeping track of objects detected by the network in spatial and temporal space.

III. BACKGROUND

In the proposal the main concepts touched upon are training object detection models to recognise items in a cart and use of RFID reader to recognise tags placed on the item. Most machine learning algorithms consist of convolutional neural networks and support vector machine(SVM) classifiers. RFID systems also differ on tag types. Both of these concepts will be explained below.

3.1 Convolutional Neural Networks

CNNs belong to a class of neural networks that allow better extraction of features from input images. Unlike classical models, CNNs take image data, train the model, and then classify the features automatically for healthier classification. There are three principal components of CNNs: convolution, max pooling, and activation function. CNNs are used in many applications like image recognition, face recognition, and video analysis [5].

Convolution layers involve applying a filter mask to the input image to extract features. The extracted feature map is sensitive to the location of the feature in the map. In order to resolve the sensitivity, down sampling can be done by pooling. Most common pooling operations are max and average pooling. Pooling involves operations of reducing the feature map in patches while maintaining feature information. CNNs are made up of a sequence of convolution and pooling layers [6].

When the image passes through the layers, the distinctly required features are kept in the convolution layers, and thanks to the pooling layers, these features are intensified and kept over the network, while discarding all the information that doesn't make a difference for the task.

3.2 Support Vector Machines

Support vector machines chained with CNNs perform classification in object detection systems. The main objective of SVMs is to determine a separating hyperplane in an N-dimensional space that can distinctly separate data points in their respective classes.

Support vectors are data points close to the hyperplane that impact the positioning and orientation of the hyperplane. By applying optimizing adjustments to these vectors the precision of the classifier can be improved [7].

3.3 RFID Systems and Tag Types

RFID systems are mostly made up of a reader ,an antenna and a tag. Tags contain circuitry to store and broadcast information [8]. The information is communicated over radio frequencies. The type of tag used defines the type of the RFID system being used.

Passive tags are smaller, cheaper and thinner or more flexible. They contain no internal power source and are activated only in the presence of a radio frequency(RF) wave that is generated by the reader. The tag's internal antenna draws energy from the RF wave and powers up the circuitry which generates a signal back into the system. This signal change called backscatter is detected by the reader's antenna and interpreted into usable information. These types of tags operate on different frequencies such as low(125-134 KHz), high(13.56MHz) and ultra high(856-960 MHz) frequency.

Active tags are more robust and are similar in composition of the antenna and circuitry with the only addition of having its own power source. Two main frequencies they operate over are 433MHz and 2.45GHz. Main categories of active tags are transponders and beacons. Transponders wait for a reader signal before it broadcasts its own signal while beacons broadcast their signal at regular intervals.

IV. THE PROPOSED SOLUTION

The proposed retail checkout system accumulates a tally of products in two phases:

1. Object detection using machine learning
2. RFID tag detection

Figure 1 below illustrates the data flow of the process. The flow on the left depicts the first phase where an image is captured of the shopping cart and provided to a machine learning model as an input. The machine learning model used here is the Darknet You Only Look Once version 3(YOLOv3) model. In this solution the cart items are assumed to be tagged with RFID chips. The flow on the right depicts the second phase where items are detected by the RFID reader as they are placed in the cart.

The outcomes of both phases are a list of detected objects that are then cross checked with each other to remove possible duplicates and are added to the customers cart for billing.

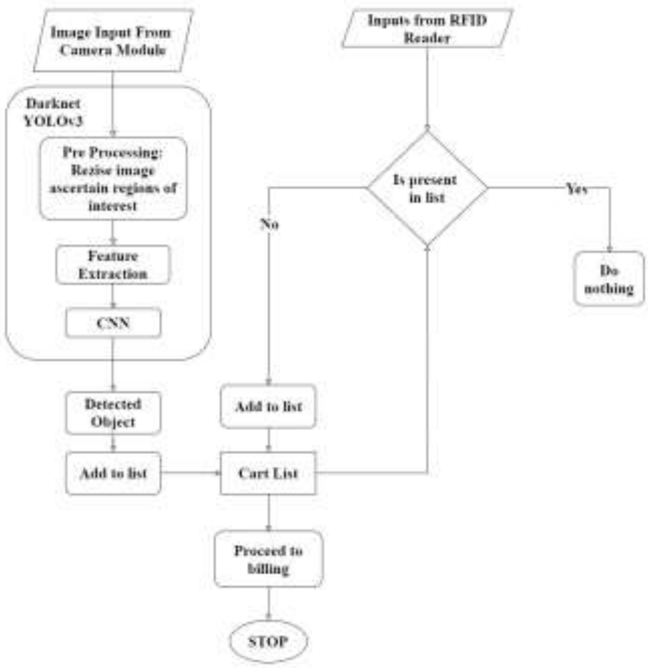

Figure 1:Data Flow of the Proposed System

V. ADVANTAGES

The solution put forward has stemmed from a need to reduce queue lines at checkout and improve checkout efficiency. Implementing this solution will help make billing in retail stores much less time consuming and lead to better customer satisfaction.

The system can be incorporated to adjust stock accordingly and link into managing the supply chain automatically as well. Prices of systems on chip devices and IoT sensors used are far more cost effective than having to buy expensive Point of Sale(POS) units and bar code readers.

VI. Components of the System

The system consists of running python scripts on a Raspberry Pi module that has an MFRC522 RFID reader module and Pi Camera connected to it. Using TensorFlow library and the Darknet YOLOv3 model image detection is carried out. We describe in brief the software and hardware components used below.

6.1 TensorFlow

TensorFlow was created by the Google Brain team and initially released in November 2015. It is a free open-source library written in python that was created for building and working with machine learning models. It has the ability to run on multiple central processing units(CPU) and graphical processing units (GPU) (with optional compute unified device architecture(CUDA) and SYCL extensions for general-purpose computing on graphics processing units) [9]. It supports multiple platforms such as 64-bit Linux, macOS, windows and Android and iOS platforms. Keras, a library built on top of TensorFlow, is used to construct deep learning algorithms.

6.2 Darknet YOLOv3

Redmon et al., [10] proposed a new approach to object detection. They designed an effective approach to determine bounding boxes and class probabilities of detected objects in a single evaluation. The YOLOv3 model can process images at 45 frames per second in real-time with a mean average precision(mAP) of 51.5 percent which is better than other real-time detectors.

Darknet is an open source framework written C and CUDA which is optimized to carry out parallel processing of CPU and GPU computations [11]. The YOLO model works by applying a single neural network to the entire image, it then divides the image into regions and predicts bounding boxes around objects with probabilities assigned for each region. The model was trained using darknet on the COCO dataset. Incremental research on improving the model to create YOLOv3 has resulted in a new network to perform feature extraction that consists of 53 convolutional layers, a new detection metric and computing a score for objectness for every bounding box using logistic regression [12].

6.3 Raspberry Pi 4 Model B

Raspberry Pi is a system on chip, miniature single-board computer designed by Raspberry Pi Foundation in association with Broadcom. There are a number of models released for numerous versions of Raspberry Pi. The first generation release is a model B followed by a simpler and cheaper Model A. Models B+ and A+ are improvements on the earlier released models. The current latest Raspberry Pi is a Raspberry Pi 4 Model B.

The Raspberry Pi 4 Model B boasts of a 1.5 GHz Broadcom BCM2711 processor, 1,2,4 and 8 GB variants of Memory with a 40 pin general purpose input output (GPIO) pins. It also has network support via Ethernet, dual band Wireless network support and Bluetooth capabilities. It has audio support via a 3.5mm audio jack. Display support is provided by two micro High Definition Multimedia Interface(HDMI) ports. It also has a 2-lane mobile industry processor interface(MIPI) for display and camera serial interfaces. It operates at a low voltage of 5V [13].

6.4 Pi Camera

The Pi camera module is a portable lightweight camera that is used in conjunction with Raspberry Pi. It communicates with Pi using the MIPI camera serial interface protocol. The camera is used due to its small payload of data transferred. It is a 5 megapixel camera that supports a 2592 by 1944 resolution. It has 1080p, 720p and 480p image quality support [14].

6.4 MFRC 522

The MFRC522 module is a highly integrated reader and writer IC for contactless communication at 13.56 MHz. The internal transmitter of the module is able to operate a reader/writer antenna designed to communicate with ISO/IEC 14443 A/MIFARE cards and transponders without the need to provide additional active circuitry. The RFID tags that are used along with the MFRC522 reader are passive in the sense that the transponder in the tags has no power source and utilizes the energy emitted by the reader antenna to run its circuitry and reflect the signal to the reader.

Communication modes supported by the module are the Serial Peripheral Interface(SPI), Serial Universal Asynchronous Receiver/Transmitter(UART) and I2C bus interface [15].

VII. Implementation

Inorder to detect objects in the shopping cart various object detect methods were experimented with. Inputs were taken as both images as well as videos inputs. In the case of video inputs frames were extracted using the openCV library an these frames were then fed to the detection models to predict object detection.

The initial model under consideration was the VGG 16 model trained on the ImageNet dataset. While it did detect objects with 90 percent accuracy, its large size and processing requirements would not be compatible to run on devices with lower processing capabilities. The next alternative obtained was the YOLOv3 model trained on COCO dataset. This model had an alternative variant called YOLOv3-tiny which had a smaller size being able to run on constrained environments but still showed an impressive detection rate of 244 frames per second on a single gpu which is faster than its larger counterpart.

In order to train the model on the products to be identified 100 images were taken of each product under consideration. Bounding boxes were drawn and classes assigned to get annotated training images that were then supplied to train the model. Training is performed for 10000 epochs for 5 classes to obtain usable accuracy of abovr 80 percent. The YOLOv3-tiny model with the trained weights were then uploaded to Raspberry Pi.

The entire setup was done as shown in Fig.2. The Pi Camera was connected to the Raspberry Pi via the camera serial interface.

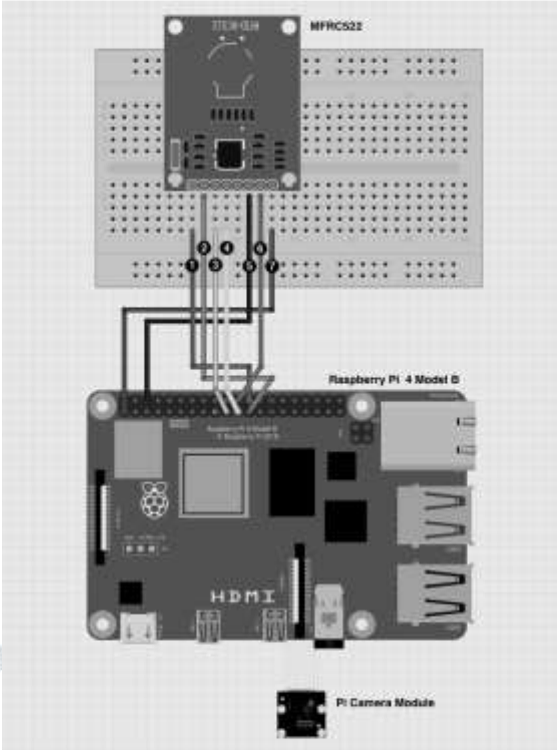

Figure 2: Setup of Raspberry Pi 4 Model B with the Pi Camera and MFRC522

MFRC522 communicates via the serial peripheral interface with the Raspberry Pi and is connected as shown in Table.1.

Table 1: Connection configurations between MFRC522 and Raspberry Pi GPIO Pins

Connection No.	MFRCC 522 Pin	GPIO Pin
1	SDA	Pin 24
2	SCK	Pin 23
3	MOSI	Pin 19
4	MISO	Pin 21
5	GND	Pin 6
6	RST	Pin 22
7	3.3v	Pin 1

The RFID tags placed on the items are written with item information. As items are placed into the cart the RFID reader notes down the item information. The Pi Camera is placed in front of the cart and a still image is taken. This image is then provided as input to the trained YOLOv3 model. The model recognises objects and provides back bounding box and class annotations which are drawn on the image. Items detected are cross checked with those read by the RFID reader and a final cart items list is generated ready for billing.

Figure 3 shows the data read by the RFID reader and Fig.4 shows the bounding boxes placed around identified objects. Objects are at times misclassified which are discarded when checked against the list obtained from the RFID reader.

```
721391411828
apple
42599310618
sugar
42599310618
sugar
143039628092
orange
```

Figure 3: Scanned tag data from the RFID reader

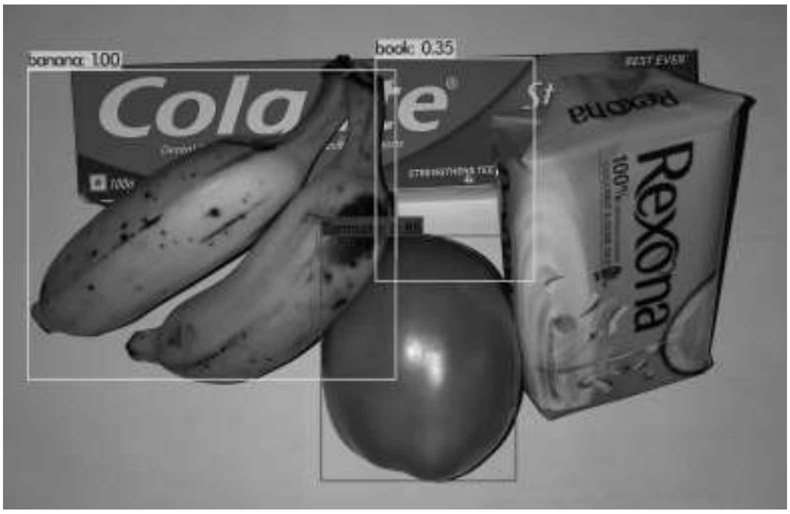

Figure 4: Cart image output with prediction bounding boxes from YOLOv3

VIII. Conclusion

Proposed solution will hasten checkout times thus improving the store efficiency. The multiple object detection done by training a YOLOv3 model grouped with reinforcing the item count with RFID tag reading will help speed up billing by possibly performing a single glance detection in an instant.

Tagging all products with passive RFID tags is not a very cheap solution and also suffers for detection range. Alternative cheaper options can be looked into to replace this component of the system.

IX. Future Scope

Future prospects could include incorporating the idea into building a virtual customer profile where a customer with payment information connected could enter a store pickup the items that would be recognised in the cart itself and walk out of the store concluding a transaction and getting billed automatically.

References

[1] https://www.reuters.com/article/us-amazon-com-store-idUSKBN25N0QF.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 886-893, 2005.

[3] R. Girshick, J. Donahue, T. Darrel and J. Malik, “Rich Feature Hierarchies for Accurate Object Detection and Se-mantic Segmentation”, in Proceedings of IEEE Conference on Computer Vision and PatternRecognition, pp. 580–587, 2014.

[4] W. Wang, J. Shen, and L. Shao, “Video Salient Object Detection via Fully Convolu-tional Networks”, Vol.27, No.1, pp. 38–49, 2018.

[5] https://towardsdatascience.com/understanding-convolutions-and-pooling-in-neural-networks-a-simple-explanation-885a2d78f211.

[6] https://machinelearningmastery.com/pooling-layers-for-convolutional-neural-networks.

[7] https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47.

[8] https://www.atlasrfidstore.com/rfid-insider/active-rfid-vs-passive-rfid.

[9] https://en.wikipedia.org/wiki/TensorFlow.

[10] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 779-788, 2016.

[11] J. Redmon, “Darknet: Open Source Neural Networks in C”, Available at: https://pjreddie.com/darknet/. Accessed on 01 May 2021.

[12] J. Redmon and A. Farhadi. “Yolov3: An incremental improvement.”, Technical Report, arXiv preprint arXiv:1804.02767, 2018.

[13] https://www.raspberrypi.org.

[14] https://components101.com/misc/pi-camera-module.

[15] https://www.nxp.com/docs/en/data-sheet/MFRC522.pdf.