e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:11/November-2021 Impact Factor- 6.752 www.irjmets.com

A REVIEW ON SMART RETAIL SHOP USING ARTIFICIAL INTELLIGENCE (AI) AND INTERNET OF THINGS (IOT)

Shruti Chetan Margaje*1, Varsha Rasal*2

*1,2Computer Engineering, NBN Sinhgad School Of Engineering, Pune, Maharashtra, India.

ABSTRACT

Smart Retail Shops are bringing about a great change in the shopping experience of the customer as well as the backend of the whole procedure taking place during shopping. Previously many issues were faced in terms of what the product is being taken from the vending machine, as it plays an important role in the smart unstaffed retail shop. Hence these problems are overcome with the emerging technologies namely Artificial Intelligence (AI) and the Internet of things (IoT). In this review paper, it is presented of how AI can be beneficial to enhance the recognition and object detection accuracy in goods detection using appropriate methods. Recent advancement in deep learning has shown high potential in variable tasks of image classification, object detection, and recognition. Taking into consideration image classification and object detection, it is a more challenging task to solve object detection. For this, some of the main methods for objection detection are examined here. And based on the result SSD algorithm is employed to analyze the multi-target features of the goods present. The SSD algorithm is further enhanced to improve the recognition accuracy by adding sub prediction structure.

Keywords: Artificial Intelligence (AI), Internet Of Things (Iot), Single Shot Multi-Box Detector (SSD), Object Detection.

I. INTRODUCTION

Recent progress in the field of the Internet of Things (IoT), Artificial Intelligence (AI), online mobile payments, and the use of the internet in day-to-day life has popularized the concept of the smart unstaffed retail shop to a greater extend. In a smart retail shop, the customer can complete the entire process with ease on their own without any assistant.

With the help of IoT, it has become possible to connect smart hardware systems to the internet. Vending Machine plays an important role here, wherein the customer can interact and shop for things. Artificial Intelligence (AI) technology is applied for the detection and recognition of objects in the vending machine. With the help of this Artificial Intelligence (AI), now it possible to find out what the object is being taken and also helps to maintain records.

In this work, the architecture of a smart unstaffed retail shop scheme is also reviewed for a better understanding of the process taking place fig 1. An integrated system of smart hardware and intelligent technology has made it possible for such an unstaffed shop scheme. In these unstaffed shops, the goods stored in vending machines are mostly canned or have caps. With the use of Artificial Intelligence, the top and body features of the items are analyzed to get the best classification results. A dual-camera system is used to capture the inside images of goods placed in the vending machine

II. BACKGROUND

The unmanned retail shop is getting great growth with the rise of online flow. In such cases two main problems which are to be addressed for commodity and customer are as mentioned:

a) Customer Identification

For information binding and identity, identification can be achieved by using smartphones and biometrics features recognition.

b) Commodity Recognition

For the identification of commodities, some of the technologies used are RFID, gravity sensing, and machine vision. In unstaffed retail shops, the vending machine acts as an important factor. Different vending machines have different techniques used for commodity recognition. Conventionally these machines face a high commodity loss rate for which it becomes difficult to manage goods efficiently. Radio Frequency Identification

(RFID) Technology is the most used technology. Later it is observed to have high operational costs. The gravity sensing method has good detection ability of the commodities placed in the vending machine but is unable of recognizing what the customer is taking. Hence with the help of Artificial Intelligence, the detection and recognition of the stock-keeping units become possible.

The basic architectural design of a smart unstaffed shop is as shown in fig 1

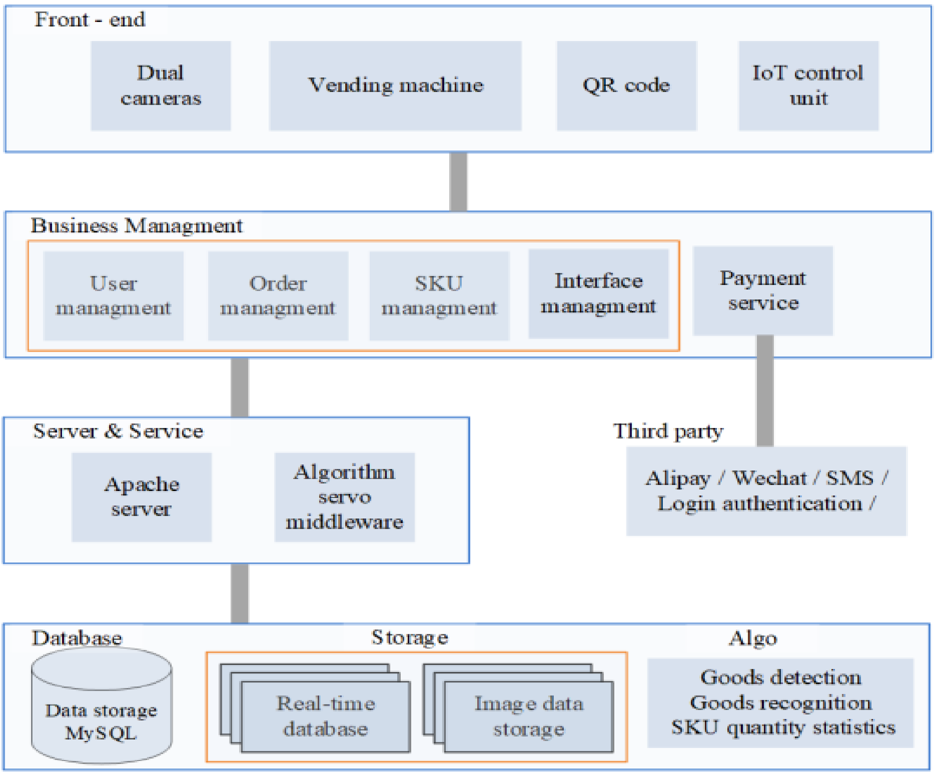

Fig 1: Basic architecture of unstaffed retail shop [1].

III. LITERATURE SURVEY

The goods in the vending machine are classified and analyzed using the Deep Learning method. AI detection and identification of the stock-keeping units are used for locating the goods in the raw image taken by the dual camera. The goods can be detected using certain parameters like color feature, mathematical morphology, texture feature, and depth feature [1]. Amongst these, for depth feature-based object detection methods some of them are RCNN, YOLO, and SSD. After this, these models are trained, SSD outperforms YOLO in detection speed and is Close to RCNN in terms of accuracy [1]. Based on these observations the SSD Model is employed. Further SSD model with a sub-prediction structure is designed to enhance the recognition rate of the goods present in the vending machine in practical scenarios.

Short briefing of the methods mentioned above:

A. Fast R-CNN

Fast Region-based Convolutional Network (Fast R-CNN) is a method that is used for object detection. It is employed to improve the training and testing speed along with increasing the accuracy. Hence it is considered comparatively faster than R-CNN. The advantage of fast R-CNN is that it has higher detection quality than R-CNN. The architecture of Fast R-CNN:

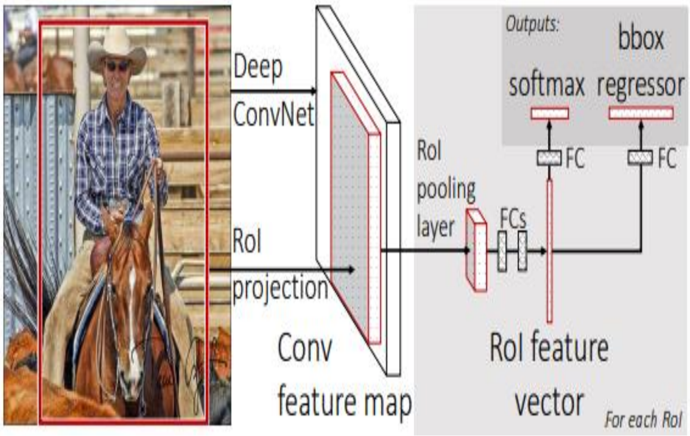

Fig 2: Architecture of Fast R-CNN [2].

At first, the image is fed to the network along. The convolutional network processes the given input image and outputs many features. It is called the conv feature map. This is because the detected features are stored relative to the original input image where they are detected. Then using the region proposal it is easy to extract the region of interest from the feature map, this process is called the region of interest projection (RoI). With the help of the region of interest pooling layer, it is possible to downsample this feature map to get a fixed-length feature map of definite height and length. The RoI feature can be then used to get two outputs. in the first output, it uses the fully connected layer and then the softmax for the classification of the image. The second output uses the fully connected layer and then the bounding box regression outputs for getting the size location of the classified image.

Drawbacks of Fast R-CNN:

- Using the selective search algorithm is slow and also time-consuming
- It takes about 2 sec per image to detect objects which sometimes may not work properly for big data in real life.

B. Faster R-CNN

Advancement of region proposal methods and region-based convolutional neural networks in object detection, Faster R-CNN is the next step that can solve some of the problems of Fast R-CNN. It is the most classic algorithm used for object detection. The advantage is that the problems of Fast R-CNN are solved using the Region Proposal Network (RPN) instead of selective search. It is much faster than Fast R-CNN because of RPN and it can be used in real-time object detection.

The architecture of Faster R-CNN:

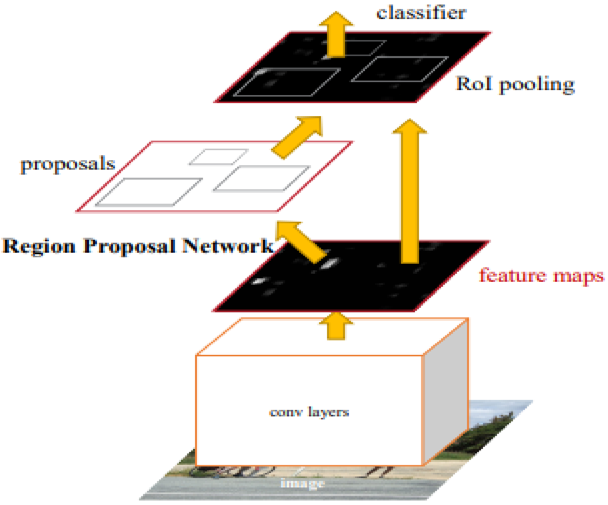

Fig 3: Architecture of Faster R-CNN [3].

Faster R-CNN architecture consists of two parts: Region of Interest (RoI) and Region Proposal Network. The convolution layer consists of the convolution and the pooling layers. It is used for extracting the feature maps. These feature maps are then later used by the Region proposal networks (RPN) layer and fully connected layers (FC). In the Region proposal networks (RPN), region proposals are generated. It uses the softmax and regression bounding box for classifying and correcting the proposals. Collected feature maps and the proposals pass through the region of interest pooling and are then sent to a fully connected layer for the determination of the target category.

The drawback of Faster R-CNN:

- When extraction of all the samples from the single image with the help of the Region proposal network, there are more chances that they have similar features. Due to this, it may take longer than expected.

C. YOLO

You only look once (YOLO) models are a series of end-to-end deep learning models for object detection. It involves a single conv neural network that splits the image into grids and then each of them is bounded by a box. The advantage of YOLO is that the speed of detection is highly improved here. With minimum background errors, it predicts accurate results.

The architecture of YOLO :

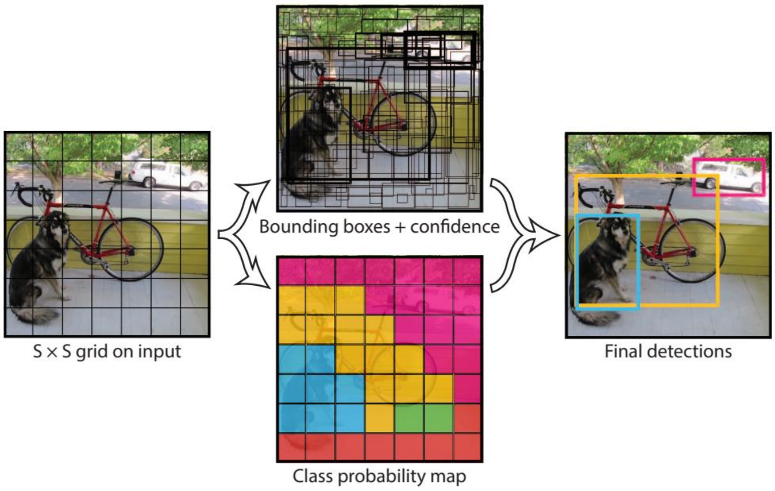

Fig 4: Basic flow of YOLO[4

YOLO works using the following techniques like Residual blocks, Bounding box regression, and Intersection Over Union (IOU). Initially, the image is divided into various boxes or grids. The object that appears with the grids is detected efficiently. The objects detected are then highlighted by bounding boxes for the outline. YOLO uses single bounding box regression to predict attributes like width, height, center, etc. Intersection over Union (IOU) is a method with the help of which it makes it easy to perfectly select bounding boxes that surround well to the objects. It eliminates bounding boxes that are not equal to the actual box.

Drawbacks of YOLO:

For small object, there are some detection problems. These problem arises in YOLO because it only predicts one type of class in each gird separately [4]

D. SSD

Single shot detector has a good balance between accuracy and speed. It is faster and accurate than YOLO. The main difference between training SSD and training other is that it uses region proposals and pooling before a final classifier. The advantage is that SSD has the highest mAP among the models targeted for real-time processing.

Comparison results:

For depth feature-based object detection method, the major approaches used are,

- Faster R-CNN

- YOLO
- SSD

After these three models are trained on dataset, the comparison between SSD, Faster R-CNN, and YOLO on Pascal VOC 2007 test set shows SSD (300×300) outperforms YOLO in detection speed, and it is close to Faster R-CNN in accuracy. Accordingly, the SSD (300 × 300) is employed to extract the location information of the goods.[1]

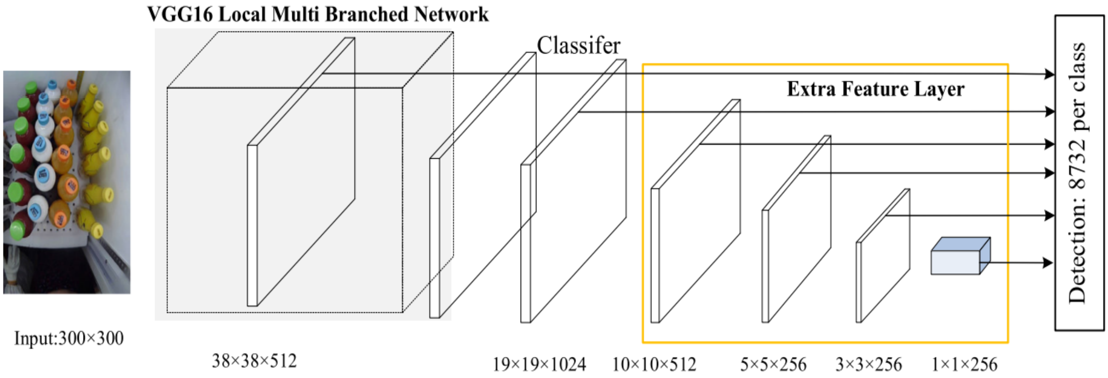

FIG 5: THE ARCHITECTURE OF SSD[5]

E. Improve SSD

SSD is mainly used to extract feature layers in basic networks and add a convolutional layer which effectively reduces the size. At every point in feature layer it is associated with an anchor of variable size corresponding to the input image. The approach used in [6] can be used to detect items in vending machine. In it , it is not able to distinguish between items that have the same shape, for this body features are used, they are model adjustments and add color feature information with special weight. In fig. 5 for detection of product the SSD(300 x 300) the network is illustrated.VGG16 local multi-branch network serves as the fundamental network, and in the yellow box is the feature extraction layer added. At each point of the feature layers it corresponds to different size of anchors. The prediction in SSD will be carried out on the feature maps selected in the previous steps besides object detection on the final feature map selected[1]. As the main difference between the original and proposed SSD model is the sub prediction structure comparison experiments are carried out between both models. They use the same training data set and inference data set. Fig. 6 illustrates the detection precision of these two models for 20 type of items.

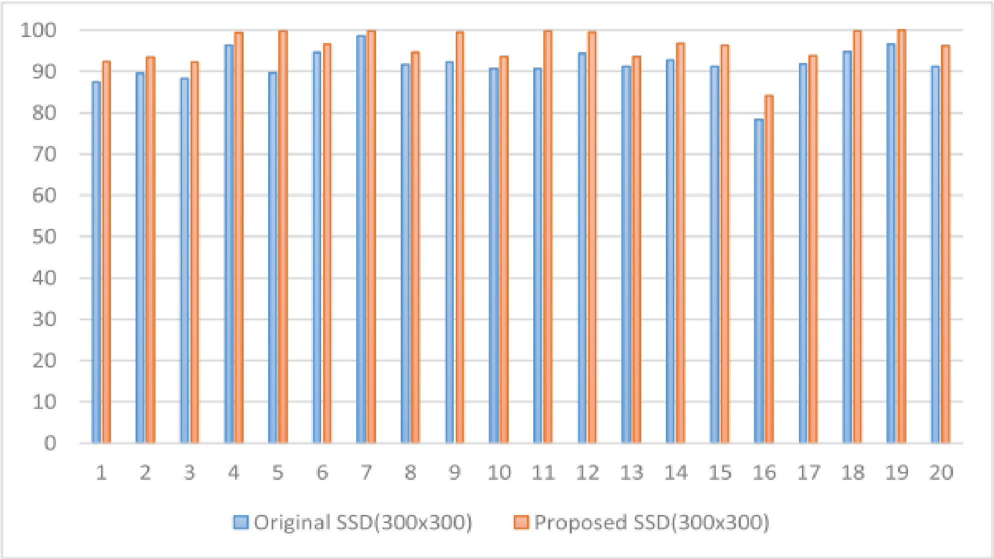

Fig 6: Illustrates of detection precision of these two models for 20 type of items [1].

Table I mean average precision and detection time between the original SSD and the proposed SSD. [1]

Models	Original SSD (300×300)	**Proposed SSD (300×300)**
mAP %	91.64	**96.1**
Time (ms)	21	**22**

The graph clearly shows the results that the proposed SSD has a better precision detection than the original SSD in fig. 6. In the proposed SSD model, convolutional processing is added to the sub-prediction structure, such addition increases the detection time by 1ms ,table 1.

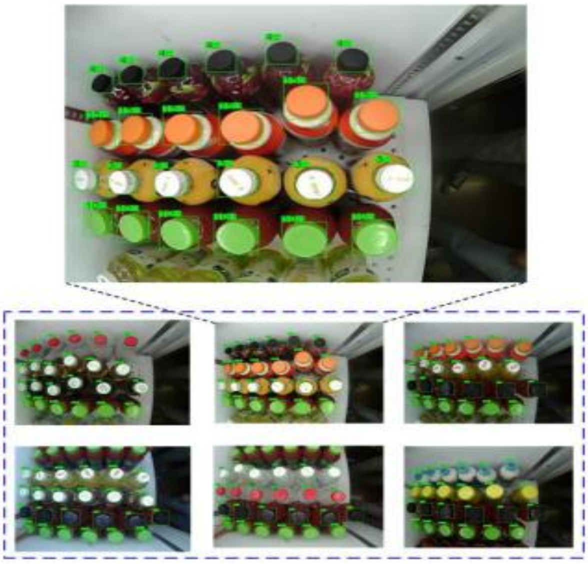

Fig 7: Good example of object detection[1]

IV. CONCLUSION

In this paper, the smart unstaffed retail shop scheme based on artificial intelligence and the internet of things is reviewed. With the use of the new technology in smart shops, it is observed that the shopping experience of the customer and the transaction volume has comparatively increased. after comparison of various object detection models, the conclusion is brought upon SSD(300 x 300) model. This model is further enhanced with a sub-prediction structure that is designed to analyze commodities efficiently conclusion.

V. REFERENCES

[1] J. Xu et al., "Design of Smart Unstaffed Retail Shop Based on IoT and Artificial Intelligence," in IEEE Access, vol. 8, pp. 147728-147737, 2020, doi: 10.1109/ACCESS.2020.3014047.

[2] R. Girshick, “Fast R-CNN,” in Proc. ICCV, 2015, pp. 1440–1448.

[3] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards realtime object detection with region proposal networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 779–788.

[5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “SSD: Single shot multibox detector,” in Proc. Eur. Conf. Comput. Vis., 2016, pp. 21–37.

[6] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2017, pp. 2961–2969.