JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Survey Paper on Retail Store Object Detection

Madhavi R.P.
(Associate Professor)
Computer Science and Engineering
B.M.S. College of Engineering
Bangalore, India

Alok Kumar Rastogi
(Student)
Computer Science and Engineering
B.M.S. College of Engineering
Bangalore, India

Rahul Khandelwal
(Student)
Computer Science and Engineering
B.M.S. College of Engineering
Bangalore, India

Varsha Suresh
(Student)
Computer Science and Engineering
B.M.S. College of Engineering
Bangalore, India

Vishal
(Student)
Computer Science and Engineering
B.M.S. College of Engineering
Bangalore, India

***Abstract*—Object detection, when applied to shelf images can solve many problems in retail stores such as observing the number of products on the shelves and comparing the organized products with the planogram continuously. However, resolving this issue is not simple. A computer, unlike the human eye, analyses pictures in two dimensions. Furthermore, the object's size, orientation, and placement in the picture might all be somewhat different. This study aims to simplify the detection of products in shelves by eliminating manual effort with assistance of deep learning algorithms.**

***Keywords*—object detection, planogram, machine learning, deep learning, SSD, YOLO, R-CNN, mAP%, FPS**

I. Introduction

Object detection is utilized in the domain of computer vision which aims to locate instances of objects within images or videos. It utilizes technologies such as machine learning or deep learning to output relevant results. It focuses on locating the object of interest within the image and then it attempts to determine the category of that object and draws a bounding box enclosing the object.

Computer Vision and Artificial Intelligence have an immense impact on the automation of processes in several industries including retail.

In retail stores, manual labor is required to continuously keep track of the available product stocks, to ensure that all the products are properly placed in their respective slots, and also to ensure that the arrangement of products complies with the planogram.

In real-life situations, there will be various products placed on the shelves and it would be a tedious task to manually keep track of everything. Due to this the need for the development of automated product recognition algorithms in retail stores is of utmost importance.

II. Literature Survey

Jun Deng et al, in [1], did analysis on object detection in the domain of deep learning and concluded that some of the most fundamental problems in computer vision, object detection have been explored in depth in recent years. Detection algorithms founded on deep learning have been broadly used in many fields. Yet, there is an abundance of problems that still need a resolution:

1) Dependence on data needs to be reduced.

2) Efficient detection of small objects needs to be achieved.

3) Detection of Multi-category objects needs to be improved.

They also did a detailed comparison of numerous object detection algorithms (Table 1) based on their mAP% and FPS when applied to various kinds of datasets. According to their observation, YOLOv2 resulted in the highest mAP% whereas YOLOv4 resulted in the highest FPS. Comparatively, SSD resulted in a good balance between mAP% and FPS.

Syed Sahil Abbas Zaidi et al, in [2], said that the best object detection models available now, are still nowhere near the saturation level when it comes to performance. The applications are increasing day by day and therefore there is a strong need of light weight models which can be deployed on phones. In this paper, They have demonstrated through tables how various detection algorithms have evolved. Even though Two-stage detectors have higher accuracy, they can't be used for real-time applications because of their slower speed. However, there has been a shift in the past few years because the one-shot detectors are becoming more and more accurate while maintaining their quick inference speeds. They are reaching the same accuracy levels of two-stage detectors. As depicted in Fig. 1, a transformer based detector i.e. Swin Transformer, at the time of publication of this paper, was the most accurate detector. If the current trend of improvement in accuracy of detector continues, eventually, there will be more accurate and faster detectors.

Jinling Li et al, in [3], introduced an improved SSD multi-object detection model that was ameliorated with respect to detection rate and efficiency. Moreover, the calculation cost and related hardware cost of the model were diminished. Its contributions can be largely mirrored in the aspects mentioned below:
1) In traditional SSD, the corelation between the predicted object category score and object positioning accuracy is poor. The improved model was observed to enhance the correlation between the object score and the positioning accuracy by adding the IoU prediction loss branch, in an attempt to improve the detection accuracy of the model.

2) To reduce the spatial redundancy and to improve the real-time performance, the convolution co-relation module based on multi-frequency feature components was implemented which reduced the calculation cost and related hardware cost of the traditional model.
3) AdaMod optimizer was introduced which altered the adaptive learning rate of the abnormal value. This resulted in improved real-time performance and also incremented the convergence speed of the model.
4) A large number of experiments were conducted and the ideal settings of the hyper-parameter ρ in the IoU detection branch and the hyper-parameter β in the multi-frequency convolution operation were prospected. Various datasets like MS-COCO and PASCAL VOC2012 were used for verification and it was observed that the improved model has comparatively better performance in different datasets.

Lu Tan et al, in [4], collected pill images and used LabelImg to make a standard PASCAL VOC format image database. Three currently dominant object detection methods, Faster R-CNN, YOLO v3 and SSD, were trained using their pill database and their performance was compared experimentally. The results show that each of the three models has its advantages and disadvantages. The Faster R-CNN model has a high MAP (87.69%), but the detection speed (FPS : 7) is not fast enough for real-time application. SSD is intermediate in performance, with scores between the other two networks on both speed (FPS : 32) and MAP (82.41%). Although YOLO v3 did not have the highest MAP (80.17%), according to their observations it can greatly improve the detection speed and achieve real-time performance (FPS : 51). They observed that in busy hospital pharmacies, pill identification requires not only a high enough MAP, but also detection speed. YOLOv3 may be the best compromise to suit their use case. This method can help pharmacists to quickly identify drugs, reduce the probability of dispensing the wrong drug, and can help improve patient safety. They concluded that YOLO v3 algorithm can meet the conditions of operating on low-performance platforms, in environments with requirements for high speed of detection, and has broad development prospects and practical application value.

Guangxing Han et al, in [5], proposed a novel meta-learning-based few-shot object detection model in this paper. Their model consists of the following two modules to tackle the low quality of proposals for few-shot classes. First, a lightweight coarse-grained prototype matching network was proposed to generate proposals for few-shot classes in an efficient and effective manner. Then, a fine-grained prototype matching network with attentive feature alignment was proposed to address the spatial misalignment between the noisy proposals and few-shot classes. Experiments done by them on multiple FSOD benchmarks demonstrate the effectiveness of their approach.

Zhong-Qiu Zhao et al, in [6], culminated that object detection based on deep learning has been a major field of interest in the past few years due to their ability to learn and the advantages they provide in dealing with occlusion and scale transformation. This paper dispenses a through review of object detection based on deep learning frameworks. These object detection models are obtained by modifying the already existing R-CNN models and they primarily focus on sub-problems such as occlusion, clutter and low resolution. The review first explains the basic object detection models and provides a brief understanding of their base architectures. Then, it focuses on three common applications which are salient object detection, face detection and pedestrian detection. Finally, they provided numerous insights into the future directions in which the object detection scenario can be taken to.

Tausif Diwan et al, in [7], classified object detection into two categories: single stage object detection and two stages object detection. They explored a number of two stage object detectors like RCNN, Fast-RCNN, and Faster-RCNN. They also explored some of their important applications. Their major focus was on single-stage object detection models specifically YOLOs. They dispensed various aspects and optimizations in the models showing how YOLO has evolved and also explained their underlying concepts. Moreover, they thoroughly explained the challenges and their motivations for specifically focusing on reviewing the single-stage object detectors. As per their observations, the YOLOs were performing significantly better than other object detection algorithms in terms of detection speed and inference time.

Wen-Sheng Wu et al, in [8], made some improvements to the YOLOv3 network and applied them to inventory cup detection and counting. The detection mAP reached 96.82% with 54.88 FPS.
The first improvement they made was setting a detection area. They only focused on goods or products in specific areas due to which they were able to ignore the products not under active consideration. At first, they trained another YOLOv3 network to find the shelf and obtain the coordinates of the detection area. They manually set the coordinates as user-defined parameters instead of detecting the shelf, to save time and computing resources. One of the advantages is that, when the shelf is changed or we are only interested in one of the layers of the shelf, the coordinates can be changed easily without retraining a network to obtain the coordinates.
The second improvement they made was by eliminating the feature maps and the structures behind them that make no contribution to the final detection results. Through this method, detection FPS increased from 48.15 to 54.88, while the mAP increased from 95.65% to 96.65%.
The last improvement they made was to refine the anchor size. After performing a clustering algorithm on their own dataset and resetting the anchor size, the mAP increased from 96.65% to 96.82%.

They concluded that the performance of YOLOv4 is better than YOLOv3. However, the model of YOLOv4 is more complex and has more parameters. Considering the limited computing resources of industrial computers and that YOLOv3 is more mature at that time, they chose YOLOv3 as the base of our research and achieved satiable results.

Ajay Talele et al, in [9], proposed a new method for obstacle detection using a single webcam. They also presented a new method of vision-based surveillance robot with obstacle avoidance capabilities for general purposes in indoor and outdoor environments. The algorithms which make use of neural networks for obstacle detection and framing the image have been used. The camera which they used had a very poor resolution because it uses only 5MP. For future work, they plan on improving this system, to act as a surveillance system in robotic as well as future applications.

Parag Tirpude et al, in [10], proposed a real time object detection model using OpenCV and Python. Their model is a web based application that mainly aims to detect multiple objects from various types of images. To achieve this goal, the shape and edge features from images are extracted. The proposed model uses a large image database for correct object detection and recognition. This system also provides an easy to understand user interface to retrieve the desired images. The system has additional features one of which is Sketch based detection. In Sketch detection, user can sketch by hand as an input. Finally, the system results output images by searching those images that the user wants.

III. Problem Identification

In retail stores, specifically dairy stores, if an item is out of stock, or if it is in stock but has been placed on the wrong shelf, it can give the customer the impression that the product isn't available and the customer might leave and purchase that same product from another retailer. This causes a direct loss in revenue. Sometimes it also happens that due to some human mistake, the arrangement of products isn't in compliance with the planogram and this may disrupt the organizational strategy crafted by the store to boost sales. To overcome this issue, the store can appoint various employees to keep track of each and every product on the shelves but this in turn makes the store cluttered. Also, if the store is financially unstable, paying so many employees is not feasible.

IV. Proposed Solution

In order to optimize the efficiency, accuracy, and security of a dairy shop, it is proposed that a new Object Detection System be implemented using the latest Image Processing Algorithms. This system will be designed specifically for use in a dairy shop environment, incorporating advanced features and capabilities to meet the unique requirements of the retail industry.

The proposed system will be based on the latest computer vision algorithms and deep learning techniques providing highly accurate and efficient object detection. It will be able to identify and locate objects of interest, such as dairy products, within an image, providing information on the location, size, and orientation of each product.

The system will also be designed to be easy to use and highly customizable, allowing retailers to tailor the system to meet their specific needs. For example, the system can be trained to detect specific objects, such as dairy products, and to ignore others, such as customers. This customization will ensure that the system is optimized for the specific requirements of the store.

In conclusion, the proposed Object Detection System for a dairy shop will provide retailers with a powerful tool for automating tasks such as inventory management, customer service, and security. By incorporating the latest in Image Processing Algorithms and deep learning techniques, the system will be able to deliver highly accurate and efficient results, improving the efficiency and profitability of the store.

The Deep Learning algorithms under consideration are SSD, YOLOv7, and RCNN.

V. Figures and Tables

Method	Backbone	Size/Pixel	Test	mAP/%	fps
YOLOv1	VGG16	448×448	VOC 2007	66.4	45
SSD	VGG16	300×300	VOC 2007	77.2	46
YOLOv2	Darknet-19	544×544	VOC 2007	78.6	40
YOLOv3	Darknet-53	608×608	MS COCO	33	51
YOLOv4	CSP Darknet-53	608×608	MS COCO	43.5	65.7
R-CNN	VGG16	1000×600	VOC2007	66	0.5
SPP-Net	ZF-5	1000×600	VOC2007	54.2	-
Fast R-CNN	VGG16	1000×600	VOC2007	70.0	7
Faster R-CNN	ResNet-101	1000×600	VOC2007	76.4	5

Table 1. Comparison of Object Detection Algorithms

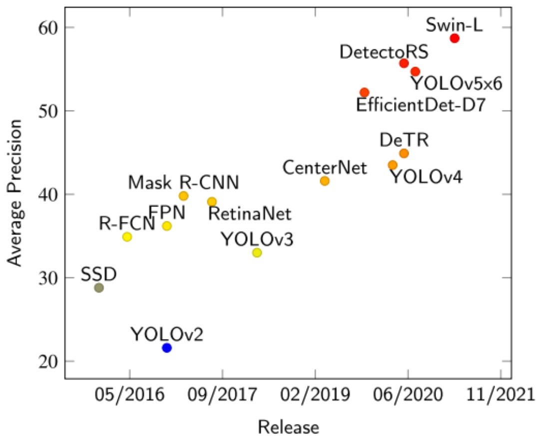

Figure 1. Performance of object detectors on MS COCO dataset

VI. Conclusion

Object detection through image processing algorithms is a powerful tool for automating the task of object recognition and localization. In a retail environment, such as a dairy shop, this technology has the potential to greatly improve the efficiency and accuracy of inventory management, customer service, and security.

By utilizing the applications of image processing techniques to extract meaningful information from an image, the Object Detection subsystem can be trained to precisely detect and locate objects of interest within the store. This can be used to automate the process of inventory management, reducing the time and effort required to manually track the quantities and locations of products. It can also be used to assist customers in finding the products they are looking for, improving their shopping experience and increasing customer satisfaction.

For a dairy shop, the ability to accurately identify and locate products within the store can help to optimize the shopping experience for customers, while also reducing the time and effort required to manage the store's inventory. Additionally, the system can also provide an extra layer of security, helping to prevent theft and other criminal activities.

In addition, By detecting and locating suspicious or unauthorized objects, the system can alert store employees to potential security threats and help to prevent theft and other criminal activities. The implementation of the Object

Detection Subsystem can be customized to meet the specific needs of a retail store. For example, the system can be trained to detect specific objects, such as dairy products, and to ignore others, such as customers. This customization can ensure that the system is optimized for the specific requirements of the store.

Overall, the use of Object Detection using Image Processing algorithms in a retail store, such as a dairy shop, has the potential to greatly improve the efficiency, accuracy, and security of the store's operations. As technology continues to advance and become more widely adopted, it is likely to become an increasingly valuable tool for retailers of all types. It can greatly benefit retailers, by improving efficiency, accuracy, and security. As technology continues to evolve and become more widely adopted, it will become an increasingly valuable tool for retailers of all types.

REFERENCES

[1] Jun Deng, Xiaojing Xuan, Weifeng Wang, Zhao Li, Hanwen Yao and Zhiqiang Wang. “A review of research on object detection based on deep learning”

[2] Syed Sahil Abbas Zaidi, Mohammad Samar Ansari, Asra Aslam, Nadia Kanwal, Mamoona Asghar, Brian Lee. “A survey of modern deep learning based object detection models”.

[3] JINLING LI , QINGSHAN HOU, JINSHENG XING, JIANGUO JU. “SSD Object Detection Model Based onMulti-frequency Feature Theory”

[4] Lu Tan, Tianran Huangfu, Liyao Wu, Wenying Chen. “Comparison of YOLO v3, Faster R-CNN, and SSD for Real-Time Pill Identication”.

[5] Guangxing Han, Shiyuan Huang, Jiawei Ma, Yicheng He, Shih-Fu Chang, “Meta Faster R-CNN: Towards Accurate Few-Shot Object Detection with Attentive Feature Alignment”

[6] Zhong-Qiu Zhao, Member, IEEE, Peng Zheng, Shou-tao Xu, and Xindong Wu, Fellow, IEEE, "Object Detection with Deep Learning: A Review"

[7] Tausif Diwan1 & G. Anirudh2 & Jitendra V. Tembhurne, "Object detection using YOLO: challenges, architectural successors, datasets and applications"

[8] Wen-Sheng Wu and Zhe-Ming Lu, "A Real-Time Cup-Detection Method Based on YOLOv3 for Inventory Management"

[9] Ajay Talele, Aseem Patil, Bhushan Barse, "Detection of Real Time Objects Using TensorFlow and OpenCV"

[10] Parag Tirpude, Pooja Girhepunje, Sagar Sahu, Sweta Zilpe, Prof. Harshita Ragite, "REAL TIME OBJECT DETECTION USING OPENCV-PYTHON".