Stock Management System Using Object Detection Techniques

Mamta Borle, Janvi Patel,Riddhi Nalawade,Smit Mehta

Abstract— Lack of availability of goods and/or the improper positioning of products on the shelves of a retail store can result in loss of sales to a retailer. Visual audits are undertaken by the retailer's staff and the staff of the FMCG product companies, (whose products are stocked in the retail shelves), to discover out-of-stock and misplaced products in a retailer's shelf. In this paper, a method of automating the process of manual inspection has been described. The solution involves real-time monitoring of the shelf using a camera. The paper also demonstrates that by applying image processing techniques it is possible to identify and count the front-facing products, as well as detect void spaces on the shelf. Images from a video stream (such as from a security camera) can also be analyzed to count the number of facings of a specific product on a shelf and identify if they are placed face-up, as should be the case. The image processing approach proposed in the paper will primarily enable proper positioning of products on the shelf in the front row. While that may seem as a limitation for inventory counting, it is actually an important parameter for product manufacturers who usually rent shelf space and positions at a premium and mandate the retailers to place specific products at specific shelves. The an incremental change that the paper proposes is to extend the use of feature extraction in image processing to highlight incorrect placement and positioning of items on the shelves. The implemented solution does not require significant additional infrastructure costs, and is easy to implement/use while being reasonably accurate.

Index Terms— Object detection, yolo, Deep learning

I. INTRODUCTION

To understand and manipulate digital videos Computer Vision (CV) is used. Face recognition, industrial inspection, image retrieval and augmented reality are playing important role. With the help of deep learning CV is useful for many applications. To manage retail products and improve users' experience product recognition is used. Barcode recognition is used not only in research but also in industries. The management of products can be easily facilitated by scanning barcode marks on each product package. Every item is having its respective barcode. The barcode should be placed properly so that it should be visible for machine to identify. Many customers said that many times as the barcode is not visible so it is not useful. To enhance the automation of product for the growth of computer technology radio frequency identification (RFID) is applied in business fields .

Retail industries are impacted by automatic product recognition implementation in grocery stores using images. Product recognition can be used to improve Planogram compliance of products. Missing items can be identified using automatic production from the shelf to remind the store staff to provide the products immediately. In relation with image classification and object detection problems, A n arduous instance is retail product recognition problems. For image classification and object detection the core solution is deep learning, especially in the domain of computer vision.

II. PRIOR WORK

Computer Vision is growing exponentially along with technology.For the continuous growth and improvements in the domain of Computer Vision a lot of work has happened. To witness the growth and improvements in this area many researchers follow different methods and approaches to a problem. Always researchers will be kept on digging to find improvements. This section explains about the previous works which have been done so far using different methodologies followed. Firstly, we start with different applications of object detection and then we get to the background of the implementing algorithm.

A. *Classic Methods*

Researchers have been drawn to product recognition using the technology with computer vision's rapid growth. By extracting features on the image of the package product recognition is realized. The composition of the product image recognition system is shown in

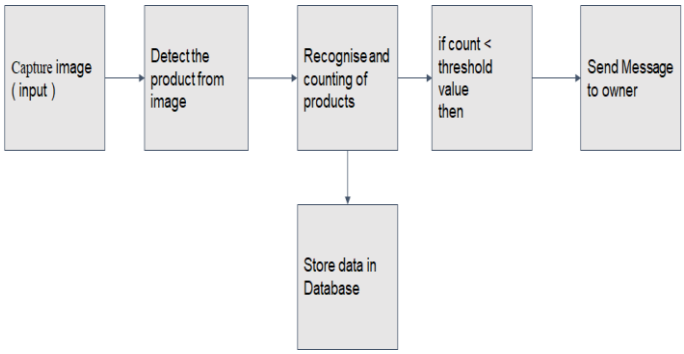
Fig (1) Getting various images from mobiles and cameras (Image capture). (2) To provide high quality images for subsequent operations noise and redundant information can be removed. It mainly includes image transformation, segmentation, and enhancement (Image preprocessing). (3) To determine the invariant characteristics in the image the analysis and processing of image data is done (Feature extraction). (4) After the feature vector or space is mapped with a certain image feature, a specific decision rule is applied to classify the low-dimensional feature to make the recognition result accurate (Feature classification). (5)The pre-trained classifier is employed to predict the category of the retail (The output of recognition).

B. *Deep Learning*

The sub field of machine learning is deep learning. To learn deep representations is the vital objective of deep learning it means to learn multilevel representation and abstraction from information.The authoritative scholars in the field of machine learning in 2006 proposed the concept of deep learning (also known as deep structured learning), the methods of unsupervised pretraining and fine-tuning are presented by Hinton and Salakhutdinov to solve the vanishing gradient problem. After that year, deep learning became a research hotspot.To optimize the initial weights for deep networks a greedy layer-wise training strategy was provided in 2007.To preserve more information among multiple layers which could restrain the vanishing gradient problem ReLU (rectified linear unit) was defined in 2011.To prevent overfitting, and it helped improve the deep network performance the dropout algorithm was proposed.

Convolutional Neural Networks

The biology research of the cat's visual cortex gives the inspiration for the success of deep learning in computer vision profits from convolutional neural networks (CNNs). In 1998 to classify the images LeCun et al. first proposed to employ convolutional neural networks. They conceived seven layered the LeNet convolutional neural network model. This model had been successfully applied to the digital identification of checks after training on a dataset which contained handwritten characters. Implementation



The real-time video, captured by the surveillance camera can be considered as a sequence of N frames and it can be denoted as V = (f1, f2, ..., fN). Each frame in the video can be treated as a colour image in RGB format, where RGB represents red, green and blue colour components. The frames in a frequent interval will be transferred to the next phase.

The next phase in the proposed scheme is scene change detection. The key idea is that the further steps like object detection, counting the products in the rack, and automated emailing or SMS sending process can be carried out only after empty space detected. Since the camera is working all the time and we are considering the frames in a frequent interval (every frame in an interval of 30 seconds), whenever the product is taking out by costumers then a scene change will be detected. No need to do all the phases in every frame. The scene change detection helps to optimize the overhead of object detection and recognition tasks.

Threatening object images are used to create image data set ,which is taken from google images. We have taken nearly 78 object images out of which the images splitted into train and test images. Using trained images we can train and evaluate model using the test performances. Firstly, we start from labelling the images.We can

use labeling tool for images.The rectangular box is created around the objects which gives co-ordinates of the object where it lies. The image data is initially stored in xml format for each image. As the number of images is more we will have a same number of xml files to avoid complexity. So, we create a csv file which has data of all the images.

After finishing training, we access the api from our side by giving the test images to detect the objects in a,n image. These images are led to the tensor flow serving server. TensorFlow Serving is a versatile, high- performance serving system intended to produce environments different machine learning models. TensorFlow Serving makes deploying fresh techniques or algorithms and experiments straight forward while maintaining same server design and APIs. The input image we have given are returned with Bounding boxes around the object.

In this phase, the number of products in the selected rack will be counted and if it goes beyond the threshold limit an SMS or an email will be triggered to inform the supervisor. The product identifier or product name also will be included in the SMS or email that will help the supervisor to plan the refilling process.

III. *ANALYSIS OF RESULTS*

A. *Results of Product Recognition and Counting of Product*

Turn on the camera equipment, and the surrounding environment is detected in real time. The system intercepts the real-time shooting contents into each frame for detection,Taking the detection at two different moments as an example, the results are shown in Figure 2. The system can identify the shooting objects as Cornflakes and Chocos.

Fig2: product Detection and Counting

B. *Recognise Empty space*

Once the system identifies that the number of products in
the product rack is less than it will send an SMS and/or email to the respective supervisor.

The Analysis is carried on a limited number of products and the results observed during the experimental Study We have considered five products and assumed very few products will be kept in the rack.We assumed the threshold for sending warning message is (halfof the maximum number of products that we can keep in a rack).

IV. CONCLUSION

A computer vision-based approach for automated moni-
toring of the products in the supermarket is introduced in
this paper. The proposed framework can be adopted in busy
supermarkets. The implementation of the proposed scheme in supermarkets will help the managing companies to reduce the required manpower, increase the profit and provides better customer satisfaction. In the proposed scheme. In this work, we have used one camera to keep track of the products in one single rack, but in a real case, a single camera may be used to cover a large area which contains racks for different products.

This issue can be taken into consideration for improving the scheme introduced in this chapter.

REFERENCES

[1]Shraddha Mane, Prof.Supriya Mangle," Moving object detection and tracking Using Convolutional Neural Networks",2018International Conference on Intelligent Computing and Control Systems(ICICCS)

[2]B N Krishna Sai,Sasikala T,"Object Detectionand Count of Objects in Image using TensorFlow Object Detection API", 2019 Second International Conference on Smart Systems and Inventive Technology(ICCSIT)

[3] Yanmei Liu 1, Ben Liu1, Yuda Chen2," Research on Image Recognition of Supermarket Commodity Based on Convolutional Neural Network", 2019 International Symposium on Computational Intelligence and Design(ISCID)

[4] Sandeep Kumar Yedla,V. M. Manikandan, Panchami V." Real Time Scene Change detection With Object Detection For Automated Stock Verification", 2020 International Conference on Device, Circuits, Systems(ICDCS)

[5]Nishant Kejriwal, Sourav Gang, Swagat Kumar,"Product Counting Using images With the Application to Robot-Based Retail Stock Assessment"

[6]Ceren Gulra Melek,Elena Battini Sönmez, Songul Albayrak"Object Detection in Shelf Images with YOLO"

[7] Eran Goldman1, Roei Herzig, Aviv Eisenschtat, Oria Ratzon, Itsik Levi, Jacob Goldberger, Tal Hassner,"Precise Detection in Densely Packed Scenes"

[8] Ramiz Yilmazer, Derya Birant,"Shelf Auditing Based on Image Classification Using Semi-Supervised Deep Learning to Increase On-Shelf Availability in Grocery Stores"

[9] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: a survey and new perspectives," *ACM Computing Surveys (CSUR)*, vol. 52, p. 5, 2019.View at: Publisher Site | Google Scholar

[10] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.View at: Publisher Site | Google Scholar

[11] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.View at: Publisher Site | Google Scholar

[12] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, USA, 2007.View at: Google Scholar

[13] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, Haifa, Israel, 2010.View at: Google Scholar

[14]G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, https://arxiv.org/abs/1207.0580.View at: Google Scholar