2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT)

Efficient Shelf Monitoring System using Faster-RCNN

1st Divyam Arora
Dept. of ECE
BNM Institute of Technology
Bangalore, India
aroradivyam007@gmail.com

2nd Keerti Kulkarni
Dept. of ECE
BNM Institute of Technology
Bangalore, India
keertikulkarni@bnmit.in

2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT) | 979-8-3503-8295-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICDCOT61034.2024.10515839

***Abstract*—In the fast paced world of today where marketing and advertisement is at its peak, an effective inventory management is essential for success. Remote Inventory management is traditionally done through Image processing methods in which the images of the shelves are taken and then processed to find the missing objects. This process creates a new set of problems relating to customer privacy. While a majority of the works concentrate on finding the missing objects, this article aims to detect the absence of objects or empty shelves, without disclosing the identity of the customers. Along with this the location of the empty shelf is also pointed out. This work uses the pre trained model Faster RCNN to detect empty regions of shelves of a retail store. Results of the simulation show an accuracy of 99% thereby indicating the use of this method for industrial deployment.**

Keywords*—Object Detection, Shelf Monitoring System, Inventory Management, Inventory Monitoring, RCNN, Faster R-CNN, Shelf Availability.

I. Introduction

In the continuous changing and evolving field of retail, the demand for an efficient system for the management of stock or inventory is escalating daily. Traditional retail stores face significant labor costs to monitor shelf inventory regularly, often postponing this operation until off-peak hours. A delay in inventory monitoring causes high sales losses when a particular item is gone from the shelf though additional stock existed in the warehouse. An ordinary convenience store faces out-of-shelf stockout rates of 5–10%, which results in a loss of up to 4% of sales retail stores. The prime challenges faced here are efficient restocking of items which further involves accurate real time detection of stock levels. This real time detection can provide us with valuable insights like popularity of a particular product, efficiency of staff in restocking the items, and may also provide us analytics regarding the gross sales of a particular store. Data shows that when a product is not available on the designed shelf space, 31% of consumers buy the product from a different store, 26% of them buy a different brand, 19% of them buy a different size of the same brand, 15% of them buy the same product at a later time, and 9% of them buy nothing [1]. There are two ways in a broader view to detect the stocks, one is to detect the amount of products present and the other is the simpler method of detecting the empty shelves. Both the techniques traditionally involve Image Processing algorithms. Though these algorithms are simple and easy to execute they need real time images to give satisfactory results. The images have to be captured either directly (which is again time consuming and costly) or taken through a CCTV footage. Hence the shift towards Machine learning algorithms, where once the model is trained real time data is no longer needed to detect the missing objects or empty shelves. While numerous methods/algorithms are implemented (detailed analysis is given in section 2), there are various drawbacks of each of them. Some of the drawbacks are insufficient data (images from all categories of the retail market scene), lesser efficiency and more time consuming. Taking all these factors into account the proposed work has the following objectives. 1. Design an efficient and cost-effective method for detecting empty shelves in the retail market scene. 2. Efficiency should be achieved without compromising the privacy of the customers. The system uses pre trained Faster RCNN model which is well know for its object detection capability. This approach combines the traditional CNN model with the ResNet-50 architecture and hence it is appropriate for binary classification tasks like detecting empty shelves. The use of RCNN has proven to be the best approach because of its balance between the accuracy and lesser computation than other models, reason of its accuracy is directly related to its unique approach of region proposals and object classification. This dynamic approach is truly a game changer and will elevate the shopping experience to new heights, this approach of inventory management is not bounded to just retail stores but can extend to all the domains that involve inventory management irrespective of size of the business. The major contributions of this works are 1. Design of a Faster RCNN algorithm for detecting empty shelves which gives an accuracy of 99%. 2. This work also provide us with valuable insights like time taken to restocking of items and staff management. This is achieved as the location of the empty shelf is also indicated which in turn helps in reducing staff strength and thus being a cost-effective solution. This paper is organized as follows. Section II contains the background of similar works or the Literature Survey. section III explains the methodology used, section IV explains the results, section V discusses the results obtained followed by conclusions in section VI.

II. Literature Survey

Out of stock (OOS) is one of the major reasons why retail markets lose business. There are various methods available in literature which deal with on-shelf availability (OSA). The simplest method uses the RFID tagging to monitor the quantity of the products on the shelf [2]. The disadvantage of such methods is that they are not scalable to too many shelves and are not cost effective [3]. Alternately, no hardware would be required if this problem is handled in the image processing domain using specialized softwares [4]. A further extension of this work uses deep leaning methods to detect the presence or absence of objects [5]. Even though the image processing techniques based on histograms are quite popular they are limited in their performance. The use of DL algorithms is effective when the images are annotated effectively. Even though the databases are often found, there is a major problem of the annotation of the images in the database. The authors [6] tackled this problem by using the deep learning architecture based on YOLO algorithm. Research has shown that under such circumstances a semi-supervised approach is more effective and comparable to all the other approaches [7]. Some other authors have come-up with a template matching approach to detect the presence or absence of products on a shelf [8]. Robots mounted with cameras can also be used for the counting of the products on the shelf [9, 15]. This approach again, lacks scalability and is not cost effective. Images obtained from the surveillance cameras can be used to detect the presence of empty shelves as shown by the authors in [10, 16]. Yet another method utilizes the fact that when a customer is standing in front of a shelf there is a high probability that he/she has bought the product. Hence analyzing the image frames after the customer has left the scene provides an insight to the state of the shelf [11,14]. Some other methods for shelf detection deal with use of planograms which are sent to the store manages in advance [12] and the use of spectral graph matching technique [13]. While all the surveyed literature has employed a variety of algorithms, all of them consider efficiency, privacy and cost-effective parameters separately. This is a major research gap as identified by the authors.

III. Proposed Methodology

The methodology adopted for the proposed work is shown in Figure 1, which is followed by a detailed explanation of all the blocks.

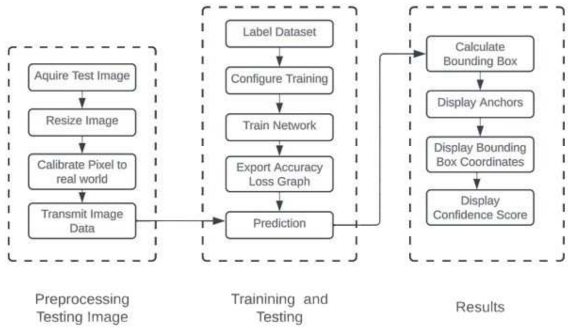


Fig. 1. Workflow for the proposed methodology

A. Collection and Annotations of Image Dataset

The dataset required for the work is downloaded from github repository (https://github.com/sayan0506/Product-Matching-with-Shelf-images-using-Yolov4-following-product-recognition-using-Resnet50v2?tab=readme-ov-file). Size of each image in the dataset varied, hence it is resized to 300 x 300 for further processing. The dataset chosen should include the images of all types of products available in a retail store. A sample of the images is shown in Figure 2, Figure 3 and Figure 4.

The dataset used in this project consist of approximately 1000 images of shelves of a retail grocery store, the images were captured under well-lit condition and are of different time stamps. Images also involve a variety of products on the shelves with different categories of shelves increasing the diversity of dataset.

Fig. 2. Dataset Image-1

Fig. 3. Dataset Image-2

Fig. 4. Dataset Image-3

The next major step is the annotation of these images. The annotation of images is carried out in MATLAB's inbuilt annotation tool The Image Labeler App as it suits the need for square shaped annotation boxes with a simple annotation of images under the label 'Empty'.

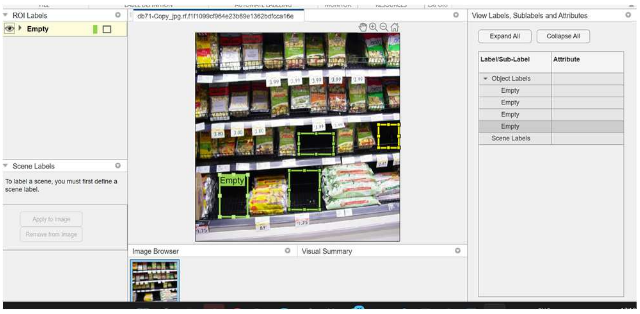
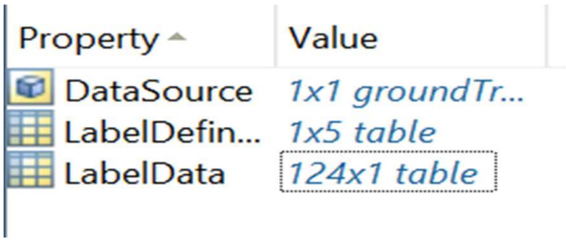


Fig. 5. Image Annotation

Figure 5 shows the process of annotation in Image Labeler App. As shown in the figure, a ROI is labelled as empty and the anchor boxes chosen are rectangle green in color. The images in the dataset are annotated in serially. After all the annotations are completed, MATLAB has inbuilt function to export the annotated images in form of ground truth labels or the table can be directly imported to the workspace. The ground truth table consists of information about the Label data, Label definitions (it has information about x- y coordinates of anchor boxes and source (it has path to input images) as shown in Figure 6.



Property	Value
DataSource	1x1 groundTr...
LabelDefin...	1x5 table
LabelData	124x1 table

Fig. 6. Annotated data imported to the workspace

B. *Preprocessing of Images*

The pre processing performed in on this dataset is to set a constant resolution value of 300x300 pixels to the input dataset, this step is to ensure there are no diversities in the input images provided to detector and it also marks a standard format to reduce further complexities while training the model, this step also ensures restoring of essential features of the image with better convergence during the training of the model.

C. *Anchor Box Prediction*

An inbuilt MATLAB function called 'estimateAnchorBoxes()' is used here. The function takes input as the pre-processed training data and estimated number of anchor boxes and evaluates an estimate value of both number and size of anchor boxes the model can encounter while testing , this step ensures that the model can test object of various sizes and improves adaptability of the code and accurate detection of empty shelves.

D. *Model and Architecture Selection*

1) Faster R-CNN

A generalized view of the RCNN model used is shown in Figure 7.

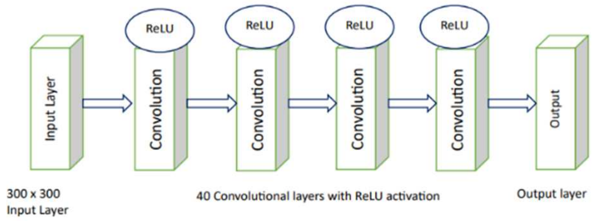


Fig. 7. RCNN model used

Convolutional Neural Networks have definitely restructured the traditional methods of object detection which faced some significant issues in object detection especially in situations involving multiple objects to be detected, the following problem has been solved with Region proposal approach of R-CNN changing the view of object detection [4]. The network that generates region proposals is known as region proposal network (RPN) in faster R-CNN. RPN produces region proposals by predicting whether the anchor boxes contain an object or not. R-CNN functions in two prime stages: proposal of most probable region to find bounding boxes and second is to classify objects, the outputs from the first stage are transferred to the second stage and thus computing the final output, this new approach ensures accurate object detection in terms of location and efficiency, the initial approaches suffered from slow training as region proposal was carried out separately.

Using R-CNN to detect empty shelves-:

- Regional approach effectively reduces the image in areas with most probability to detect empty shelves
- RCNN excels at precise positioning of the objects, which turns out to be very beneficial in detecting empty shelves accurately.
- The potential derivatives of RCNN are capable of adapting wide range of variations such as scale and aspect ratio of the input data.
- The 'estimateAnchorBox' function as discussed above is also plays an essential role in accurate detection of empty spaces.
- The R-CNN efficiently uses the input images and is able to discriminate them on the basis of their respective annotations and predicting the label.
- The usage of pre trained model also aids in the work as the model is already trained for a large dataset, this makes the job easy to train on a comparatively smaller dataset subjective to the needs.
- R-CNN's architecture is proven to produce results which can be interpreted easily.

2) *ResNet-50:* ResNet-50 stands for Residual Network with 50 layers, is shown in Figure 8. This particular architecture truly stands as a benchmark for other architectures, ResNet uses a special technique of residual connections which is in simple words are direct paths or shortcuts for the information to flow, making the learning of the model a lot effective and less time consuming, this

method of learning also reduces the loss of important details to a very large extent [5]. This model simplifies network training while addressing the issue of network degradation. Other architectures like VGG are known for its simple approach and uniform architecture and GoogLeNet which uses inception modules and involves pipelining in multiple filter sizes making both VGG and GoogLeNet a computationally intensive approach, whereas MobileNet which is designed to work in constrained environments is less intensive as it uses limited resources like memory.

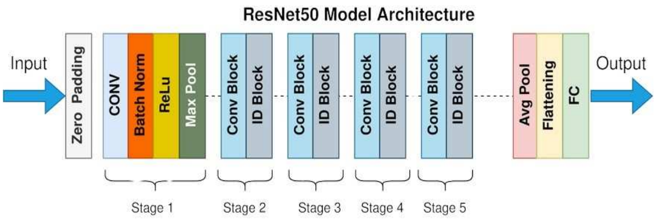


Fig. 8. ResNet-50 model

On comparing all the above three architectures with ResNet-50 it is concluded that ResNet-50 turns out to be the best approach for a task like empty shelf detection as it has a robust balance between accuracy and efficiency in computations which avoids the trade off between two as mentioned in above architectures. Talking about the activation layer used is ‘activation_40_relu’ to extract vital features of the images and also keeping in account the global and local context of the operation., the other layers were not chosen because they didn’t align with the task of classification and this particular layer lands a sweet spot computing both vital overall and finer little details required for the task of empty shelf detection.

E. *Training of Model and Optimization*

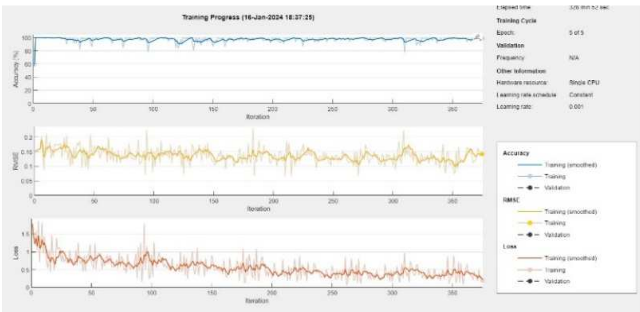


Fig. 9. Model training

The model training is shown in Figure 9. The training and optimization of the model involves various parameters that affect the extent of learning of the model and thus also decides the accuracy and performance of the final model, in the following work, ‘sgdm’ which stands for Stochastic Gradient Descent with Momentum is used as an optimizer which updates the model’s parameters in each iteration based on the gradient loss function allowing convergence for the optimal solution, other parameters involved are number of epochs which refers to the amount of times the entire dataset is passed through the training dataset ,mini-batch size which refers to the number of samples taken in every iteration which further optimizes memory usage and efficiency of computations ,also improves the speed and ability of the training. The above figure shows the three plots namely Accuracy vs Iteration, Loss vs Iteration, RMSE vs Iteration which helps us in the visualization and monitoring the performance of the model.

IV. Results

The simulation was run in MATLAB and the results are shown in the following figures. Not only is it important to detect the empty shelf it is also important to know the location of the empty shelf.

A. *Displaying anchor boxes and their respective coordinate*

Fig. 10. Empty Shelves detected

Fig. 11. No empty shelves detected

The above Figure 10 clearly shows the detection of regions of empty shelves in the respective images. To test the integrity of the model it is tested with an image containing no empty shelves and resulting Figure 11 clearly shows the desired output with no empty shelves being detected. The predicted anchor boxes are well aligned with the empty spaces detected by the model.

```
Empty Shelf 1: [x, y, width, height] = [269, 409, 120, 106], Score: 1.00
Empty Shelf 2: [x, y, width, height] = [170, 413, 52, 92], Score: 0.99
Empty Shelf 3: [x, y, width, height] = [352, 100, 84, 96], Score: 0.99
Empty Shelf 4: [x, y, width, height] = [532, 252, 111, 112], Score: 1.00
Empty Shelf 5: [x, y, width, height] = [232, 156, 60, 49], Score: 0.58
```

Fig. 12. Accuracy of the model

The above Figure 12 represents the coordinates of anchor boxes of Figure 10 with the confidence score of each anchor box detected ,the confidence score tell us about the degree upto which the model is confident about a generated output.

B. *Visualization of anchor boxes*

An alternate method for the visualization of the results is shown in Figure 13.

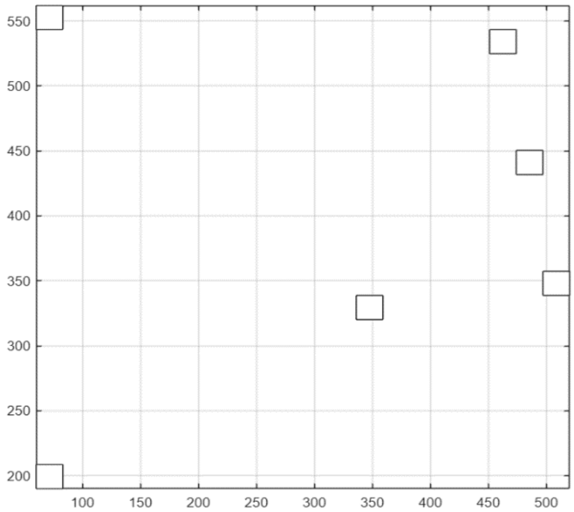


Fig. 13. Location of the empty shelf

The Figure 13 is a 2D histogram representation of anchor boxes of figures 10 based on their respective coordinates this method helps in clear visualization of distribution of empty shelves all over the shelf.

With this method the overall accuracy is found to be 99%. The time taken to train the model is approximately 6 hours, whereas the prediction is achieved as fast as 24 seconds. A comparison of the results with the other works is shown in Table 1.

TABLE I. RESULTS COMPARISON

Reference	Algorithm	Accuracy
[6]	YOLO	86.22%
[10]	CNN	89.6
Proposed Method	Faster RCNN	99%

V. DISCUSSION

As it is evident by the results, the implemented model shows promising results in detecting empty shelves accurately both respective to the empty region size and localization, the model is tested on various kinds of images containing diversity of products as well as on diversity of shelf types, each of the results of detected empty shelves have the confidence score of almost a hundred percent proving how well the model is trained to adapt all kind of conditions involving shelves with lot of empty spaces ,shelves with no empty spaces or even shelves with moderate amount of empty spaces the results were as accurate as it can get ,talking about the post analysis of the data , 2D histograms of the anchor boxes detected has been implemented, this helps us in the following ways:

- Clear visualization of distribution of regions of empty shelves.
- This analysis can provide us insights about the sale of a particular product.
- For instance if more of the empty shelves are seen accumulated at a single portion of shelf that tells us lot about the psychology of people,effects of advertisment of multiple products on them.

Histograms are comparatively easier to interpret and can provide valuable insights by giving a standard output.

VI. CONCLUSION

The main objective of this work is to employ an efficient and cost-effective method for empty shelf detection in the retail market scene without compromising the privacy of the customers. This is achieved in this work by implementing a faster RCNN model on a dataset which encompasses majorly all the products in the retail scene. This dataset also does not have any customers on the scene. The achieved effieiciency is 99% which is an indication that this method can be deployed in actual retail markets for empty shelf detection. As a future work, this work can be extended to other types of shops which can include the same or a better algorithm.

ACKNOWLEDGMENT

The authors would like to thank B N M Institute of Technology and VTU for providing a platform to conduct this research work.

REFERENCES

[1] Higa K, Iwamoto K. “Robust Shelf Monitoring Using Supervised Learning for Improving On-Shelf Availability in Retail Stores”. Sensors (Basel). 2019 Jun 17;19(12):2722. doi: 10.3390/s19122722. PMID: 31213015; PMCID: PMC6631981.

[2] Bottani, E.; Bertolini, M.; Rizzi, A.; Romagnoli, G. Monitoring on-shelf availability, out-of-stock and product freshness through RFID in the fresh food supply chain. Int. J. RF Technol. Res. Appl. 2017, 8, 33–55.

[3] Michael, K.; McCathie, L. The Pros and Cons of RFID in Supply Chain Management. In Proceedings of the 4th Annual 4th International Conference on Mobile Business, ICMB, Sydney, NSW, Australia, 11–13 July 2015; pp. 623–629

[4] Moorthy, R.; Behera, S.; Verma, S.; Bhargave, S.; Ramanathan, P. Applying image processing for detecting on-shelf availability and product positioning in retail stores. In Proceedings of the ACM International Conference Proceeding Ser., Kochi, India, 10–13 August 2015; pp. 451–457.

[5] Higa, K.; Iwamoto, K. Robust estimation of product amount on store shelves from a surveillance camera for improving on-shelf availability. In Proceedings of the IST 2018—IEEE International Conference Imaging Systems and Techniques Proceeding, Kraków, Poland, 16–18 October 2018; pp. 1–6.

[6] Ramiz Yilmazer and Derya Birant, “Shelf Auditing Based on Image Classification”,Sensors2021,21(2),327;https://doi.org/10.3390/s21027

[7] Zhu, X. Semi-Supervised Learning, Encyclopedia of Machine Learning and Data Mining; Springer: Berlin/Heidelberg, Germany, 2017; Volume 3, ISBN 9781489976871.

[8] Satapathy, R.; Prahlad, S.; Kaulgud, V. Smart Shelfie-Internet of shelves: For higher on-shelf availability. In Proceedings of the 2015 IEEE Region 10 Symposium TENSYMP, Ahmedabad, India, 13–15 May 2015; pp. 70–73.

[9] Kejriwal, N.; Garg, S.; Kumar, S. Product counting using images with application to robot-based retail stock assessment. In Proceedings of the IEEE Conference on Technologies for Practical Robot Applications, Woburn, MA, USA, 11–12 May 2015; pp. 1–6.

[10] Higa, K.; Iwamoto, K. Robust estimation of product amount on store shelves from a surveillance camera for improving on-shelf availability. In Proceedings of the IST 2018—IEEE International Conference Imaging Systems and Techniques Proceeding, Kraków, Poland, 16–18 October 2018; pp. 1–6.

[11] Higa, K.; Iwamoto, K. Robust shelf monitoring using supervised learning for improving on-shelf availability in retail stores. Sensors 2019, 19, 2722.

[12] Liu, S.; Tian, H. Planogram Compliance Checking Using Recurring Patterns. In Proceedings of the 2015 IEEE International Symposium on Multimedia (ISM), Miami, FL, USA, 14–16 December 2015; pp. 27–32.

[13] Liu, S.; Li, W.; Davis, S.; Ritz, C.; Tian, H. Planogram compliance checking based on detection of recurring patterns. IEEE Multimed. 2016, 23, 54–63.

[14] Falcão J, Ruiz C, Pan S, Noh HY and Zhang P (2020) FAIM: Vision and Weight Sensing Fusion Framework for Autonomous Inventory Monitoring in Convenience Stores. Front. Built Environ. 6:568372. doi: 10.3389/fbuil.2020.568372

[15] Mahdi, F.P., Motoki, K. & Kobashi, S. Optimization technique combined with deep learning method for teeth recognition in dental panoramic radiographs. Sci Rep 10, 19261 (2020). https://doi.org/10.1038/s41598-020-75887-9

[16] Zhang, Luying, Yuchen Bian, Peng Jiang, and Fengyun Zhang. 2023. "A Transfer Residual Neural Network Based on ResNet-50 for Detection of Steel Surface Defects" Applied Sciences 13, no. 9: 5260. https://doi.org/10.3390/app13095260