SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I7P118

Volume 11 Issue 7, 180-187, July 2024

Original Article

Vision-Based Empty Shelf Detection in Retail with Real-Time Telegram Notifications for Efficient Restocking

Shital Pawar[1], D.B. Jadhav[2], Deepali Godse[3], Rohini Jadhav[4], Shruti Thakur[5]

[1,5]*Department of Computer Engineering, Bharati Vidyapeeth's College of Engineering for Women, Maharashtra, India.*
[2]*Department of Mechanical Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Maharashtra, India.*
[3]*Department of Information Technology, Bharati Vidyapeeth's College of Engineering for Women, Maharashtra, India.*
[4]*Department of Information Technology, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Maharashtra, India.*

[1]*Corresponding Author : shital.pawar@bharatividyapeeth.edu*

Received: 13 May 2024 Revised: 11 June 2024 Accepted: 11 July 2024 Published: 27 July 2024

Abstract *- Empty spaces and fewer items on shelves of stores and big marts often dissatisfy the customer by making the items unavailable when needed. Empty spaces and fewer items on shelves of stores and big marts often disappoint customers by making items unavailable when needed. This also reflects the commitment of store staff to their work. As a result, there is a decrease in sales and a breakdown of trust between sellers and customers. Object detection is used to identify empty spaces and shelves with fewer items. Commonly used algorithms for object detection include CNN, YOLO, and SSD. Large, freely available standard datasets such as Pascal (plate number 1) and Pascal (plate number 2) are utilized, each containing around 20 classes for shelf item detection. Items are labelled as 'Out of Stock' along with their names. This labelling helps visually represent the items. Object detection often requires GPUs and a webcam. The system has developed a dataset containing four classes of grocery items. The labels for the items have been derived from their respective images, with annotations stored in separate image files. The system has been trained using the YOLOv5 algorithm. The output, consisting of images showing empty shelves or low item counts, has been connected to the Telegram API to notify store staff to restock as needed, streamlining the restocking process. This versatile application can be used for inventory management, research, and development and can also be integrated with commercial retail stores, utilizing CCTV cameras for monitoring.*

Keywords *- Computer vision, Deep Convolutional Neural Network, Machine Learning, Object detection, YOLO algorithm.*

1. Introduction

Efficient inventory management is essential for maximizing sales and customer satisfaction in the fast-changing retail industry. Empty shelves can lead to lost sales, dissatisfied customers, and a negative impact on the store's reputation. Keeping track of shelf inventory using traditional methods is labor-intensive and usually not effective in preventing stockouts [1-4]. Advancements in computer vision and machine learning have recently provided new opportunities for streamlining inventory management. However, many existing solutions are costly and require extensive infrastructure changes, limiting their adoption [5-7].

Computer Vision and Artificial Intelligence have an immense impact on the automation of processes in several industries, including retail. Despite these impacts, existing solutions often face several challenges. Many systems still find a need to build their dataset to deliver good results as they use standard available datasets (T. Ahmad et al.,2020) [5]. Due to limitations in computational resources, achieving high accuracy with the current system is not possible (Fuchs et al., 2019) [1]. In response to these challenges, this research proposes an approach for empty shelf detection using You Only Look Once version 5 (YOLO V5), an advanced object detection model known for its speed and accuracy [8, 9].

Unlike previous systems, this solution leverages readily available webcam technology and integrates with Telegram for immediate restocking alerts, providing a cost-effective and easily implementable solution. By addressing the limitations of high cost and poor adaptability, this system offers significant improvement over existing methods [9].

However, there are certain limitations to the system. The performance of a system may be affected by varying lighting

conditions, shelf arrangements, and occlusions. Achieving real-time processing capabilities while balancing resources limits the system's usability. Additionally, the system is not designed to detect empty shelves or count the number of objects. Thus, it is limited to a few objects [10-12]. Future improvements include adding support for more object classes, efficiently detecting a wider range of objects, improving the system's ability to process large datasets, enhancing its object counting capability, and integrating efficiently with the Telegram API to send images with detected objects along with messages [13, 14]. The objectives and scope of this study are:

- To develop an efficient vision-based system for
- detecting empty shelves in retail environments.
- To integrate the detection system with Telegram to send real-time notifications to store personnel.
- To evaluate the accuracy and reliability of the detection system in various retail scenarios.
- To assess the effectiveness of the real-time notification system in improving restocking efficiency.

The key contributions of the authors are as follows:

- Development of a Vision-Based Empty Shelf Detection System: Implemented a deep learning model, You Only Look Once version 5, to detect empty shelves in retail environments accurately. A comprehensive dataset of retail shelf images has been created to train and test the model.
- Integration with Real-Time Notification System:
- Developed a mechanism to send real-time notifications via Telegram when an empty shelf is detected. Ensured notifications are prompt, enabling immediate action for restocking.
- Improvement in Inventory Management: Automated the process of shelf monitoring, reducing reliance on manual checks and Enhancing the efficiency and accuracy of inventory management in retail stores.

The paper structure provides acomprehensive overview. The Introduction section outlines the context of the study, defines the problem statement, and states the research objectives, and contributions. The Literature Survey examines existing solutions and identifies gaps in current research. The Proposed System details the data collection process, model training and testing, and the complete process. The results and discussion section presentsthe performance metrics and gives visual representations. The Conclusion, at last, recaps the main discoveries and proposes potential paths for future studies.

2. Literature Survey

After reviewing multiple papers, the authors noted that no dataset was created from scratch; rather, pre-existing datasets were used, which affected the results. Additionally, the version of the You Only Look Once model was often found to be unsuitable for the retail detection system. Real-time testing on shelves was not conducted. This allowed the authors to expand the scope of the system. A comprehensive analysis of the reviewed papers is provided below.

Fuchs et al. [1] have published a paper that explores the utilization of convolutional neural network architectures in conjunction with an open-source dataset. This dataset comprises 300 images of vending machines containing 15,000 labeled instances of 90 distinct products.

The main contribution of the study lies in the open-sourcing of this labeled image dataset, which serves to advance research in the field of computer vision for product identification. Moving forward, the authors suggest that higher accuracy in product identification can be achieved, particularly with improved content specification and readily available data. This opens avenues for future research to enhance the precision and efficiency of computer vision systems in identifying packaged products, thereby benefiting various applications within retail and beyond.

Caiet. al. [2] have used the Classification Confidence Network (CLCNet) inspired by R-CNN on a Locount dataset. The researchers collected a significant amount of data from 28 different stores and apartments, resulting in a big object localization and counting dataset known as the Classification Confidence Network, or The Locount Dataset. Their work primarily focuses on the Classification Confidence Network method, which employs a coarse-to-fine multistage process to classify and regress bounding boxes, ultimately predicting the instance numbers contained within these boxes. The findings highlight the efficacy of the Classification Confidence Net-work method, as it achieves remarkable results with an APls score of 43.5% on the Locount dataset, surpassing all other methods examined. The authors suggest that further improvements in accuracy are possible by increasing the number of iterations.

Diwan et al. [3] have published their study. The paper offers insights into the utilization of You Only Look Once and its comparison with other deep learning techniques/models commonly employed in object detection. These techniques include deep neural networks, Convolutional Neural Networks, RNNs, and their architectural variants, such as LSTM and GRU, GANs, and various types of autoencoders. The findings of the study reveal accuracies of 63.4% and 70% for You Only Look Once and Fast-RCNN, respectively. Also, the inference time is significantly faster, around 300 times, in the case of You Only Look Once. The scope of the paper suggests that advancements in Convolutional Neural Networks should be minutely experimented with for further enhancement in single-stage object detectors, paving the way for continued improvement in object detection methodologies and applications.

Jiang et al. [4] provides a comprehensive overview of advancements in the You Only Look Once algorithm. The work delves into the techniques applied in You Only Look Once algorithm developments, including batch normalization and multi-scale training. Through their findings, the authors highlight the distinctions and commonalities observed between Convolutional Neural Networks and various versions of the You Only Look Once algorithm. The scope of the paper suggests a focus on more in-depth implementations, particularly in comparing scenarios, thus providing insights into applications and performance of different You Only Look Once algorithm iterations across diverse contexts. This emphasis on scenario analysis can further elucidate the strengths and limitations of You Only Look Once-based object detection systems, contributing to a deeper understanding of their efficacy in real-world settings.

3. Proposed System

3.1. System Architecture

The system architecture integrates data processing, image processing, and the You Only Look Once algorithm to analyze store shelves efficiently. It captures and processes images, employs You Only Look Once for object detection, and integrates with the Telegram API for inventory updates and staff notifications. The user interface provides a streamlined experience for staff to monitor real-time data and take necessary actions, ensuring accurate stock management (see Figure 1). The algorithm detects objects by identifying specific-terms, shapes, or features within the image.

3.1.1. Telegram API Integration

The detected empty spaces trigger an automatic update to the inventory management system. Upon identifying empty shelves (with one object), the system generates notifications for store staff. These notifications include details about the location of the space and recommendations for restocking. This ensures that the inventory records reflect the real-time status of each shelf, aiding in accurate stock management.

3.1.2. User Notification Dispatch

The processed information is presented on a user interface, allowing staff to monitor and interpret the data easily. The Notifications include details about the space, such as shelf location and suggested actions.

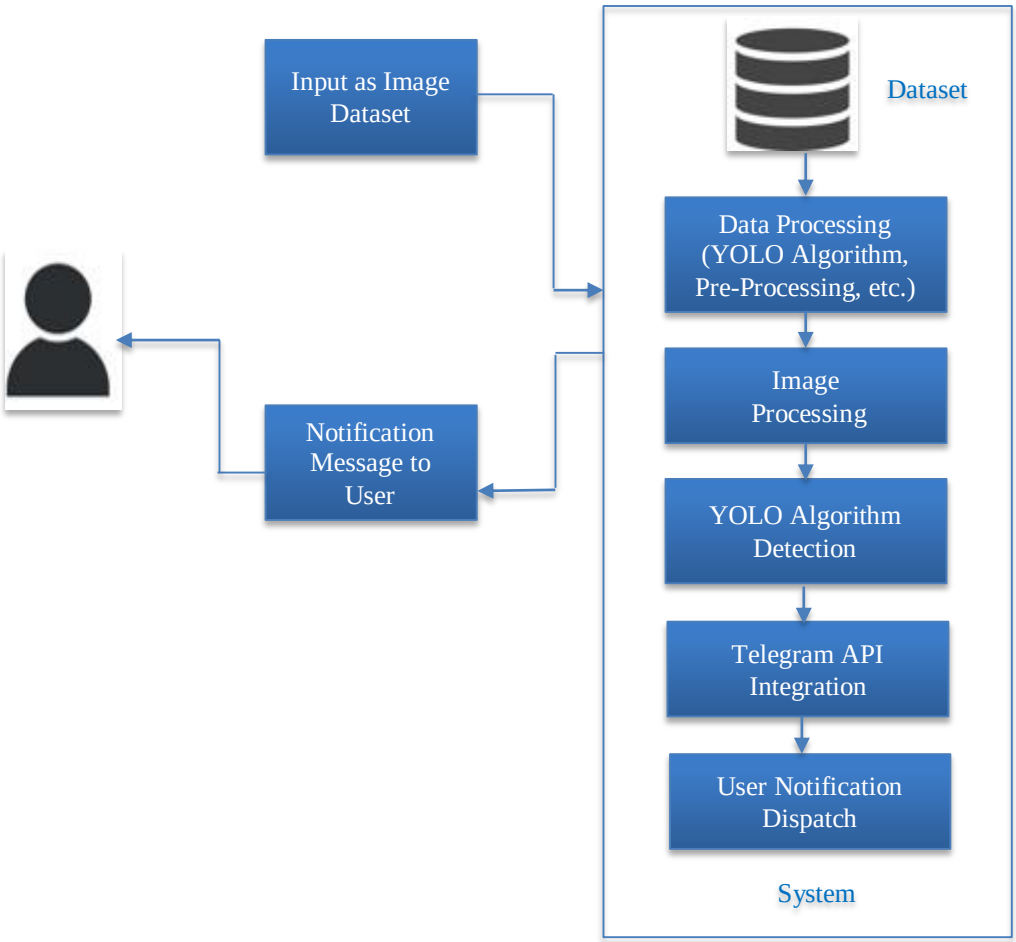

Fig. 1 System architecture

3.1.3. Image Processing

The retail space uses strategically positioned cameras to capture images of store shelves. The raw images undergo pre-processing to improve quality and prepare them for analysis. This includes operations such as resizing, normalization, and noise reduction to ensure consistent input for the next processing steps.

3.1.4. You Only Look Once Algorithm Detection

The You Only Look Once algorithm is employed for object detection. It divides the image into a grid and, for every grid, predicts bounding boxes and associated class probabilities. The algorithm focuses on identifying and localizing objects, particularly empty spaces or out-of-stock items on the store shelves.

3.2. Dataset

The system gathered data from a retail shop consisting of 4,088 images with corresponding labels. The dataset was divided into training and testing sets at a 70:30 ratio, with 2,862 images and labels for training and 1,226 for testing. This arrangement enables the model to learn from a diverse range of training data while providing a separate test set for evaluation. For efficient restocking, only one or two images of each item were taken, and no empty shelves images were included. The images display four different products in various conditions on the shelves, presenting a realistic depiction of retail environments. These detailed images (refer to Figure 2) enhance the model's ability to accurately detect out-of-stock items and improve its generalization capability.

Fig. 2 Samples from dataset

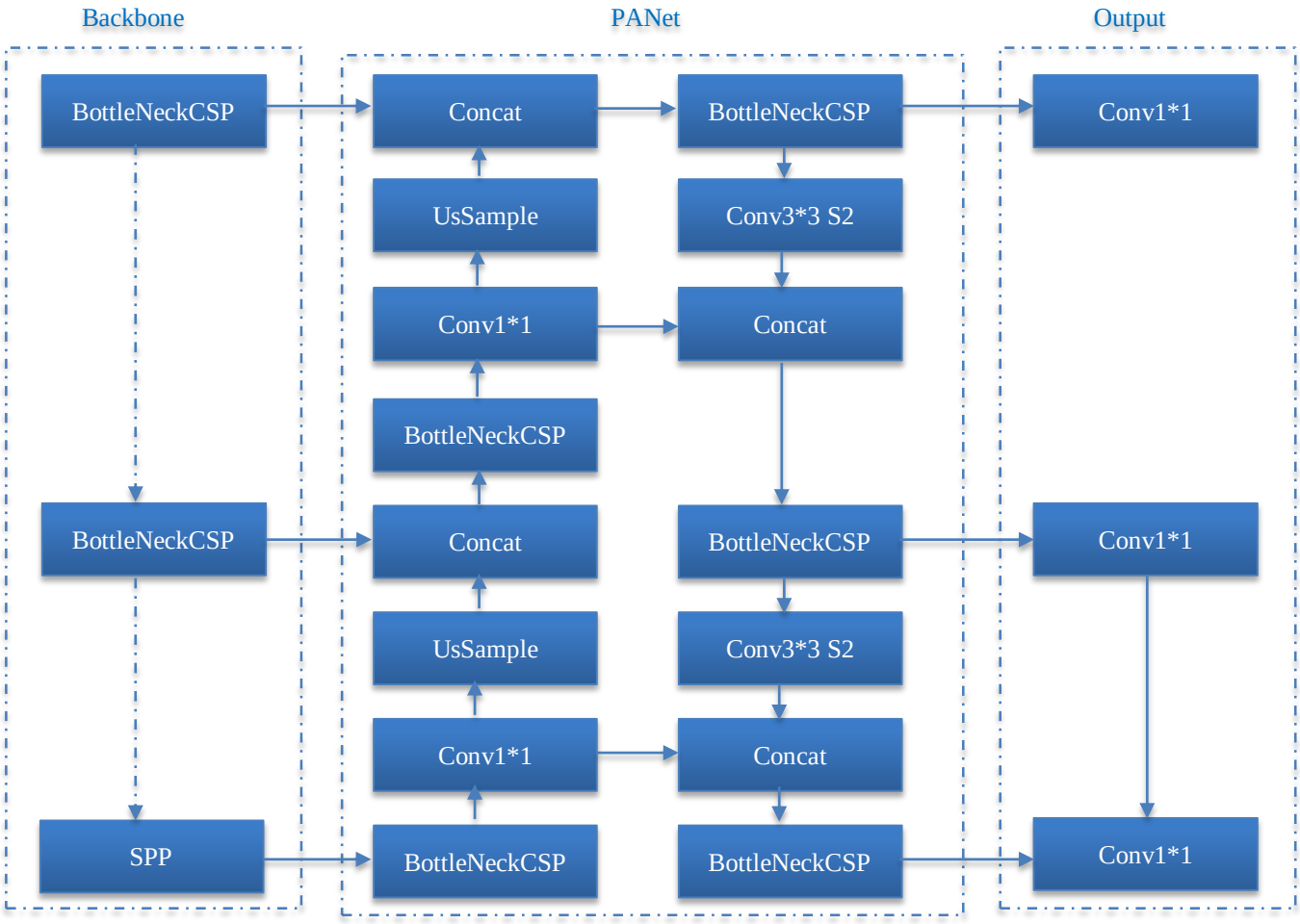

Fig. 3 Overview of you only look once version 5s

3.2.1. Algorithm Used

The algorithm used is known for its speed and accuracy. You Only Look Once version 5 is a version of the You Only Look Once algorithm developed by Ultralytics, which aims to improve upon previous versions with enhancements in both performance and ease of use (see Figure 3). Key features of You Only Look Once Version 5 are:

- Architecture: You Only Look Once version 5 utilizes a deep CNN architecture, similar to its predecessors.
- Model Variants: You Only Look Once version 5 comes in various model sizes, such as You Only Look Once version 5s and much more, with increasing complexity and accuracy. Users can choose a model variant based on speed and accuracy.
- Training Pipeline: You Only Look Once Version 5 provides a simplified training pipeline compared to previous versions, making it easier for users to train their custom object detection models.
- Performance: You Only Look Once version 5 aims to achieve better performance in terms of both speed and accuracy compared to previous versions.
- Framework Integration: You Only Look Once version 5 is implemented using PyTorch, which provides flexibility and usability for developers.

4. Result and Discussion

The confusion matrix is a valuable tool for determining the performance of the model on the test data across the four chosen labels: Dettol Handwash, Ariel, Lakme, and Himalaya Facewash. The matrix gives a visual representation of the model's ability to correctly classify each label, where the diagonal values represent the correctly classified instances for each category. In this case, the matrix shows a diagonal value of 1 for each label and zeros elsewhere, indicating perfect classification accuracy with no instances of misclassification across the four categories. This demonstrates the model's robust performance and ability to distinguish between different product labels with high precision in Figure 4.

The system achieved an accuracy of 96%. The model achieved high metric scores across different classes, demonstrating its robustness in accurately detecting out-of-stock items. TP = True Positive, FP = False Positive, FN = False Negative, TN = True Negative.

- Precision: Precision is calculated as the ratio of true positive detections to the sum of true positive and false positive detections. It measures the accuracy of the model's positive predictions.

  Precision = TP/TP+FP

- Recall: Recall is calculated as the ratio of true positive detections to the sum of true positive and false negative detections. It measures the model's ability to identify all relevant objects.

  Recall = TP/TP+ FN

  mAP50 (Mean Average Precision at IoU 0.5): mAP50 calculates the mean average precision for multiple classes at a fixed Intersection over Union (IoU) threshold of 0.5. This metric represents the model's accuracy across different object classes at a specific IoU level. mAP50 = at IoU 0.5

- mAP50-95 (Mean Average Precision at IoU range 0.5-0.95): mAP50-95 is calculated as the mean average precision across different IoU thresholds ranging from 0.5 to 0.95 in steps of 0.05. This metric provides a comprehensive overview of the model's performance across various IoU levels. mAP50-95 = at IoU range 0.5-0.95

The graph of Model Accuracy illustrates the model's learning progression, with improvements in performance over time. It gives the real values from the training process, which is represented in the given graph in Figure 5. Precision and recall measure the model's accuracy and completeness, while mAP0.5 and mAP0.5-0.95 provide insights into mean average precision at specific IoU thresholds. This graph shows the model's overall robustness and its effectiveness in learning from the data.

The graph of Model Loss shows how the model learns and generalizes over time. The training loss demonstrates the model's improved ability to minimize errors during training. The validation loss indicates that the model generalizes well to unseen data. The convergence of both curves towards zero suggests that the model is effectively learning and achieving low error rates for both training and validation data in Figure 6.

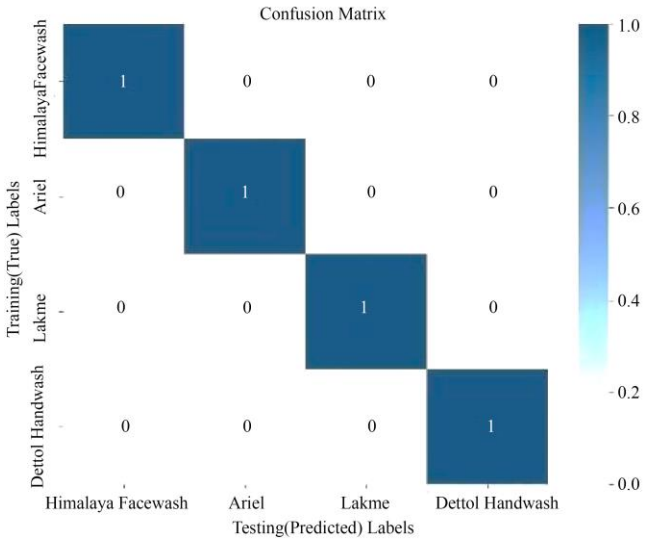

Fig. 4 Confusion matrix

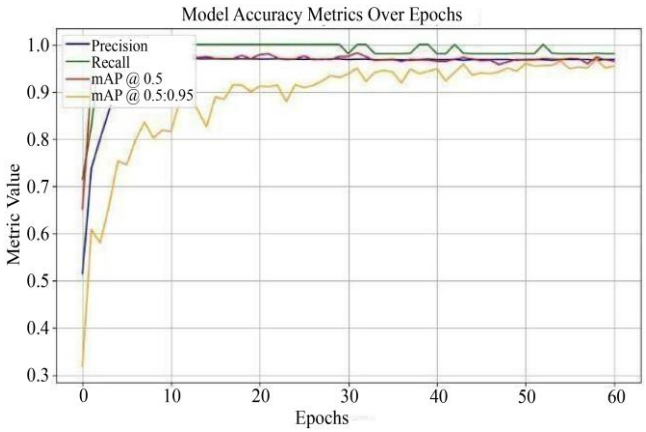

Fig. 5 Model accuracy graph

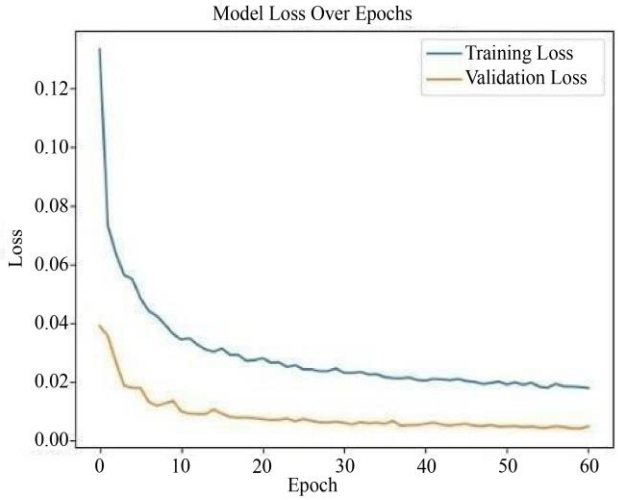

Fig. 6 Model loss graph

Fig. 7 Detected images

The results were achieved by creating a dataset from scratch. Extensive training was provided to the model by the use of GPUs. Using good GPUs (A100, V100) over 100 epochs provided 96% accuracy. Proper detection was ensured. Efficient integration of the model with Telegram was done by creating a bot on Telegram. Thorough testing was done until satisfactory results were obtained.

Figures 7 and 8 represent the output of the system. Figure 7 represents the detected objects with bounding boxes and respective labels. Figure 8 is a screenshot of the Telegram App in which images of detected objects are Ariel and Dettol Hand wash have been sent with a message by the system.

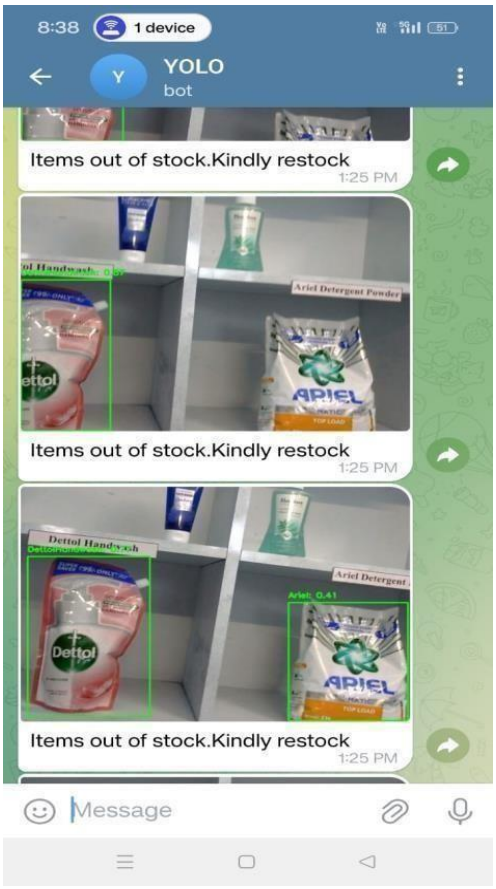

Fig. 8 Detected image with message on telegram

Table 1. Model comparison

Criteria	YOLOv5	YOLOv8
Accuracy	96%	96%

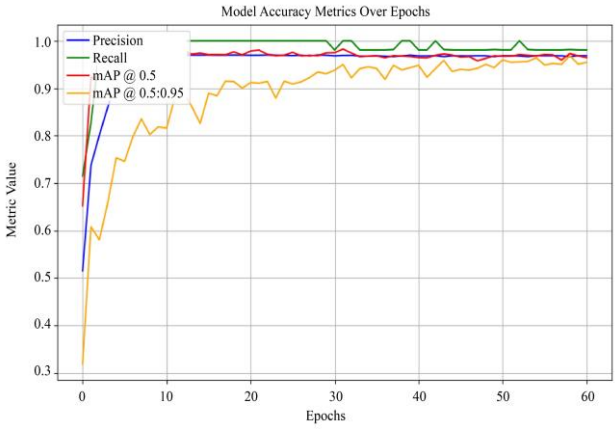

Fig. 9 YOLOv5 accuracy graph

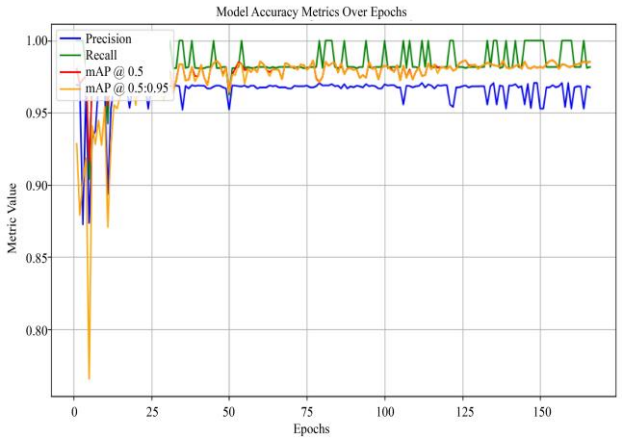

Fig. 10 YOLOv8 accuracy graph

The two graphs in Figures 9 and 10 show the comparison of the YOLOv5 and YOLOv8 models. The models are compared on the basis of the accuracy (mAP50-95) metric. Both have 96% accuracy. The YOLOv8 has a metric range starting from 0.6 while YOLOv5 has from 0.3. This shows that v8 has a good metric range. However, YOLOv5 took 60 epochs, while the other took around 160 due to its complexity.

5. Conclusion

In a dynamic retail landscape, integrating the You Only Look Once version-based object detection system marks a significant leap forward in inventory management and customer satisfaction. This innovation not only streamlines restocking processes but also enhances customer experiences by ensuring products are readily available.

The integration of the You Only Look Once version-based object detection system revolutionizes retail inventory management and enhances customer satisfaction. Leveraging the You Only Look Once version's real-time accuracy, it efficiently identifies empty spaces and low stock, streamlines restocking, and ensures product availability. Its adaptability in crowded environments showcases its practicality. From dataset creation to deployment, the system promises to redefine retail management, with the potential for future expansions and additional functionalities.

In the future, there is potential to improve and expand the system. More advanced versions of the You Only Look Once (YOLO) model can be developed to enhance the system's capabilities. This could involve adding more classes to increase the size of the dataset.

Additionally, there is a need to allocate slightly more computational resources and budget to accommodate these changes. Moreover, the detected objects can be counted more effectively using better techniques, which is currently a limitation.

References

[1] Klaus Fuchs, Tobias Grundmann, and Elgar Fleisch, "Towards Identification of Packaged Products via Computer Vision: Convolutional Neural Networks for Object Detection and Image Classification in Retail Environments," *Proceedings of the 9th International Conference on the Internet of Things*, pp. 1-8, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[2] Yuanqiang Cai et al., "Rethinking Object Detection in Retail Stores," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 2, pp. 947-954, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] Tausif Diwan, G. Anirudh, and Jitendra V. Tembhurne, "Object Detection Using YOLO: Challenges, Architectural Successors, Datasets and Applications," *Multimedia Tools and Applications*, vol. 82, pp. 9243-9275, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Peiyuan Jiang et al., "A Review of Yolo Algorithm Developments," *Procedia Computer Science*, vol. 199, no. 4, pp. 1066-1073, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Tanvir Ahmad et al., "Object Detection through Modified YOLO Neural Network," *Scientific Programming*, vol. 2020, no. 1, pp. 1-10, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] S. Geethapriya, N. Duraimurugan, and S.P. Chokkalingam, "Real-Time Object Detection with Yolo," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 8, no. 3S, pp. 578-581, 2019. [Google Scholar] [Publisher Link]

[7] F. Sultana, A. Sufian, and P. Dutta, "A Review of Object Detection Models Based on Convolutional Neural Network," *Intelligent Computing: Image Processing Based Applications*, pp. 1-16, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Wang Zhiqiang, and Liu Jun, "A Review of Object Detection Based on Convolutional Neural Network," *2017 36th Chinese Control Conference (CCC)*, Dalian, China, pp. 11104-11109, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[9] Zhong-Qiu Zhao et al., "Object Detection with Deep Learning: A Review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212-3232, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[10] Xinrui Zou et al., "A Review of Object Detection Techniques," *2019 International Conference on Smart Grid and Electrical Automation (ICSGEA)*, Xiangtan, China, pp. 251-254, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[11] Rayson Laroca et al., "A Robust Real-Time Automatic License Plate Recognition based on the YOLO Detector," *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, Brazil, pp. 1-10, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[12] Yunong Tian et al., "Apple Detection During Different Growth Stages in Orchards Using the Improved YOLO-V3 Model," *Computers and Electronics in Agriculture*, vol. 157, pp. 417-426, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[13] Yonten Jamtsho, Panomkhawn Riyamongkol, and Rattapoom Waranusast, "Real-Time License Plate Detection for Non-Helmeted Motorcyclist Using YOLO," *ICT Express*, vol. 7, no. 1, pp. 104-109, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[14] Jian Han et al., "Target Fusion Detection of LiDAR and Camera Based on the Improved YOLO Algorithm," *Mathematics*, vol. 6, no. 10, pp. 1-16, 2018. [CrossRef] [Google Scholar] [Publisher Link]