Received: 24 February 2023 | Revised: 30 March 2023 | Accepted: 10 April 2023 | Published online: 13 April 2023

RESEARCH ARTICLE

Artificial Intelligence and Applications
2023, Vol. 1(3) 175–179
DOI: 10.47852/bonviewAIA3202798

BON VIEW PUBLISHING

Object Detection and Tracking for Crate and Bottle Identification in a Bottling Plant Using Deep Learning

Leendert Remmelzwaal[1],*

[1]*University of Cape Town, South Africa*

Abstract: This paper presents a practical implementation of supervised object detection techniques for real-world manufacturing applications, specifically for crate tracking, bottle counting, and bottle inspection in a bottling plant. The proposed model architecture utilizes a two-stage tracking process, employing a wide-angle camera and advanced object detection algorithms to overcome the limitations of traditional convolutional neural networks. The first stage of the model tracks the crates, while the second stage identifies the bottles within the crates. The accuracy of the proposed approach is validated to be over 99.9%. The paper details the dataset preparation, model architecture, training procedure, and evaluation results.

Keywords: deep convolutional neural networks, machine learning, object detection, manufacturing, bottling

1. Introduction

Convolutional neural networks (CNNs) have demonstrated excellent performance in object classification, detection, and segmentation on established image datasets and have recently been applied in various manufacturing applications (Sun et al., 2021). Object classification, object detection, and object segmentation are three fundamental computer vision tasks that play a critical role in various applications such as autonomous driving, robotics, and video surveillance. While all three tasks involve identifying objects within an image, they differ in their level of granularity and complexity.

Object classification is the task of assigning a single label or category to an entire image or a region of interest within an image (Krizhevsky et al., 2012; Simonyan & Zisserman, 2015). This task involves training a classifier to recognize specific visual patterns or features associated with different object categories (He et al., 2016; Szegedy et al., 2017). For instance, given an image of a dog, a classifier trained to recognize dog images would predict the label "dog" for the entire image.

Object detection, on the other hand, involves identifying the location and category of objects within an image (Girshick et al., 2014; Lin et al., 2017; Li et al., 2022; Redmon & Farhadi, 2018; Wang et al., 2018). Unlike object classification, object detection requires localizing the object(s) of interest by drawing bounding boxes around them (Liu et al., 2016, 2017). The goal of object detection is to identify all instances of objects in an image, as well as their precise spatial locations (Bochkovskiy et al., 2020; He et al., 2019). For instance, given an image of a street scene with several cars and pedestrians, an object detector would identify and localize all the cars and pedestrians present in the image.

Object segmentation is the most fine-grained of the three tasks, as it involves identifying the exact pixel-level boundaries of objects within an image (Long et al., 2015; He et al., 2017). Object segmentation requires the algorithm to segment an image into different regions corresponding to different objects and to assign a distinct label or mask to each of these regions (Chen et al., 2018; Long et al., 2015; Ronneberger et al., 2015). For instance, given an image of a person walking on a beach with a dog, an object segmentation algorithm would segment the image into distinct regions corresponding to the person and the dog and assign a unique mask to each of these regions.

Object classification, object detection, and object segmentation all involve identifying objects within an image; they differ in their level of granularity and complexity, with object classification requiring the least amount of granularity and object segmentation requiring the most. Each of these tasks has its own unique set of challenges and techniques, and understanding their differences is crucial for developing effective computer vision systems.

In this paper, we chose to apply supervised object detection to a practical application, namely crate tracking, bottle counting, and bottle inspection in a bottling plant. This decision was made because it offers sufficient granularity in detection, while consuming the least processing power.

Our client specializes in the manufacturing of bottled products and has an extensive global presence with manufacturing plants across different regions of the world. As part of their manufacturing process, the client required a reliable solution that could accurately track the movement of crates and count the number of bottles entering and leaving their manufacturing plant. The purpose of this requirement was to generate accurate reports on raw and processed materials that could help the client to

*Corresponding author: Leendert Remmelzwaal, University of Cape Town, South Africa. Email: leen@firststep.ai

effectively manage their manufacturing operations and make informed decisions regarding their production processes. By implementing an automated system to track and count their products, the client would be able to streamline their operations, reduce errors, and increase efficiency.

Generally speaking, the more information the model provides, the larger the model, the longer the training, and the longer inference takes. We aim to show that our approach achieves high accuracy with minimal computation and can be easily applied in industrial settings.

2. Model Architecture

While CNNs have demonstrated impressive results in object detection, one of their limitations is their inability to effectively detect objects that they have not been trained on Zhang et al. (2016). To address this issue, our approach was to collect a diverse and balanced dataset (He & Garcia, 2009; Lee et al., 2019) that contains a wide range of samples for the objects we aim to classify. To this end, we collected and annotated over 2000 images of crates and 24,000 images of bottles, providing a robust and diverse dataset for training and testing our models.

To further improve the accuracy and effectiveness of our object detection system, we designed a two-stage tracking process (Sun et al., 2021), which consists of two separate object detection models working sequentially (see Figure 1). In recent years, there has been a significant increase in the use of computer vision techniques for object detection in manufacturing processes. Two-stage object detection, which involves first generating region proposals and then classifying these regions, has shown promising results in improving the accuracy and speed of detection systems. In manufacturing, these systems can be used to detect defects, monitor product quality, and optimize production processes. For example, in semiconductor manufacturing, two-stage object detection has been used to identify defects on wafers with high accuracy and efficiency (Chen et al., 2018). Similarly, in the automotive industry, two-stage object detection has been used to detect and classify defects on car bodies during the production process (Chen et al., 2020).

The first model was responsible for tracking the crates, while the second model identifies the bottles within the crates. This approach allows for more granular and precise object detection, as well as providing greater flexibility and adaptability to different real-world scenarios. By combining a diverse dataset with advanced object detection techniques, our system is able to overcome the limitations of traditional CNNs, delivering high levels of accuracy and reliability in object detection tasks.

In the initial phase of our object tracking process, we employed a wide-angle camera to track the crates as they moved along the conveyor belt. The camera was set to capture images at a high frame rate of 14 FPS, which ensured a smooth and reliable tracking performance under challenging conditions. Images were processed in real time by an EDGE processing unit. To further enhance the accuracy and precision of our system, we utilized a state-of-the-art object detection model using the Tensorflow 2.x framework that achieved an impressive mean average precision (mAP) score of 0.93, as measured by IoU@0.05:0.95 (see Figure 2(a)). This was achieved after 247 epochs, and we carefully monitored the training process to ensure that the model achieved optimal performance while minimizing the loss function, which is a crucial aspect of deep learning-based approaches. The best model was chosen from epoch 247 because the system accuracy (mAP score) and model loss did not improve for 100 epochs thereafter. Each model was Float 16 quantized for optimal performance on EDGE devices.

By employing a high-speed camera and advanced object detection algorithms, we were able to achieve highly accurate and robust tracking of the crates, which forms the foundation for our subsequent bottle detection and classification stages.

The second stage of the tracking process was carefully designed to improve the accuracy of bottle detection. The process involved first cropping the crate from the high-resolution frame and utilizing only the crate image for detecting bottles. The decision to

Figure 1
Two-stage object tracking and detection

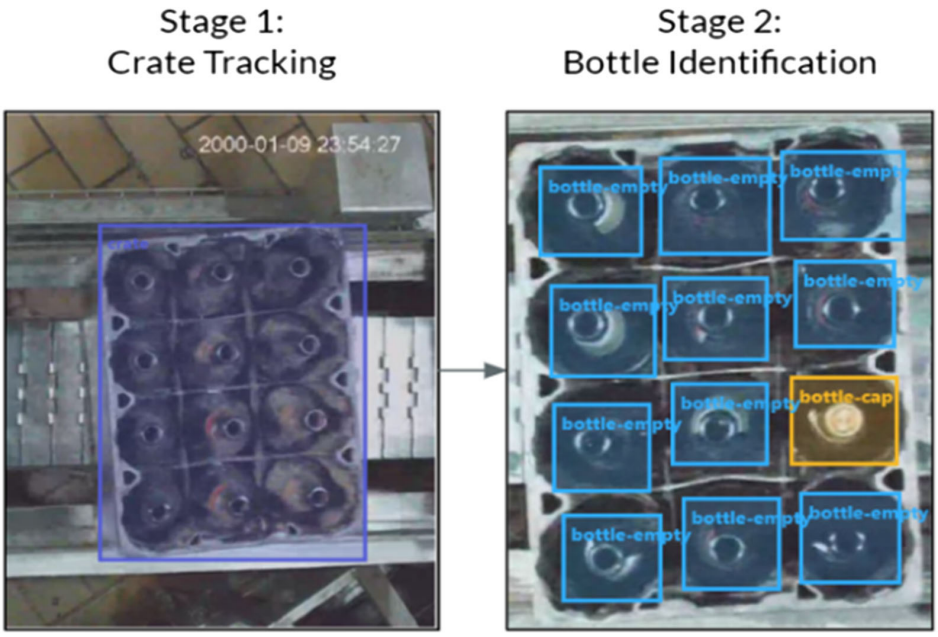
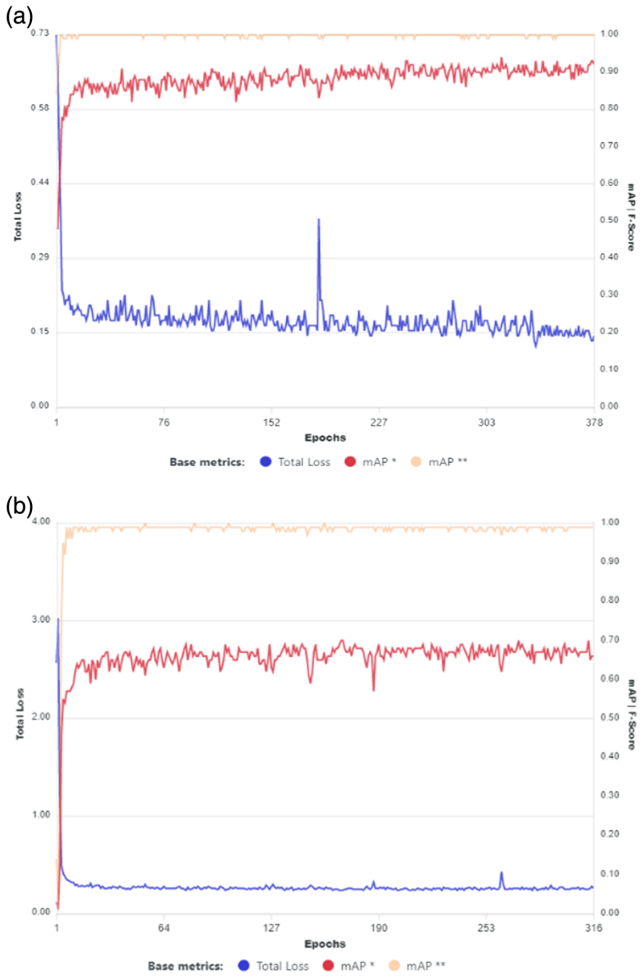


Figure 2
(a) Training results for crate detection and (b) training results for bottle detection



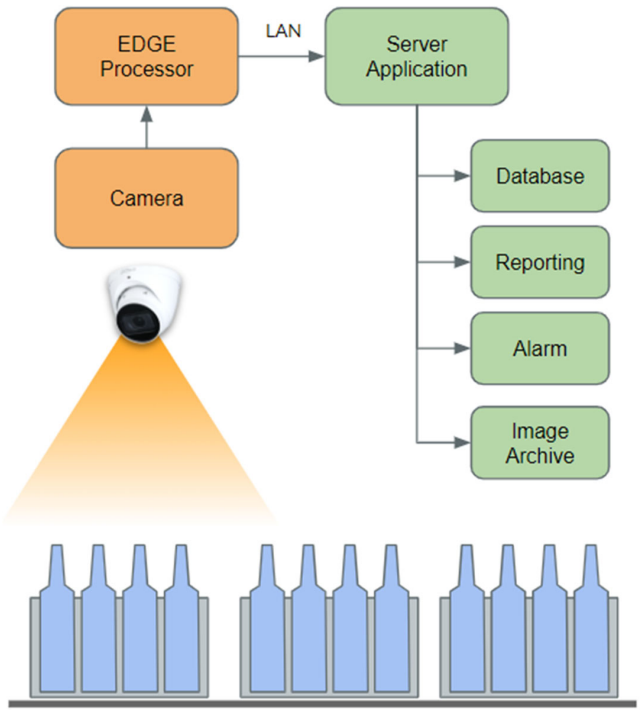
crop the crate was made to eliminate any potential confusion or background noise that may arise from the high-resolution image. Detection of the bottles occurred only when the crate passed directly underneath the camera, allowing for a clear view of the crate pockets. The artificial intelligence (AI) model was then trained to detect missing bottles, empty bottles, and capped bottles to ensure that any abnormalities or anomalies in the bottling process could be quickly detected and addressed. The second model achieved an impressive mAP (IoU@0.05:0.95) score of 0.99, with a loss of 0.24 after 169 epochs (see Figure 2(b)). The lengthy training time of the model speaks to the thoroughness of the training process, and the model's high accuracy speaks to the efficacy of the process in detecting bottles with precision and consistency.

3. System Architecture

The overall system architecture includes a HD camera and EDGE processor sending processed data and images to a server on the same network (see Figure 3).

Deploying AI models on EDGE Internet of Things (IoT) devices has been gaining attention in recent years due to their ability to perform AI processing at the edge without relying on a central server (Khan et al., 2019; Li et al., 2019; Shi et al., 2016).

Figure 3
System architecture



In this context, EDGE devices are chosen to perform AI processing as they are capable of carrying out computing tasks closer to where the data are generated, reducing the latency associated with data transmission to a remote server (Mao et al., 2021). This results in quicker decision-making, better resource allocation, and reduced costs. In our study, we deployed our AI model on an EDGE IoT device to perform real-time AI processing on the device itself. The EDGE IoT device used in the project had the ability to perform both Central processing unit (CPU) and Graphics processing unit (GPU) computations, making it capable of running AI inference in real time at the edge. To ensure uninterrupted service, the device was connected to an uninterruptible power supply power line.

A HD camera was connected to the device, and the client's network was connected via an ethernet cable. Additionally, the EDGE IoT device was designed to store inference results and images locally in case it became disconnected from the network temporarily. These stored results and images were placed in a queue, ensuring that the system could continue functioning until the network connection was restored. The device was equipped with advanced features to guarantee the reliability and accuracy of the system's output. The processed data logs were then sent to a central server for archiving purposes only, enabling data collection and analysis for future improvements.

Generating reports for clients is an essential part of data analysis, allowing them to gain insights into the data collected (Kim et al., 2015; Wu et al., 2014). In our study, we collected data on a central server, and once collected, we generated reports for the client. These reports were tailored to their specific requirements and included daily and hourly reports, providing them with up-to-date information on the data collected. Additionally, the client was provided with access to the database, enabling them to extract data and generate their own reports. This allowed them to analyze the data in more depth and draw insights relevant to their business needs. The client's access to the

database also provided them with a level of control over the data, allowing them to conduct their own analyses and extract the data they required, which further improved their ability to make informed decisions based on the data collected.

4. Validation Methodology

To evaluate the performance of the AI model, a rigorous validation test was conducted over a period of 5 consecutive days. The model was tasked with tracking a total of 128,000 crates and scanning over 1.5 million crate pockets, a substantial amount of data that tested the model's ability to handle real-world scenarios. The validation testing was conducted by two separate trained individuals using an intuitive application, ensuring the reliability and accuracy of the test results. To guarantee a 95% confidence level and a 5% margin of error, a representative sample of bottles and crates was sampled during the validation testing, providing a comprehensive assessment of the model's performance. These validation test results demonstrate the robustness and reliability of the AI model in accurately detecting full and empty bottles in crates, providing a practical and effective solution for industrial applications.

5. Results

In this project, the primary goal was to develop an AI model that could accurately detect full and empty returnable bottles in crates with a high level of precision. To assess the performance of the model, we conducted a thorough validation test, which yielded remarkable results. Specifically, we found that the model achieved a validated accuracy of 99.9996% for crates of full bottles, which was an impressive achievement. Similarly, for crates of empty returnable bottles, the model demonstrated a validated accuracy of 99.9206%, which was highly satisfactory. It is worth noting that the requirements for this project stipulated a target accuracy of 99.9%, which was comfortably surpassed. These results attest to the model's effectiveness in accurately detecting full and empty bottles in crates, thus meeting the project's accuracy requirements and providing a reliable solution for industrial applications.

6. Future Work

In order to further enhance the capabilities and performance of the object detection model in industrial applications, several avenues for future work are being pursued. One key area of development is the creation of an interface that allows users to validate continuously and to re-train the model automatically when a new dataset has been created by the client. This approach will enable the AI model to evolve over time, improving its performance and accuracy as new data are gathered, which is a critical requirement in the fast-paced manufacturing industry. Furthermore, we plan to investigate the potential benefits of object segmentation techniques to further enhance the accuracy of bottle detection (He et al., 2017). Another promising avenue for future research is the use of unsupervised learning techniques to reduce the need for manually annotated data, which can significantly speed up the training process and improve the scalability of the model. In addition, we aim to explore the use of more advanced GPUs to accelerate the training and inference process (Chollet, 2017), allowing for even faster and more efficient data processing. Finally, we plan to deploy the model across additional bottling plants and monitor its performance over time, ensuring that the model remains effective and reliable in real-world manufacturing environments. By pursuing these areas of research and development, we hope to further advance the capabilities and performance of the object detection model, providing practical and effective solutions for the bottling industry and other manufacturing applications.

Acknowledgement

The AI modeling and training tools utilized in this project were made available by https://firststep.ai/. The necessary funding for the project was secured from a private entity, whose identity must remain undisclosed.

Conflicts of Interest

The author declares that he has no conflicts of interest to this work.

References

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint:2004.10934*.

Cao, Y., Qin, K., Shao, Z., Liu, M., Liu, B., & Gao, R. X. (2021). A dual-task learning method for efficient defect detection in car-body production. *Journal of Intelligent Manufacturing, 32*(4), 1011–1024.

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 834–848.

Chen, Y., Wang, J., Guo, Y., Lu, C., Cheng, Y., Guo, J., & Chen, X. (2020). A deep learning framework for wafer defect detection based on improved mask R-CNN. *IEEE Transactions on Semiconductor Manufacturing*, *33*(4), 565–572.

Chollet, F. (2017). *Deep Learning with Python*. USA: Manning Publications Co.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 580–587.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering, 21*(9), 1263–1284.

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, 2961–2969.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Khan, W. Z., Ahmed, E., Hakak, S., Yaqoob, I., & Ahmed, A. (2019). Edge computing: A survey. *Future Generation Computer Systems, 97,* 219–235.

Kim, K. J., Ahn, J. H., & Kwon, H. Y. (2015). Towards big data analytics-enabled watermarking for cloud-based multimedia contents. *Multimedia Tools and Applications*, *74*(11), 3699–3715.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems,* 1097–1105.

Lee, H.B., Lee, H., Na, D., Kim, S., Park, M., Yang, E. & Hwang, S.J. (2019). Learning to balance: Bayesian meta-learning for imbalanced and out-of-distribution tasks. *arXiv preprint:1905.12917.*

Li, X., Liang, J., Li, S., Shen, S., & Liu, L. (2022). Neural architecture search for object detection: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(1), 73–89.

Li, Y., Li, L., Li, Y., Wu, X., & Li, M. (2019). An intelligent edge computing system for smart home. In *2019 IEEE International Conference on Artificial Intelligence and Computer Applications*, 55–59.

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2117–2125.

Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, 2980–2988.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In *European Conference on Computer Vision*, 21–37.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440.

Mao, Y., You, M., Zhang, J., & Tang, J. (2021). Edge intelligence: Architecture and technologies. *Artificial Intelligence and Applications*, *37*(1), 17–28.

Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv preprint:1804.02767*.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241.

Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, Y. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, *3*(5), 637–646.

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*.

Sun, J., Wu, J., Wu, C., Zhang, L., Zhang, X., & Wang, Y. (2021). Object detection in smart manufacturing: a comprehensive review. *Journal of Intelligent Manufacturing*, *32*(3), 629–656.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, inception-ResNet and the impact of residual connections on learning. In *Proceedings of the AAAI Conference on Artificial Intelligence 31*(1), 4278–4284.

Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7794–7803.

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, *26*(1), 97–107.

Zhang, L., Zhang, L. & Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine, 4*(2), 22–40. http://doi.org/10.1109/MGRS.2019.2908800

How to Cite: Remmelzwaal, L. (2023). Object Detection and Tracking for Crate and Bottle Identification in a Bottling Plant Using Deep Learning. *Artificial Intelligence and Applications* 1(3), 175–179, https://doi.org/10.47852/bonviewAIA3202798