Inventory Monitoring at Distribution Centers Proposal

AbdelKarim Eljandoubi

August 19, 2023

1 Domain Background

A Distribution Centre serves as a spacious warehouse where products and goods are frequently kept in stock, ready for redistribution to wholesalers or retailers. These centers play a vital role in ensuring a smooth operational flow within a company's supply chain network, connecting the company, its suppliers, and ultimately the consumers.

Within the Distribution Center lies an expansive area containing numerous goods arranged in storage containers referred to as bins. These centers buzz with constant activity as goods are continuously moved in and out. Hence, it is of utmost importance that this movement transpires efficiently, promptly, and within specified timeframes. The effective management of these processes directly influences the final pricing of products delivered to end users and indirectly affects overall inventory costs by minimizing stock holdings.

2 Problem Statement

The manual task of keeping track of sales and inventory records within a Distribution Centre can become exceedingly burdensome, susceptible to errors, and notably inefficient—particularly for companies operating multiple Distribution Centers and stores that handle a substantial volume of products on a monthly basis. Achieving accurate, efficient, and cost-effective management of this process is crucial. Otherwise, the potential consequence is the inability to meet customer demand, leading to escalated costs and diminished revenue.

This challenge can be significantly alleviated through the application of Computer Vision and Machine Learning. These technologies enable the identification and categorization of diverse images found within specific storage bins. This, in turn, facilitates the calculation of item quantities, which holds the potential to aid in forecasting stock levels. Strategic inventory management has assumed heightened significance, and its complexity continues to grow as product development advances.

3 Solution Statement

The project will be addressed by capitalizing on advanced techniques in deep learning, specifically Convolutional Neural Networks (CNNs), along with machine learning approaches. The dataset at hand comprises a substantial collection of approximately 535,000 images depicting bins containing varying quantities of objects. Each image is accompanied by a metadata file that furnishes crucial details such as the object count, dimensions, and object types.

The primary objective of this undertaking is to create a model capable of classifying the number of objects within each bin. To execute this classification task effectively, the chosen approach involves employing a pre-trained Convolutional Neural Network (CNN). This CNN model will be subjected to further training and fine-tuning procedures using Amazon SageMaker.

The central purpose of this project is to establish robust evidence that machine learning techniques can indeed be harnessed to devise a system capable of inventory tracking. This system holds the potential to ensure the accuracy of delivery consignments by verifying that the correct number of items is included.

4 Datasets and Inputs

Data serves as the lifeblood of any organization. To successfully undertake the projects outlined below, it is imperative to gather and utilize specific dataset components from sources like the Amazon Bin Image Dataset, AWS, and the internet.

Dataset Details:

1. This dataset encompasses approximately 535,000 images in jpg format, representing various bins. Each image is paired with corresponding metadata detailing the items present within it. The metadata is presented in json format. This dataset is accessible through an S3 bucket named "aft-vbi-pds" within the US-east-1 AWS region.

2. Each image is assigned a unique numerical identifier, linked to its corresponding metadata. For instance, an image labeled as 1111.jpg will have a metadata file named 1111.json. This one-to-one correspondence facilitates data organization.

3. The quantity of items within each bin exhibits variability across different bins. For the project, only a subset of this data will be utilized. The selection process aims to include bins with item counts ranging from 1 to 5. Notably, the distribution of classes within this subset is not evenly balanced. A visual representation of the subset is provided below.

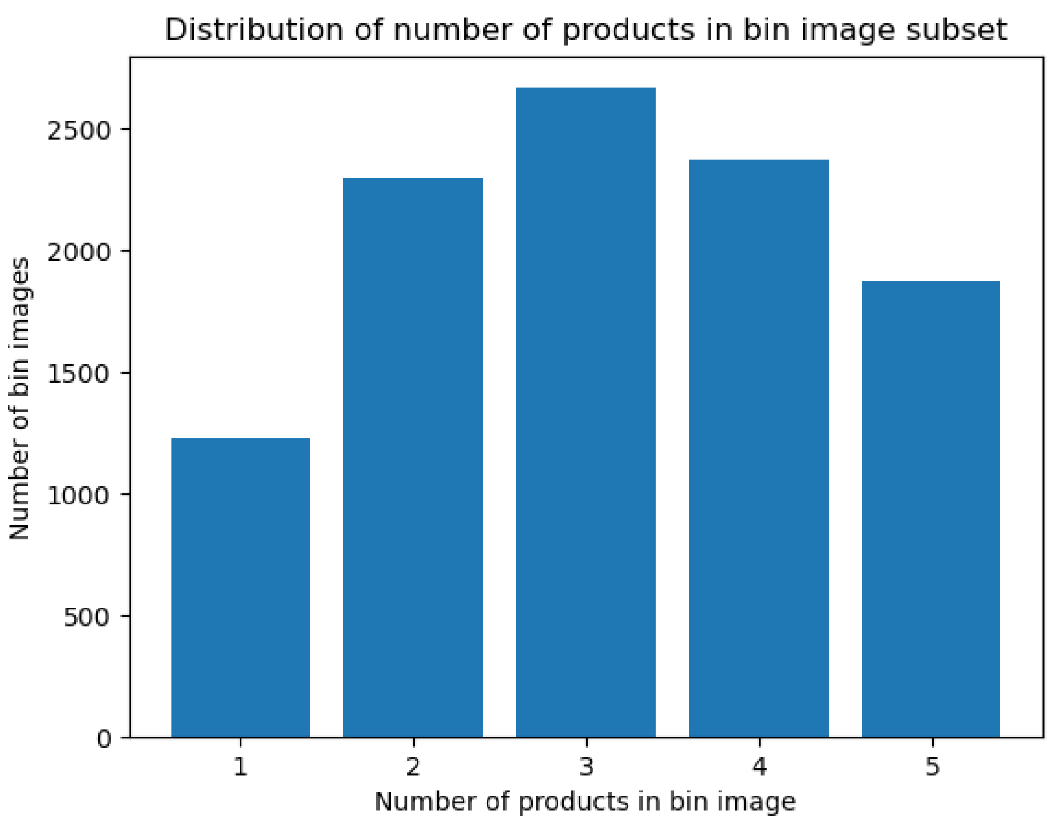

Figure 1: Distribution of number of products in bin image subset.

4. Within the metadata (json) files, essential details are included, such as the associated image's name, the anticipated item quantity, along with dimensions and units of measurement for each individual item. These elements contribute to a comprehensive understanding of the dataset.

5 Benchmark Model

During the hyperparameter tuning process, all variants of the ResNet architecture will be assessed. Subsequently, the variant that attains the highest testing accuracy score will be identified and chosen for implementation. This selected variant will then be utilized for the intended purpose.

6 Evaluation Metrics

Given that this constitutes a classification problem, the general accuracy of the classification holds value as a means to assess the effectiveness of the trained model's performance.

$$\text{accuracy}(\mathbf{g}, \mathbf{p}) = \frac{1}{n} \sum_{i=1}^{n} 1_{g_i = p_i} \tag{1}$$

Assigning $1_{x=y}$ as the indicator function, and denoting $\mathbf{p}$ as the prediction and $\mathbf{g}$ as the ground truth.

7 Project Design

The objectives of this project encompass training a machine learning model to classify the quantity of objects within a bin, while simultaneously demonstrating a comprehensive machine learning engineering pipeline. To accomplish these goals, the subsequent steps will be undertaken:

1. **Instance Selection:** Instances for computation will be chosen considering factors such as computational power and cost efficiency.

2. **Data Gathering and Preparation:** Data will be sourced from a database and subsequently uploaded to Amazon S3 to facilitate model training. The dataset will be partitioned into training (80%), validation (10%), and testing (10%) subsets. Preprocessing steps will involve operations such as resizing, random flipping, conversion to tensor format, and normalization before forwarding the data to the model.

3. **Hyperparameter Tuning:** SageMaker's hyperparameter tuning functionality will be harnessed to systematically explore a range of hyperparameters, aiming to identify the optimal set that enhances model performance.

4. **Training Phase:** Leveraging SageMaker and a dedicated training script, the chosen model will undergo fine-tuning and training using the identified optimal hyperparameters. The resulting trained model will be stored within the S3 repository.

5. **Debugger and Profiler Utilization:** SageMaker's debugger and profiler tools will be employed to gain insights for potential enhancements and to log critical metrics.

6. **Model Deployment:** The trained model will be deployed onto a SageMaker endpoint, enabling predictions to be made by providing an image as input.

In essence, this project is geared towards not only achieving accurate object classification within bins but also showcasing the robustness and comprehensiveness of the entire machine learning engineering process, from data acquisition and preprocessing to model selection, training, deployment, and performance evaluation.

References