Capstone Project Proposal for Inventory Monitoring at Distribution Centres

Tan Nguyen
AI Engineer at FPT Software AI Center
TanNP3@fpt.com

1. The Domain and Background of the Project

In the realm of Robotics, the genesis of our problem arises, particularly within its application in industrial domains such as Inventory Monitoring and Supply Chain Operations. Corporations engaged in managing physical cargo and overseeing supply chains have been actively pursuing automation to enhance efficiency and precision. Take Amazon, for instance, a colossal hub for diverse goods delivery. Within its expansive warehouses, the vast quantities of stored items necessitate inventory management on a grand scale. Manual inventory checks, given the enormity of these quantities, demand a significant investment in human resources and are susceptible to errors.

Enter robots, the champions of Inventory Monitoring. Equipped with Machine Learning Models, these robots adeptly perform tasks such as Object Detection, Outlier & Anomaly Detection, among others. Once trained, these models are easily scalable, offering a cost-effective solution applicable to warehouses and distribution centers, facilitated by industry-grade robots.

Transitioning to distribution centers, robots play an integral role in object transportation, often utilizing bins to carry multiple items. Yet, occasional mishandling can lead to discrepancies between recorded bin inventory and actual contents. A pivotal need emerges for a system that not only ensures precise inventory tracking but also guarantees the complete delivery of consignments by accurately counting items in each bin.

In the pursuit of this objective, our project endeavors to develop a robust model capable of performing item counts within individual bins. Such a system holds the potential to revolutionize inventory tracking, thereby safeguarding accurate delivery of consignments while maintaining item integrity.

2. The Problem Statement

As noted earlier in the domain description, distribution centers commonly employ robots to transport items. These items are housed within bins, with each bin having the capacity to hold between 1 and 5 objects.

The central challenge that this project endeavors to address involves the accurate enumeration of items contained within these bins. The significance of resolving this challenge cannot be understated, as it holds substantial practical applications. By crafting a model capable of processing images of bins and providing an accurate count of enclosed objects, we have the

potential to offer a transformative solution. This accomplishment would lead to the full automation of a pivotal stage within the Inventory Management process.

3. Dataset and Inputs

The Amazon Bin Image Dataset contains 536,434 images and metadata from bins of a pod in an operating Amazon Fulfillment Center. The bin images in this dataset are captured as robot units carry pods as part of normal Amazon Fulfillment Center operations. This dataset has many images and the corresponding medadata.

The image files have three groups according to its naming scheme.

- A file name with 1~4 digits (1,200): 1.jpg ~ 1200.jpg
- A file name with 5 digits (99,999): 00001.jpg ~ 99999.jpg
- A file name with 6 digits (435,235): 100000.jpg ~ 535234.jpg

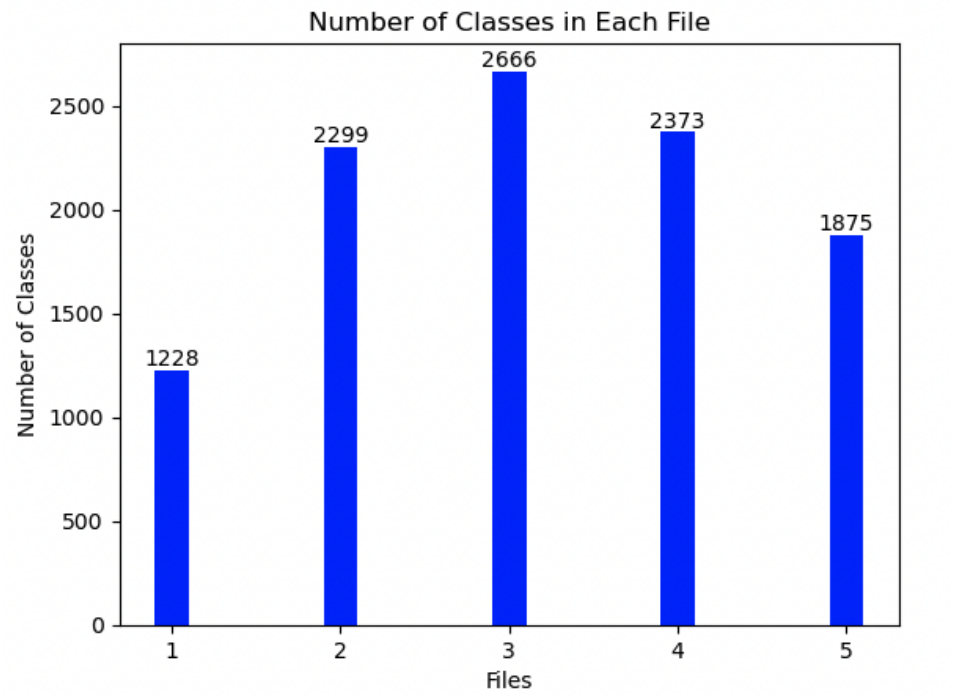
Sample metadata file:


```
{
    "BIN_FCSKU_DATA": {
        "B00PLKV5H6": {
            "asin": "B00PLKV5H6",
            "height": {
                "unit": "IN",
                "value": 6.799999999999999
            },
            "length": {
                "unit": "IN",
                "value": 7.0
            },
            "name": "HBD Thermoid NBR/PVC SAE30R6 Fuel Line Hose, 5/16\" x 25' Length, 0.3125\" ID, Black",
            "quantity": 1,
            "weight": {
                "unit": "pounds",
                "value": 3.0
            },
            "width": {
                "unit": "IN",
                "value": 7.0
            }
        },
        "B00WTI3SGO": {
            "asin": "B00WTI3SGO",
            "height": {
                "unit": "IN",
                "value": 1.2
            },
            "length": {
                "unit": "IN",
                "value": 7.6
            },
            "name": "The Witcher 3: Wild Hunt - PC",
            "quantity": 1,
            "weight": {
                "unit": "pounds",
                "value": 0.6
            },
            "width": {
                "unit": "IN",
                "value": 5.4
            }
        }
    },
    "EXPECTED_QUANTITY": 2,
    "image_fname": "500.jpg"
}
```


The “EXPECTED_QUANTITY” field is the total number of objects in image. However, since this dataset is too large to use as a learning project, and due to cost limitations on the Udacity AWS Portal, we will be using a subset of the data provided to us by Udacity itself.

Within this specific dataset segment, there exist 5 distinct categories, each corresponding to the quantity of items contained within a bin: 1, 2, 3, 4, and 5. The cumulative count of images within this particular subset amounts to 10,441.

Image illustrate distribution of sub-dataset which Udacity provide



4. Solution Statement

To solve our problem statement, we will use the task of Computer Vision, to come up with a Machine Learning Model, which given an image from our dataset, can identify the number of objects present in it. Essentially, we would be using Multi-Class Image Classification, with Number of Objects from 1-5 as an individual class.

In this project, we use pre-trained Transformer-based model because Transformer has achieved promising performance in computer vision such as BeIT, ViT.

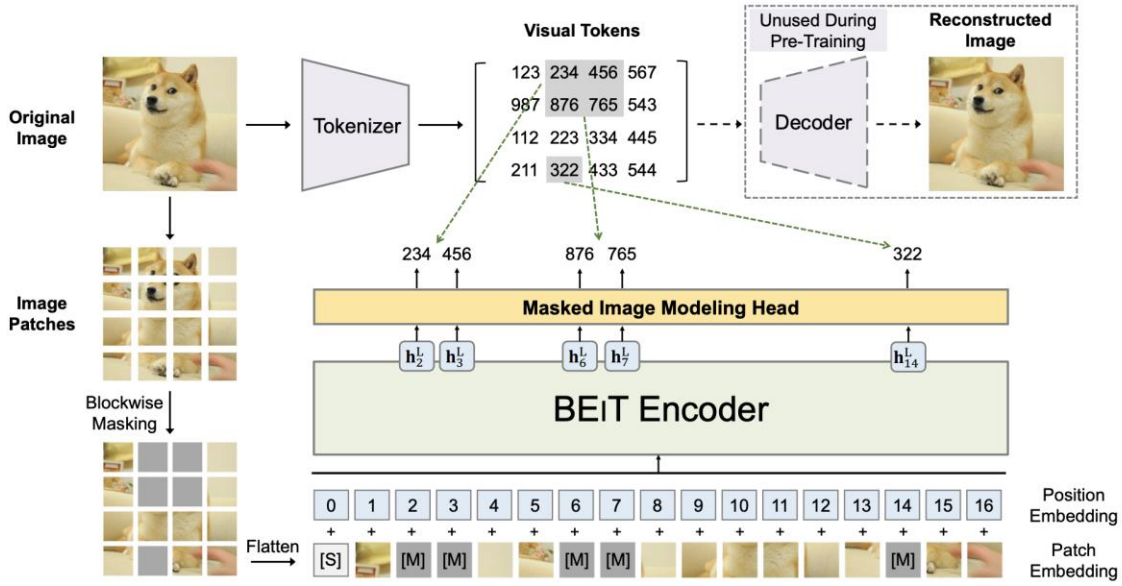


Fig 1: How to train BeIT

Because this research prove BeIT has better result than ViT on ImageNet dataset which is classification task. So we think we will use pre-trained BeIT model to fine-tune model with Amazon Bin Image Dataset.

Models	Model Size	Resolution	ImageNet
Training from scratch (i.e., random initialization)			
ViT_{384}-B [DBK+20]	86M	384^2	77.9
ViT_{384}-L [DBK+20]	307M	384^2	76.5
DeiT-B [TCD+20]	86M	224^2	81.8
$DeiT_{384}$-B [TCD+20]	86M	384^2	83.1
Supervised Pre-Training on ImageNet-22K (using labeled data)			
ViT_{384}-B [DBK+20]	86M	384^2	84.0
ViT_{384}-L [DBK+20]	307M	384^2	85.2
Self-Supervised Pre-Training on ImageNet-1K (without labeled data)			
iGPT-1.36B† [CRC+20]	1.36B	224^2	66.5
ViT_{384}-B-JFT300M‡ [DBK+20]	86M	384^2	79.9
MoCo v3-B [CXH21]	86M	224^2	83.2
MoCo v3-L [CXH21]	307M	224^2	84.1
DINO-B [CTM+21]	86M	224^2	82.8
BEIT-B (ours)	86M	224^2	83.2
$BEIT_{384}$-B (ours)	86M	384^2	84.6
BEIT-L (ours)	307M	224^2	85.2
$BEIT_{384}$-L (ours)	307M	384^2	**86.3**

The platform that will be used is AWS, more specifically the following services:

- - S3: It's all about storage. Whether it is data storage or model storage.

- - SageMaker Studio: it’s all about logic of the pipeline: data preprocessing, modeling, training, evaluation.

5. Benchmark Model

1. Amazon Bin Image Dataset (ABID) Challenge:

A notable technique employed in this study involves transforming the continuous task of object counting into a classification problem through effective exploratory data analysis (EDA). The ABID dataset is partitioned into two tasks: a moderate task and a hard task. The hard task encompasses object counting across all images, while the moderate task selectively focuses on images containing up to five objects. This classification-based approach effectively simplifies the problem into six classes (ranging from 0 to 5). The architecture of choice for training in this paper is ResNet34.

2. Amazon Inventory Reconciliation Using AI:

Building upon the concepts of the previous benchmark, this research introduces advancements that enhance the classification-based approach. Similar to its predecessor, the task's conversion into a classification problem is pivotal. The exploration of various linear and non-linear models, including logistic regression, classification trees, Support Vector Machines (SVMs), and Convolutional Neural Networks (CNNs), is a distinguishing feature of this study. Among these models, ResNet50, a prominent CNN architecture, emerges as the most effective choice for achieving optimal results.

3. Inventory Monitoring at Distribution Centers:

In this benchmark, the primary focus centers on devising a model that balances accuracy with simplicity to streamline cost-effectiveness. Several strategic techniques are employed to enhance performance, including weighted random sampling, data augmentation, model refinement, and multi-resolution training. The selected architecture for training is EfficientNet B0, aligning with the goal of achieving accuracy while minimizing complexity.

6. Evaluation Metrics

As for model evaluation, the following metrics will be calculated:

- F1 $= \frac{2*Precision*Recall}{\text{Precision+Recall}}$

Where:

- $Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$
- $Recall \frac{True\ Positive}{True\ Positive + False\ Negatie}$

Given the presence of class distribution disparities within the dataset, relying solely on accuracy as a performance metric is inadequate. Instead, the F1 Score has been adopted to provide a more comprehensive evaluation of model effectiveness, particularly in addressing the repercussions of imbalanced classes. While accuracy is not a suitable measure when facing dataset imbalances, there are instances where certain subsets of the data exhibit minimal imbalances. In such cases, it becomes plausible to overlook the adverse impacts of the imbalance and proceed with utilizing accuracy as a metric of assessment.

7. Project Design

- Data augmentation
- Hyperparameter tuning
- Model training with best hyperparameters
- Model evaluation

8. References

1. https://github.com/awslabs/open-data-docs/tree/main/docs/aft-vbi- pds
2. https://github.com/udacity/nd009t-capstone- starter/blob/master/starter/file_list.json
3. https://github.com/pablo-tech/Image-Inventory-Reconciliation-with-SVM-and-CNN
4. https://arxiv.org/pdf/2106.08254.pdf