Machine Learning Engineer Nanodegree

Capstone Proposal

Margaret Chan

December 15, 2024

Proposal

Domain Background

Computer vision is a field in AI that mimics how humans process visual information, it analyzes visual inputs to identify objects, recognize patterns, and make decisions. It is widely used in industries, such as healthcare, self driving vehicles, robotics, and more.

Advancements in deep learning algorithms, neural networks, and compute power, together with large data sets made available like ImageNet[1], have made it possible to replace or alleviate what used to be labor intensive and error prone processes with computer vision systems, which can help improve accuracy, efficiency, and lower cost.

Problem Statement

Distribution centers often use robots to move objects as a part of their operations. Objects are carried in bins which can contain multiple objects. Robots often need human operators to enter data about the contents of the bins, such as the number of objects to start with, which is a slow, mundane, and error prone process. In this project, we want to improve this process by developing a computer vision model that can identify the number of objects in a bin.

Datasets and Inputs

In this project, we will use images in the ‘Amazon Bin Image Dataset’ to train a computer vision model, images in this dataset were captured as robot units carry pods in an Amazon Fulfillment Center.

Official documentation of the dataset is at:

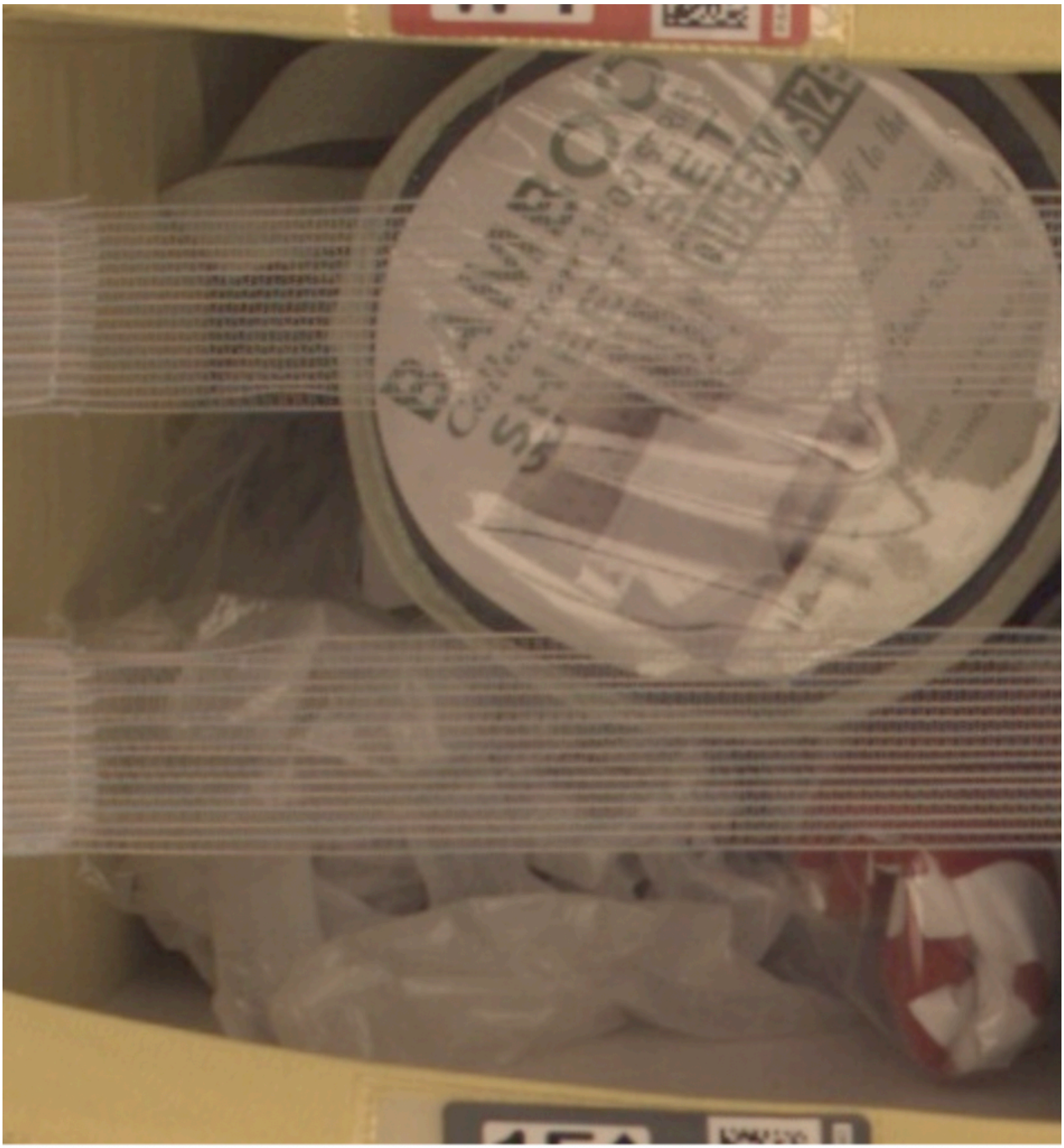
https://registry.opendata.aws/amazon-bin-imagery/

The dataset is hosted at a public S3 bucket:

s3://aft-vbi-pds

The dataset has >500,000 JPEG images of a bin with one or more objects inside, for each image, there is a corresponding metadata JSON file, which describes the object(s) with details like item name and dimensions etc, although for the purpose of this project, we are only interested in the object count.

Sample image:



The respective metadata JSON file:


```
{
   "BIN_FCSKU_DATA": {
      "B0040723CC": {
         "asin": "B0040723CC",
         "height": {
            "unit": "IN",
            "value": 1.599999998368
         },
         "length": {
            "unit": "IN",
            "value": 15.099999984598
         },
         "name": "Belkin 12 Outlet Surge Protector with 8-Foot Power Cord
(BV112230-08)",
         "normalizedName": "Belkin 12 Outlet Surge Protector with 8-Foot Power Cord
(BV112230-08)",
         "quantity": 1,
         "weight": {
            "unit": "pounds",
            "value": 2.2
         },
         "width": {
            "unit": "IN",
            "value": 6.99999999286
         }
      },
      "B00NUFF5JI": {
         "asin": "B00NUFF5JI",
         "height": {
            "unit": "IN",
            "value": 2.399999997552
         },
         "length": {
            "unit": "IN",
            "value": 8.299999991534
         },
         "name": "San Francisco 49ers fans. Proud to be a Hater Red T-Shirt (Medium)",
         "normalizedName": "San Francisco 49ers fans. Proud to be a Hater Red T-Shirt
(Medium)",
         "quantity": 1,
         "weight": {
            "unit": "pounds",
            "value": 0.022046001186074866
```

```
        },
        "width": {
          "unit": "IN",
          "value": 7.699999992146
        }
      },
      "B00PMEODQG": {
        "asin": "B00PMEODQG",
        "height": {
          "unit": "IN",
          "value": 6.199999993676
        },
        "length": {
          "unit": "IN",
          "value": 12.299999987454
        },
        "name": "High Strength Natural Bamboo Fiber Yarns Egyptian Comfort 1800
Thread Count 4 Piece QUEEN Size Sheet Set, GREY-BLUE Color",
        "normalizedName": "High Strength Natural Bamboo Fiber Yarns Egyptian
Comfort 1800 Thread Count 4 Piece QUEEN Size Sheet Set, GREY-BLUE Color",
        "quantity": 1,
        "weight": {
          "unit": "pounds",
          "value": 3.0
        },
        "width": {
          "unit": "IN",
          "value": 7.49999999235
        }
      },
      "B01B4Q4POU": {
        "asin": "B01B4Q4POU",
        "height": {
          "unit": "IN",
          "value": 5.49999999439
        },
        "length": {
          "unit": "IN",
          "value": 9.699999990105999
        },
        "name": "Rebecca Minkoff Love Silver Hardware Convertible Cross Body, Black,
One Size",
        "normalizedName": "Rebecca Minkoff Love Silver Hardware Convertible Cross
Body, Black, One Size",
        "quantity": 1,
        "weight": {
```

```
            "unit": "pounds",
            "value": 0.7999999993291277
         },
         "width": {
            "unit": "IN",
            "value": 6.99999999286
         }
      }
   },
   "EXPECTED_QUANTITY": 4
}
```


Solution Statement

In this project, we will use the images from the 'Amazon Bin Image Dataset' to train a computer vision model with CNN, and use the model to predict the number of objects in a bin, along with the probability value which indicates how confident the model is about the prediction.

Benchmark Model

I plan to call it a win (or stop training) when the training / validation and testing accuracy reaches 0.8 (or 80%), which is an acceptable value for this non-healthcare use case.

Evaluation Metrics

I will use accuracy:

$$Accuracy = \frac{\sum_{i=1}^{T} (p_i = l_i)}{T} \quad where\ p\ is\ predicted,\ l\ is\ label,\ T\ is\ test\ set\ size$$

Project Design

The product will be a REST API, where the client (distribution center robots) can call to make prediction, a few AWS services will be leveraged for development, among them is SageMaker where the computer vision CNN model will be trained and deployed, as illustrated in the diagram below.

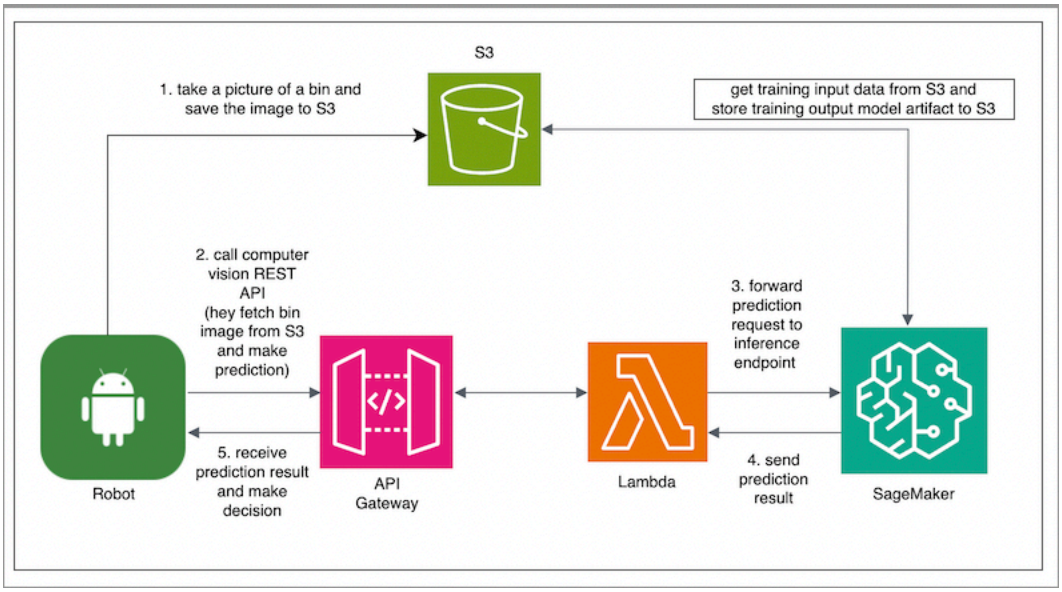


Fig. 1 Architecture Diagram

Reference

1. J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848. keywords: {Large-scale systems;Image databases;Explosions;Internet;Robustness;Information retrieval;Image retrieval;Multimedia databases;Ontologies;Spine}, *https://ieeexplore.ieee.org/document/5206848*