Inventory Monitoring at Distribution Centers

Domain Background

Most distribution centers all over the world use robots to move objects from one place to another. These robots use bins which contains multiple objects. Determining the number of objects in each bin can be very valuable in order to check that the process is working as expected.

The main goal of this project is to build a ML model that can count the number of objects in each bin in order to track inventory and check that bins have the appropriate number of items in order to reduce stock mismatches.

Problem Statement

Based on the background, it can be seen that the problem to be resolved here is related to image classification. A ton of images have been provided by our client (Amazon) and a ML model will be built in order to identify the number of objects in each bin.

Datasets and inputs

The dataset used for the project will be the Amazon Bin Image Dataset which contains more than 500,000 images and metadata from bins of a pod in an operating Amazon Fulfillment Center. Both, images and metadata, will be used to develop the model.

Images will be used as the main input in the training phase. Metadata will be used in order to arrange the pictures in a way that the ML model will identify the number of possible classes.

It is important to remark that the dataset will be split into training and test so we can evaluate the performance of the model while training.

The number of classes used for this project will be directly related to the number of objects which should be identified in every picture. In this case, we will use 6 classes:

- Class 1 for pictures without objects
- Class 2 for pictures with 1 object
- Class 3 for pictures with 2 objects
- Class 4 for pictures with 3 objects
- Class 5 for pictures with 4 objects
- Class 6 for pictures with 5 objects

Solution Statement

The solution will consist on a ML model which will identify the number of objects in each image. In order to build this ML model, SageMaker will be used, training a model using a ResNet neuronal network.

ResNet model is widely used for image classification which is pretrained and can be customized in order to categorize images from different use cases. To adapt this pretrained model to our use case, different training jobs will be launched in AWS SageMaker. In addition, hyperparameters tunning jobs will also be launched in order to find the most appropriate combination of hyperparameters for our use case.

Benchmark Model

Others have worked on the same data to build their own models. Specifically, we can find two GitHub projects related to this problem:

1. Amazon Bin Image Dataset (ABID) Challenge by silverbottlep
2. Amazon Inventory Reconciliation using AI by Pablo Rodriguez Bertorello, Sravan Sripada, Nutchapol Dendumrongsup

As can be seen in the conclusions of both projects, the obtained accuracy is 55% approximately with a RMSE of 0.94. The main goal of this project will be trying to find a model with a similar performance.

Evaluation Metrics

For this case, we will use both, the model accuracy (based on the images which are correctly identified by the model) and the Root Mean Square Error (RMSE) metrics in order to evaluate how the model is performing.

These metrics will be used at the end of each epoch in order to observe how the model is improving its results.

Project Design

1. Download data from Amazon S3
2. Put the data in an appropriate folder according to the number of objects contained in every image.
3. Split the dataset into training and test
4. Upload both, training and test dataset to S3.
5. Create a `train.py` file in order to train the model. This fill will load a pretrained ResNet neuronal network and will modify it in order to set the appropriate number of classes according to our dataset.
6. Launch the training using SageMaker.
7. Wait for the model to be trained and observe the results.
8. Use hyperparameters tunning in order to obtain which hyperparameters combination offers the best model performance.
9. Deploy the model in an endpoint so we can start using it to identify the number of objects in new images.