Capstone proposal

Sebastian Aranguiz
Dec 4th, 2023

Machine Learning Engineer Nanodegree

Domain Background

Distribution centers are increasingly relying on robotics to streamline their operations, and many of these robots are now equipped with cameras that can capture images of objects being shipped. However, accurately counting objects in bins remains a significant challenge for distribution centers. Inaccurate inventory counts may lead to stockouts, overstocking, and discrepancies in delivery requests.

Problem Statement

There is a need for automated object counting systems that can accurately count items in bins at distribution centers. Such systems would utilize computer vision techniques to process images of bins and determine the number of objects present.
One approach is using image-based methods, which employ machine learning algorithms to analyze two-dimensional images. The algorithms are trained on labeled datasets to identify and count objects in images [1].

Solution Statement and Project Design

This project aims to develop an image classification model to predict the number of items in a bin image, ranging from one to five items. The model will output the probability of each item count category. Model performance will be assessed during training, validation, and testing phases.
The project primarily focuses on implementing a practical Machine Learning engineering pipeline using AWS Sagemaker.
The following steps will be implemented as a solution:

- Conduct exploratory data analysis (EDA) on image samples to gain insights into the data distribution, identify patterns, and detect anomalies.
- Compare the performance of a recently developed model architecture against traditional computer vision models to determine its effectiveness.
- Optimize the hyperparameters of the chosen model to enhance its predictive accuracy and ensure optimal performance.
- Train, debug, and profile the model to refine its functionality and identify any potential issues (using spot instances, if available).
- Deploy the model to a production environment and perform inference to utilize its predictive capabilities in real-world scenarios.

Datasets and Inputs

For this project, we will use a subset of the *Amazon Bin Image Dataset* [2], which consists of a large set of images (500,000) of bins containing one or more items. Due to computing power constraints, only a portion of the dataset will be used for training.
Each image has associated metadata, including the number, dimensions, and types of objects. Our main focus, nevertheless, will be on classifying the number of items present in each bin (1 to 5 objects).

Benchmark Model

ResNet is a popular pre-trained model among computer vision and it's widely used for image classification. In this project proposal, it is aimed to apply **ResNeXt**, a newer image classification model which was introduced in the *"Aggregated Residual Transformations for Deep Neural Networks"* paper [3]. It claims to have better results than its ResNet counterpart. Thus, a quick benchmark will be conducted as an experimental prototype to confirm the better performance of *resnext101-32x8d* against a traditional *resnet50* model.

Evaluation Metrics

Model performance will be evaluated primarily based on accuracy on the training, validation, and testing datasets. Accuracy is a measure of how well a classification model performs. In simple terms, accuracy is the proportion of predictions that the model makes correctly [4]. Formally, accuracy is defined as:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

References

[1] Liu, X., Ma, L., Zhang, T., & Feng, J. (2019). Object counting in images of distribution center bins using deep learning. In 2019 International Conference on Computer and Communication Systems (ICCCS) (pp. 1031-1035). IEEE.
[2] Amazon Bin Image Dataset. Registry of Open Data on AWS.
https://registry.opendata.aws/amazon-bin-imagery/
[3] Xie et al (2017). Aggregated Residual Transformations for Deep Neural Networks.
[4] Google's Machine Learning Foundational course. Classification: Accuracy.
https://developers.google.com/machine-learning/crash-course/classification/accuracy