Capstone Project Proposal

1. Domain Background

A warehouse management system is a set of policies and processes intended to organize the work of a warehouse or distribution center, and ensure that such a facility can operate efficiently and meet its objectives (Wikipedia 2024a). Today, warehouses are constantly growing in size and complexity and it's important to develop efficient processes that can accurately track the inventory. A major player in the field of warehouse management is Amazon. I'm interested in how a company, such as Amazon, has been able to build global warehouse and supply chain management systems, and I want to look into how to streamline and improve processes in these systems.

2. Problem Statement

Industrial robots are used in warehouses to move items as part of operations. These items are carried in moveable shelf units containing a number bins. Each bin contains one or more items. The aim of this project is to build a machine learning model that can count the number of items in each bin. A system like this can be used to track inventory and to ensure consignments have the correct number of items. As we are dealing with counting the number of items in a bin, the problem is quantifiable and measurable.

3. Solution Statement

The warehouse robots are equipped with cameras and take photos of the bins in the moveable shelf units. Amazon has made these images publicly available.

The solution is to use a **supervised machine learning algorithm** to classify these images according to the number of items in each bin. In the model, each class corresponds to the number of items detected in the image. A pre-trained convolutional neural network can be used or a customized neural network architecture can be developed. A suitable place to train and evaluate this model is by means of a Jupyter Notebook in Amazon SageMaker.

4. Datasets and Inputs

The project will use the **Amazon Bin Image Dataset**. This dataset is publicly available and contains 500,000 images of bins containing one or more items. The bin images are captured while the robots move shelf units around the warehouse. This forms part of the normal operations in an Amazon Fulfillment Center. Information about the dataset can be found in following location:

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

Furthermore, each bin image has a corresponding metadata file, containing information about the image, such as, the number of items, its dimension and object type. Below are sample images from the dataset.

Figure 1: Sample images from the Amazon Bin Image Dataset

To reduce the cost of training a subset of the full dataset will be used. This subset will be further divided into 3 categories: training, validation and testing. The training category contains 90% of images, the validation category contains 5% of images and the testing category contains 5% of images.

5. Benchmark Model

A suitable benchmark model is **ResNet** (Residual Network). It is a deep learning model in which the weight layers learn residual functions with reference to the input layers. The model was developed in 2015 for image classification. It won that year's ImageNet Large Scale Visual Recognition Challenge (Wikipedia 2024d). **ImageNet** is a large visual database designed for use in visual object recognition software. It serves as a dataset benchmark and contains more than 14 million hand-annotated images and more than 20,000 classes (Wikipedia 2024e).

In the project, transfer learning can be used to reduce the time and cost of training. In transfer learning, the weights in a pre-trained version of ResNet will be frozen. Then, an output layer will be added with the correct number of classes. Only the weights in this output layer will be trained.

6. Evaluation Metrics

A suitable evaluation metric in this project is the **test accuracy**. As the name suggests, the test accuracy is measured on images reserved for testing only.

For a binary classification problem, accuracy is measured by means of following formula (Wikipedia 2024b):

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

where,

TP are true positives
TN are true negatives
FP are false positives
FN are false negatives

This method can be expanded to measure the accuracy of a multiclass classification problem. In our case, there are 5 classes, where each class corresponds to the number of items in the bin. There are also other types of evaluation metrics that can be used in this project, for example, **test loss**. Test loss represents the 'price paid' for inaccuracy of predictions on the test dataset (Wikipedia 2024c).

7. Project Design

In general, a machine learning workflow consists of following steps:

- Preprocessing
- Training
- Evaluation
- Deployment

During preprocessing, the data is collected and stored. If necessary, the data is then cleaned and refined. The dataset is then divided into training, validation and testing. Once the preprocessing has been completed, training is then carried out on the dataset designated for training. In terms of algorithms, this project will use ResNet, which is a popular, high-performing algorithm for image classification (Wikipedia 2024d). The training may involve several iterative steps, including tuning of the hyperparameters. During training the model is evaluated using the evaluation dataset. After the training has been completed, the model is evaluated on the test dataset, which is untouched and has not been used during training. If the evaluation is satisfactory, the trained model is then deployed. As we are using SageMaker in this project, the deployment consists of creating a SageMaker endpoint that can be queried by the user.

References

Wikipedia 2024a, https://en.wikipedia.org/wiki/Warehouse_management_system

Wikipedia 2024b, https://en.wikipedia.org/wiki/Precision_and_recall

Wikipedia 2024c, https://en.wikipedia.org/wiki/Loss_functions_for_classification

Wikipedia 2024d, https://en.wikipedia.org/wiki/Residual_neural_network

Wikipedia 2024e, https://en.wikipedia.org/wiki/ImageNet