Machine Learning Engineer Nanodegree

AWS Capstone Proposal

Ledezma, Lucia del Valle

2023

1. Domain Background:

The domain background of this project is Computer Vision.

Computer vision is a field of artificial intelligence that enables computers to derive meaningful information from digital images, videos and any other visual inputs [1]. For this project the subfield related to is Image Classification, it is the task of identifying images and categorizing them in one of several predefined distinct classes [2]. A convolutional neural network is a method of deep learning that takes an input image and assigns importance (learnable biases and weights) to various objects in the image, distinguishing one from the other [3]; furthermore, it has demonstrated exceptional image classification, recognition, segmentation, and retrieval [4].

This particular problem requires us to use images as input, and we need to classify each of them into classes, so the knowledge of this domain will help us to solve the problem, and I choose to use CNN to achieve the solution.

2. Problem statement:

In a distribution center, all the objects are managed by robots, these robots are responsible to take every object into bins. How can we count how many objects there are in each bin?. The objects can be countered by a system in order to get an inventory. This involves knowing how to classify the images into object's amount labels by using Computer Vision techniques.

3. Dataset Inputs:

For this project, The Amazon Bin Image Dataset is going to be used. This dataset contains almost

500000 images and metadata from bins of Amazon Fullfillment Center. In this case I will only use a subset of it.

4. Solution:

Train a Convolutional Neural Network in order to get a model that classifies the images into the classes corresponding to the number of objects these have. A pretrained model is going to be used for this proposal.

5. Benchmark Model:

Maybe any simple CNN could be used for solve this problem, but in this case as it is mentioned before, resnet18 is going to be the base of this benchmark model I will train, this kind of model are already tested, so a transfer learning is required, and this option is more efficient, because we only have to train the model in the new classes then compare it with the proposed model.

6. Evaluation metrics:

Accuracy is one of the most important evaluation metrics, this represents the percentage of correct predictions of the model:

$$Accuracy = (Correct\ Predictions)/(\ total\ predictions)$$

By having the confusion matrix, it could being explained as:

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

7. Project design:

I am going to use the AWS SageMaker environment for training and deploying the solution.

7.1 Data:

I am going to recover the dataset from the source of AWS. I will split the dataset into train, eval and test. then I am going to upload them into S3; to prepare the data, I am going to resize each image to the same size and apply some transformation.

7.2 Model training:

I will train the benchmark model by using the splitted dataset. After that I will train a convolutional neural network by using some pretrained model such as resnet50. Both of these steps are going to be developed by using Pytorch Framework.

7.3 Model evaluation and comparison:

I will evaluate the model in a test set and compare its performance with the benchmark model.

7.4 Model Deployment:

Once a model is ready, I will deploy it into a Sagemaker endpoint and make predictions of sample images.

References

[1] What is computer vision?. [Online]. Available: https://www.ibm.com/topics/computer-vision

[2] What is image recognition and computer vision?. [Online]. Available: https://www.ibm.com/topics/computer-vision .

[3] Khan, A.; Sohail, A.; Zahoora, U.; Qureshi, A.S. A Survey of the Recent Architectures of Deep Convolutional Neural Networks. Artif. Intell. Rev. 2020, 53, 5455–5516.

[4] Liu, X. Recent progress in semantic image segmentation. Artifical Intell. Rev. 2019, 52, 1089–1106.