Capstone Project Proposal

By

Akpojicheko Eyekpegha

Domain background

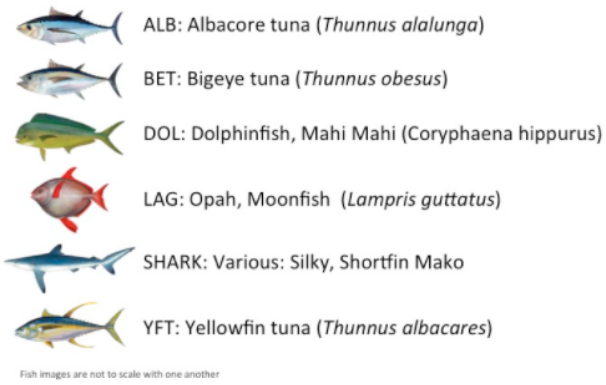
The capstone project, which is gotten from the kaggle competition- **The Nature Conservancy Fisheries Monitoring**, uses computer vision to solve a problem in the area of aquatic science and oceanography. As a fisheries graduate and data annotator, I am interested in completing this programme with the use of amazon sagemaker. Fish classification is very useful especially in the wild. For instance, when large fishing vessels harvest fish, a significant amount of the harvested stock may be untargeted in the first place. Over the years, there have been a lot of misuse and wastage of aquatic resources, since these fishing vessels find it difficult to recognise targetted fish in deep parts of the ocean or sea

Problem statement

On kaggle, the competition is organized by the Nature conservancy with the aim of finding better ways to monitor unregulated fishing activities that are threatening marine ecosystems and global seafood supplies. This further poses challenges to aquatic resource management.

Datasets and inputs

The dataset is made available by Kaggle, assessed by the link https://www.kaggle.com/c/the-nature-conservancy-fisheries-monitoring/data It consists of train images (3792 images), and 2 test datasets for 2 stages of training. Available in this dataset are eight classes to predict from. The Eight target categories comprise six fish species (in the image below) with two other classes - other (fish species other than the targetted six classes of fish) and No Fish (meaning that no fish is in the picture). To start up the project, I would unzip and upload these files to my s3 bucket to make them available for training with sagemaker. I will conduct some feature engineering where necessary to get better results. The train images would be used for the training job, after which testing is done twice for the 2 test datasets



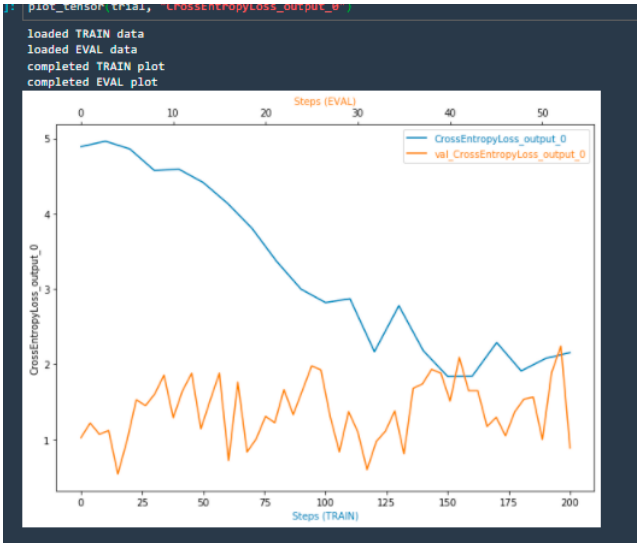
This photo is gotten from the kaggle competition.

Solution statement

The proposed solution is to create a deep learning model that is able to identify accurately the presence of fish (basically six fish species) with a high percentage accuracy by using amazon sagemaker. This solution can enable countries to better allocate human capital to management and enforcement activities, instead of fish classification.

Benchmark model

The solution will be compared to having an effective cost function, this would be shown graphically in the notebook. My model in the capstone project ought to reduce its cross-entropy loss function significantly during training. The graph below shows a significant reduction in cross-entropy loss during training (this is gotten from my project 3 notebook)



With effective feature engineering and hyperparameter tuning, even a better reduction than this can be achieved

Evaluation metrics

The evaluation metric for this project, which is used to determine the efficiency of the model will be done through debugging and profiling. With these metrics, the performance of the model can be evaluated effectively

Project design

The workflow to approaching the solution is as follows:

With the right exploratory data analysis, I will explore the data to understand, prepare and clean up the data (if necessary), making it easy to train the model to obtain the desired solution

I will train the model by specifying the sagemaker environment and commence training by creating a pytorch estimator, setting hyperparameters before calling model.fit(). Sagemaker debugging and profiling will also be done on the model. The model will be finally deployed to an endpoint.