Introduction

With the advent of modern technologies, we have seen unprecedented growth, yet this growth is proving to be problematic. An area that has been heavily impacted by these modern technologies is our oceans. This project seeks to create a solution to an issue facing commercial pelagic long liners in the Pacific Ocean. An industrial long liner can set over 500 miles (800 km) worth of line with over tens of thousands of hooks. With so many hooks, long liners have trouble keeping track of catches, and this is where this deep learning project seeks to help. By using deep learning image labelling, this project aims to be able to identify fish species in an image. The data set was retrieved from https://www.fishnet.ai/download from the v0.2.0 Dataset (September 2020) and has 12 GB of images containing different fish and animal species from commercial long liners operating in the Pacific. The goal of this project is to be able to accurately return the species present in an image and be able to predict the species in an image.

Dataset

The original CSV ("*foid_labels_bbox_v020.csv*") contains 9 categories the image name, the bounding box, its corresponding dimensions, and the corresponding labels with the bounding boxes. For the purpose of multi-labeling images, in this project only *label_l2* and image name is utilized. The categories in *label_l2* are:

> *HUMAN, NoF, YFT, ALB, OTH, BILL, DOL, BET, SKJ, LAG, SHARK, OIL, PLS, TUNA, WATER* [each label represents either a kind of fish or an object that appears on the camera]

Due to RAM complications, this project takes a randomly selected group of 10,000 images, which were saved into a CSV file ("*Subsetted_Fish.csv*"). The fish data frame takes the image id and 15 categories based on *label_l2*. Each column in the fish data frame is a *label_l2* category, and if the image contains one of those categories it has a 0 in the respective column else it has a zero (binary encoding of contains? Y/N).

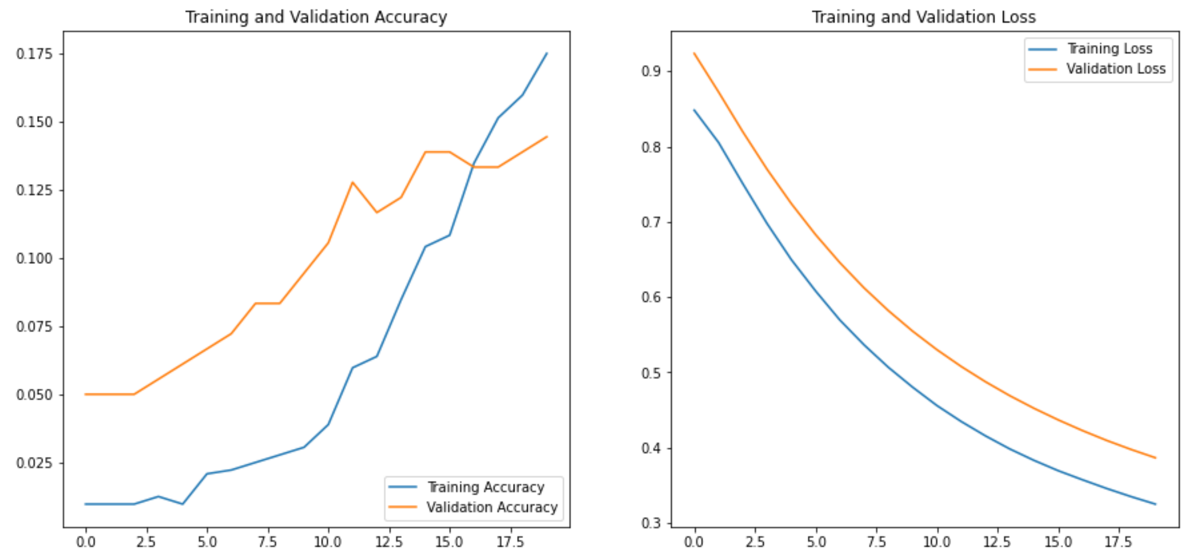
Proposed Method

The model utilizes images that are in color and have been scaled into sizes of (400,400,3) arrays, and then scaled down by 255, thus keeping their color. Having the images stay the same color was vital because the only way to differentiate certain fish is different patterns of their colors, which grayscale would not capture in as much color as RBG. The images were then stored in an array, *X*, which has the dimensions of (1000, 400, 400, 3). The next array created was the array that stored labels called *Y*, which had the dimensions of (1000, 14). The fourteen in this array represents the categories of *label_l2* without the human category.

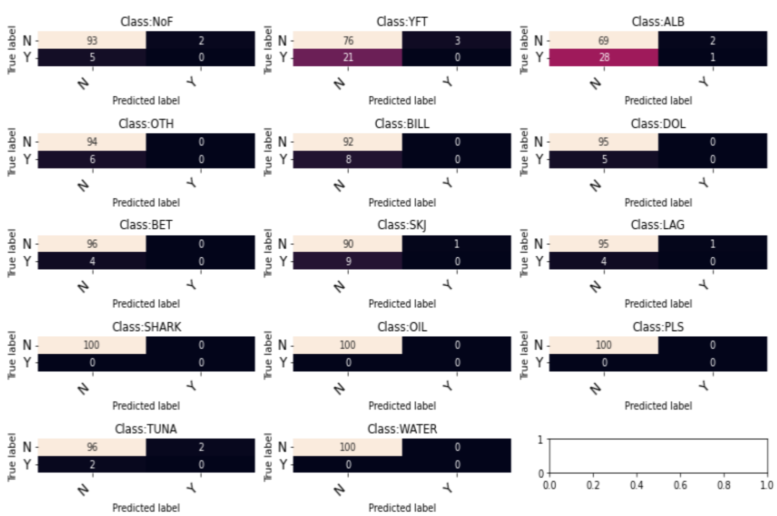
After much trial and tribulation, the decision was made to use transfer learning rather than a model from scratch. Utilizing the **MobileNetV2**, the model takes in the image of size (400, 400, 3) then uses the weights associated with the ImageNet database. We also utilize a 2D pooling. The model output layer has 14 neurons (one for each category) and uses *sigmoid* to normalize the values. When compiling the model, the **SGD** optimizer was used with a learning rate of 0.0001. In addition to the **SGD** optimizer we used the loss function called, "binary cross entropy" because if each category occurs it is either a zero or one (each category is encoded in binary dummy values).

Model Evaluation

The model is highly inaccurate. The function does give concrete results, but it has large issues with overfitting. The overfitting of this model quickly becomes quite extreme, so underfitting is the preferred option (and the option we went with). The graph below shows the increase in accuracy as the epochs increase and the decrease of loss which shows that additional epochs are not having a detrimental effect on the model.



Our proposed model is roughly 20% accurate. This needs to be improved for the model to be utilized. In order to see how our model fares in each specific category we created a confusion matrix. The multilabel confusion matrix compares the predicted values against the tested value and finds that the model is highly inaccurate in the categories that are present (which are few due to the sparsity).



We also used statistical measures to examine how the model's prediction ability fares. First, we examined the coverage error, which ideally should be close to 14 the number of categories, our coverage error, however, was 10.79. The other metric utilized was exact match ratio. The exact match ratio is the number of rows that much perfectly. The ratio is 22%, which is not great but shows that one fifth of all the images were classified 100% correctly.

Previous Solutions

In order to raise the accuracy without overfitting there were a variety of different solutions proposed. The first attempt was to create my own model without transfer modeling. The model predicted that every image had a YFT and a Human (a serious issue of overfitting returning a 98% accuracy). The next step was to use freezing layers, which also caused overfitting returning that each image had nothing in it. In addition to freezing layers, different learning rates were also tried, which once again brought the "no categories" present over fitting issue. The large issue with other solutions were overfitting, which predicted patterns based on the sparseness of the dataset rather than what was in the image.

Discussion and Results

Fish identification is a difficult task. Fish depending on their age can have different fin structures, coloring, and can have radically different sizes. On top of this even trained marine biologists cannot differentiate fish tuna species from each other without examining the fish's liver (i.e., yellowfin tuna and bigeye tuna). Using image labelling technology is not a silver bullet solution due to the variability in fish species. The proposed model's accuracy is a strong indicator about some of the challenges of image recognition and deep learning. One of the first questions the model raises is: Is image recognition right for this project? Arguably no, but image recognition can be used as a broad indicator to see catch number for maybe not fish species but maybe genus. A possible new project would be to differentiate between billfish (swordfish, marlin, spearfish) and tunas (bluefin Tuna, albacore, skipjack tuna). Another issue is the distribution of how often the classes appear. Most of the classes appear rarely in an image, which makes it difficult to train the model because an accurate model can predict an image saying that it has none of the classes in it and still be right a significant portion of the time. By having broader categories (i.e., by selecting by genus), the occurrence issue would be partially alleviated. In addition, we could weight the images so that there is more diversity in the training set, so the model can train more effectively. Using image recognition and multi-labelling does have merit to further protect the environment and our fisheries and would be best served with more research. Due to the limitations of resources (RAM limitations for one) and data this project faced major challenges. With more research and a reexamining of what the goal of our project is, a more accurate and helpful model is possible. However, this model serves as an indicator about the hardships of multi-labelling and serves as useful reminder that deep learning is not necessarily a silver bullet solution.