Mher Oganesyan

Melissa Cheung

Jasmine Cao

Adan Constanzo

Smitkumar Kaushikkumar Patel

Project Progress Report

The project is Humpback Whale Identification. We are provided a CSV file with the whale ID and the image file name. The images provided are of the whale’s tail fin; each image being a different resolution and a colored photo. We are to take the provided data to train a model and predict the species of whale by their fin. For the training images, we are provided over 25,000 images.

Adan is researching into algorithms that can help us with learning from the data since we are handling images. Some of our options are SVM, ANN, and possibly some other algorithms that we did not cover in class. Smit is looking into how we can work with the data since the images are different sizes, or at least what’s the best method in doing so. Two options are to either resize the images or to grab the most import features/information of the image as it comes in, if the latter is possible. In addition, since the images are in color, that will also affect the data processing; we need to see if we must take additional steps to handle that. Jasmine is looking into the output that is expected for the Kaggle submission. It seems as if they want five predictions for the output, therefore we must figure out what is needed to be done to get to that output. Melissa and Mher are to put together all the gather information with the rest of the group to write the code and analyze the results.

We do not have any results thus far, but we are putting together code very similar to the latest homework to see what kind of results we can get. Based on those results, we can figure out what needs to be done to get better results.