# CSCI 481/597J Project Pre-Proposal

Jonah Douglas, Ethan Lindell, Archan Rupela

May 6, 2020

1. We are trying to solve bird identification using a data set of 180 birds. In our personal experience it seems common for people to have a short-term interest in identifying birds around them. While on a walk, hike, or viewing from their house they might see a bird they want to know more about. The problem is that few go through the effort of identifying the bird through a field guide or existing online resource. Being able to take a picture of the bird to identify it and then researching further would improve availability of education and ease of investigation.

2. As taken from the Kaggle Description of the Data, the data set is of 180 bird species. There are "24497 training images, 900 test images(5 per species) and 900 validation images(5 per species. All images are 224 X 224 X 3 color images in jpg format"(Kaggle). It is noted that there are at least 100 training image files for each species but overall there is a varying number of files per species, meaning the training set is not balanced. The Description notes that "One significant imbalance in the data set is the ratio of male species images to female species images. About 80 percent of the images are of the male and 20 percent of the female. Males typical are far more diversely colored while the females of a species are typically bland...Almost all test and validation images are taken from the male of the species. Consequently the classifier may not perform as well on female specie images." (Kaggle). A flaw we identified is that there is not a label for the location of the bird. This could significantly improve the model because a tag for location could train a more accurate model, for example if there are two similar birds but they are found in entirely different locations, it would be a distinct feature to set the birds apart.

3. The data set on Kaggle is rather restrictive to the addressed problem above. A variant of the problem could be identifying whether a bird is a male or female. Given a different data-set, one could also attempt another type of species identification on plants, flowers, or another animal. These types of problems seem more difficult in the variance of flowers and plants throughout the season. As for other animals, identifying bird species might be easier in effort because there are more distinct species, unlike dogs who breed between species. Additionally, people typically will be less likely to be able to identify birds than dogs.

4. So far the project team consists of three members, Jonah Douglas, Ethan Lindell, and Archan Rupela. In order to meet the 4-5 person team size, yes, we would like help finding members. Thanks!