Game Camera Image Analysis Using a CNN

Camera traps, also known as game cameras, are an important tool for monitoring wildlife populations for conservation and management purposes. The cameras allow for wildlife data to be collected with minimal human disturbance, yielding more accurate data for species presence and interactions. Camera traps produce an enormous number of images, and sifting through these images to identify individual species can be time consuming and tedious. Training machine learning models to identify wildlife in camera trap images can therefore save companies an enormous amount of time and resources. The general focus of this project was to use deep learning to train a convolutional neural network (CNN) model that can accurately predict the animal species in a camera trap image.

Data Wrangling

While camera traps are used for all types of wildlife, monitoring bird populations can be particularly challenging, since they are more difficult to capture and track than other terrestrial animals. I used a Kaggle dataset containing images of 16 common Himalayan birds. There was very little meta-data associated with these images, and images were all properly labeled and sorted, so this step was minimal.

Exploratory Data Analysis and Preprocessing

I examined sample images in each class and determined that there was a large degree of variation in the types of bird photos caught by the camera traps. The pictures varied in the complexity of the background vegetation, the pose/angle of the bird and how obscured the bird was by vegetation, as well as bird size and distance from the camera. Some images also contained multiple birds. Given the limited size of the dataset (149 images in the training dataset), the variation in the images posed challenges, especially for a 16-class image identification problem.

I scaled all the images to a smaller size, and used ImageDataGenerator for preprocessing images, which augments the data by randomly shifting, rotating, and flipping the images.

Figure 1: A sample of the images for two different species, illustrating how variability in the image data.

Model Selection

I used a convolutional neural network (CNN) model for analyzing and classifying bird images into 16 classes. CNN models are a common type of deep learning model used for classifying images. I tried out the following models and the assessed model accuracy and validation loss metric for each:

Model 1: CNN model
Model 2: VGG16 pre-trained CNN model

The CNN model performed very poorly on the validation and test data (even after going from a base model with 5 layers to a model with 11 layers). I then fit the pre-trained VGG16 model, which improved training and validation accuracy by over 30%, but was still prone to overfitting due to the small size of the dataset (Figure 2 below). It also still performed poorly (5% accuracy) on the test data. The categorical cross-entropy loss function measures the dissimilarity between the predicted class probabilities (obtained from softmax activation function) and the true class labels, and is commonly used for multi-class classification problems.
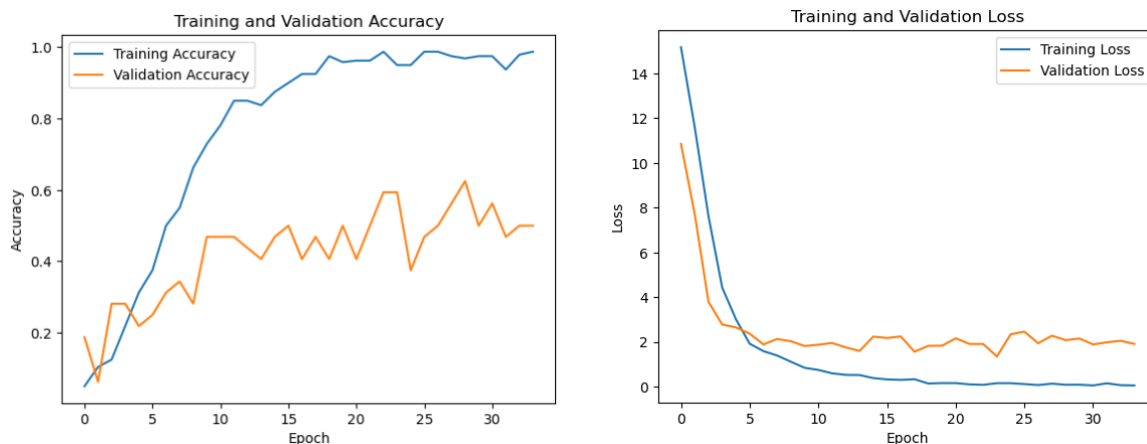
Figure 2: Model accuracy and loss function across epochs. While the loss function decreases towards 0, the training model accuracy climbs steeply towards a value of 1, suggesting overfitting.

Summary and Future Research:

This was a good exercise in image classification and gaining and understanding of CNN models. The main points I learned from this project were:

1. The architecture behind constructing a CNN model, and the increase in model accuracy offered by using a pre-trained model such as VGG16 on image data.
2. The importance of having a large dataset for good results, especially for image classification. While this was an interesting learning exercise, it ultimately did not lead

to good model results because of the limited size of the dataset, especially for a multi-class classification problem.

3. Next steps: I am interested in further exploring other pre-trained models such as the ViT transformer model for image data, but on a larger image dataset.