Udacity Machine Learning Nanodegree

Capstone Proposal - Dog Breed Classifier

Brett Schwarz

December 2020

Domain Background

Machine learning has been around for many years. The phrase was first coined by Arthur Samuel in 1952[1]. However, it hasn't gained much traction until the last decade. This is due to advances in computing power, more sophisticated techniques, more data, and the explosion of use cases.

One of the most exciting use cases for machine learning is image identification and classification. This has many applications and has improved steadily over the last decade. There are many different uses within this domain, including:

- Facial recognition in social media. For example, Facebook's facial recognition.
- Stock photo websites utilize image recognition for categorizing and searching purposes.
- Image searching for finding similar images for businesses. For example, businesses can integrate visual search[2] to enhance their product or service.
- Marketing and advertising campaigns.
- Automation of tasks. For example, Google's address recognition for maps.

[1] https://www.dataversity.net/a-brief-history-of-machine-learning/
[2] https://imagga.com/solutions/visual-search

Machine learning has always fascinated me because it seemed like magic to me. Especially image recognition. Facial recognition was first developed by the U.S. government in the mid-1960s[3]. Before studying Machine Learning, I didn't really understand how this was possible. However, even now that I understand what is involved to implement image recognition, it still fascinates me.

This project will focus on classifying dog breeds, based on input photos. It will also be able to identify if the photo is of a human, and if so, will match the human to the closest dog breed.

One of the most popular algorithms used for image recognition is Convolutional Neural Networks (CNN). CNNs borrowed from nature on how they process images but using a neural network to process the data. The CNN can trace its roots back to the neocognition, which was proposed by Kunihiko Fukushima in 1980. This work was inspired by work from Hubel and Wiesel on the study of vision in mammals[4].

Problem Statement

The specific problem that this project is to process user supplied images. It first must detect if the image has a dog or human in it. If there is a dog in the image, then it will determine which breed of dog it belongs to. If there is a human in the image, then the closest dog breed matching the human will be picked.

Datasets and Inputs

The data sets are provided by Udacity. They fall into two categories:

1. Dog images - there are a total of 8351 dog images. There are datasets for train, test and validate, with each having 133 different dog breeds. The train dataset has roughly 20 images per dog breed, and seems to be balanced.
2. Human images - there are a total of 13233 human images. There are 5749 different people represented in those pictures. Some people have multiple images, and others just have one. So, this dataset is imbalanced.

Both data sets are needed, since we want to detect both dogs and humans in pictures, so we need enough data to train the model. There will be a total of 133 breeds to detect from.

---

[3]https://www.bloomberg.com/quicktake/facial-recognition#:~:text=Facial%20recognition%20technology%20was%20first,intelligence%20agencies%20and%20the%20military.
[4] https://en.wikipedia.org/wiki/Convolutional_neural_network

Example of a Human Image (Gwyneth Paltrow) and a Dog Image (St Bernard).

Solution Statement

The solution is to accurately predict the dog breed of an image, if there is a dog in the image. If the image is of a human, then the solution is to match that human with the closest dog breed that resembles the human. A Convolutional Neural Network (CNN) will be used. For the human images, a haarcascades classifier will be used based on OpenCV's implementation[5]. For the dog images, a pre-trained VGG-16 model will be used[6]. Once the image has been identified, then the CNN model will be used to predict the breed that best matches.

Benchmark Model

The benchmark for this project will be the results from a CNN built from scratch. The results will be improved upon after the benchmark by using transfer learning, and tweaking the model (i.e. find optimal epochs).

Evaluation Metrics

The model will be evaluated based on the accuracy of the results. The expectation is that the accuracy will be low, even after transfer learning and tweaking, since predicting dog breeds is challenging (even for humans).

---

[5] http://docs.opencv.org/trunk/d7/d8b/tutorial_py_face_detection.html
[6] http://pytorch.org/docs/master/torchvision/models.html

Project Design

The first task will be to make sure the data has been preprocessed. For example, it is standard practice for image recognition to convert the images to grayscale. We then find detectors for both dogs and humans. As mentioned above, I will use the haarcascades to detect humans, and the pre-trained VGG-16 model to detect dogs. Then we are able to construct the CNN to predict images that are given to the model. I will try to improve upon the default model, by using transfer learning, and tweaking other parameters such as the number epochs. The goal is to have a working model that has a higher accuracy score than the default model.