Capstone Project Proposal

Dog Breed Classifier
Machine Learning Engineer Nanodegree

Muhammad Ovais **3rd Dec, 2020**

Domain Background

Visual classification of objects is one of the most useful applications of machine leaning. With the modern advancements in deep learning over the last eight years, this domain has been instrumental in enabling computers to see and identify all kinds of objects pretty much like humans do, solving a wide variety of problems in our lives.

One such problem is identifying any particular dog's breed. The classification is fairly challenging due to the significant overlap of features found in different dog breeds. As a result, even experienced humans sometimes fail to differentiate between certain breeds, e.g. a Brittany and a Welsh springer spaniel.

Since the start of this decade, researchers have been using this particular problem as the benchmark for the evaluation of their novel multi-classification approaches. (Liu, et al. 2012) succeeded to achieve 67% recognition rate on this dataset using their Part Localization Technique. Another faster approach on a smaller dataset was proposed in (Prasong and Chamnongthai 2012) using the size and position of key features in images and applying PCA. Our approach, however, will be different as we will be taking advantage of deep neural networks to achieve a more accurate, yet computationally intensive solution.

Problem Statement

The goal is to make such a CNN-based classifier that performs three main tasks:

- ***Detect Humans:*** It should be able to detect a human's face when given an image of a human.
- ***Detect Dogs:*** If a dog's image is provided, it should be able identify it as a dog.
- ***Detect Dog breeds:*** Moreover, it should be able to accurately identify a dog's breed when a dog's face is detected. In case a human's face is detected, it should output the dog's breed closest to the human's face in resemblance.

Datasets and Inputs

The datasets can be obtained from following links: Human dataset, Dog dataset.

The human dataset contains 13233 total images divided into 5750 folders each named after the name of corresponding person.

The dog dataset consists of 8351 total images divided into 3 folders – train (6680), test (836) and validation (835) images. Each of these folders is further divided into 133 subfolders named after respective dog breeds.

An important insight to remember is that none of the above datasets are balanced—containing varying number of images for different dog breeds and humans.

Solution Statement

For the detection of a dog face in an image, we will be using a Convolutional Neural Network (**VGG-16**). A convolutional neural network, or CNN, is a deep learning neural network designed for processing structured arrays of data such as images. Convolutional neural networks are widely used in computer vision and have become the state of the art for many visual applications such as image classification. [1]

For human face detection, we will be using OpenCV's pretrained **Haar cascade classifier**. Haar Cascade is a machine learning object detection algorithm used to identify objects in an image or video and based on the concept of features proposed by Paul Viola and Michael Jones in their paper **"Rapid Object Detection using a Boosted Cascade of Simple Features"** in 2001. [2]

Finally, in order to detect a dog's breed we will be trying two approaches—creating a CNN from scratch, and using the output features from the already trained dog detector as input to another CNN (transfer learning). The selection of a particular CNN architecture for the latter step will be based on the evaluation of multiple architectures (e.g. VGG-16, ResNet-50, ResNet-101, LeNet, etc.) in terms of accuracy, loss and training time.

Evaluation Metric

For a multiclass classification like this, conventional metrics like accuracy usually fail to represent the performance of the model to an acceptable extent. Therefore, the logarithmic loss (or Log-Loss) function, as defined by the following equation, will be used as the determinant evaluation metric for our models.

$$F = -\frac{1}{N}\sum_{i}^{N}\sum_{j}^{M} y_{ij} \cdot Ln(p_{ij})) = \sum_{j}^{M}\left(-\frac{1}{N}\sum_{i}^{N} y_{ij} \cdot Ln(p_{ij}))\right)$$

Benchmark Model

The performance benchmarks for our final models are as following:

---

[1] Convolutional Neural Network (https://deepai.org/machine-learning-glossary-and-terms/convolutional-neural-network)

[2] Deep Learning Haar Cascade Explained (http://www.willberger.org/cascade-haar-explained)

- The CNN built from scratch for the dog-breed must have at least 10% accuracy. This is to satisfy the condition that the built model is in fact performing classification based on detected features and not just giving results at random, because the probability of a random guess being correct 1 in 133 (0.75%).
- The result of our final model built using transfer learning must have at least 60% accuracy. However we will try to attain an accuracy of 85% and above.

Project Design

The design workflow of the project can be divided into following steps:

1. Import the necessary libraries and datasets. Preprocess and split the dataset into train, test and validation data. Analyze the respective datasets and decide if augmentation could be helpful.
2. Create human face detector using Haar Cascade classifier.
3. Create dog detector using a pretrained CNN architecture (e.g. VGG-16).
4. Create dog-breed CNN classifier from scratch as well as using transfer learning evaluating different possible architectures.
5. Use the above classifiers in writing an algorithm that – if a dog is detected, output the predicted breed, and if a human is detected, returns the resembling dog breed.

References

Liu, J, A Kanazawa, D Jacobs, and P Belhumeur. "Dog Breed Classification Using Part Localization." *Computer Vision – ECCV 2012.* Florence, Italy: Springer-Verlag Berlin Heidelberg, 2012.

Prasong, P, and K Chamnongthai. "Face-recognition-based dog-breed classification using size and position of each local part, and PCA." *9th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology.* Phetchaburi, 2012. 1-5.