Udacity Machine Learning Engineer Nanodegree Capstone Proposal

Taylor Isbell
Nov. 13, 2020

Project Background

The background of the project described in this proposal pertains to the task of image classification via computer vision. Computer vision is an active area of research which involves developing algorithms that give computers the ability to extract information from imagery. This problem is one that we, as humans, solve almost constantly through our vision, but is surprisingly non-trivial to teach a machine how to solve. Computer vision is an extremely powerful technology and is currently being used in things like Tesla's self-driving car technology, Apple's FaceID, Google Translate, and many others.

Problem Statement

The goal of this project is to develop a model that can correctly predict the breed of a dog given only an image. The model should be able to accept an image of a dog OR a person, and it will then decide: 1. Whether the image contains a dog or a person, and 2. what dog breed the dog OR person most closely resembles.

Datasets and Inputs

The data needed for this project has already been collected and organized by Udacity. For the human images, the Labelled Faces in the Wild (LFW) dataset will be used. For the dog images, the Stanford Dogs dataset will be used. Both datasets are benchmark datasets that have been specifically curated, edited, and labeled for the use of training and testing computer vision models.

Solution Statement

The solution to this problem should be an end-to-end data pipeline that is able to take in an unprocessed image of a dog or human, perform the necessary transformations, feed the transformed image into the model(s), and return an inference which predicts "human" or "dog" *and* the dog breed that the human or dog most closely resembles.

Benchmark Model

This blog post very clearly details the use of a deep learning model to classify images of handwritten digits. The dataset used in this benchmark model (MNIST) is a standard dataset used widely in computer vision and deep learning. It is a dataset containing 60,000 small square grayscale images of handwritten single digits ranging between 0 and 9. The goal of this model is to be able to correctly classify these

images by the digit that is written. Because this dataset is so widely used and understood, there are dozens of documented models, including this one, showing accuracy of 98-99% on the test dataset.

Evaluation Metrics

The primary metric used in this project will be accuracy. The definition of accuracy for the sake of this problem is defined below:

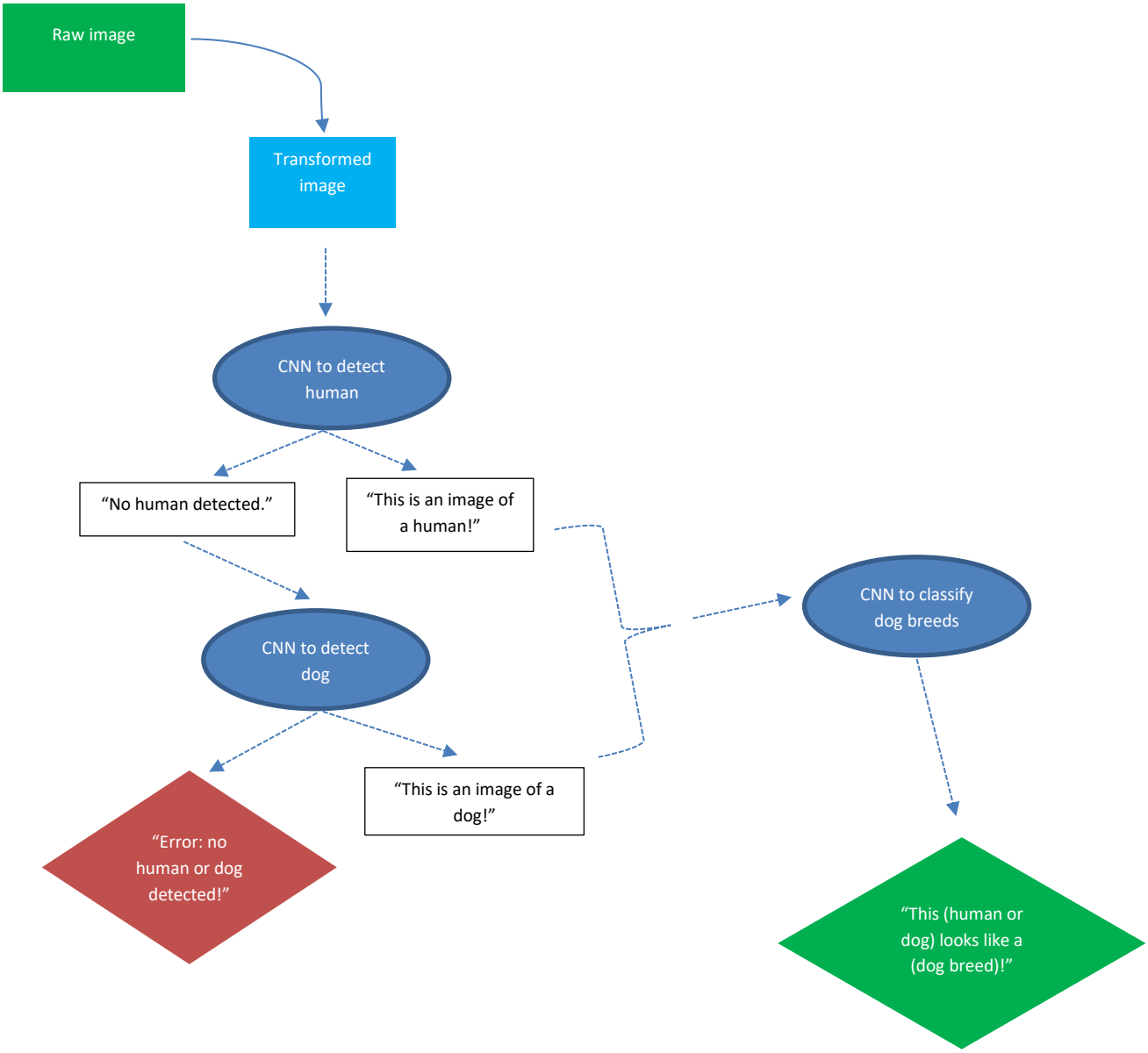
$$Accuracy = \frac{Images\ correctly\ classified}{Total\ images}$$

The details of the models used in this project will be discussed below, but in summary, there will be models developed to accomplish two different tasks: binary classification (i.e. human vs dog detection) and multi-class classification (i.e. inferring the correct dog breed). Since the former task only has two options, it is expected that the accuracy of those model(s) will be much higher than the multi-class model(s). In fact, the dog breed dataset includes images of 133 different dog breeds, which means that random guessing would result in less than 1% accuracy (as opposed to 50% for binary classification)

Project Design

The specific model type used for this problem will likely be a convolutional neural network (CNN). A CNN is a special kind of neural network that uses computational layers known as “convolutions” that reduce the complexity of spatially organized data like imagery. Because of this, these models are ideal for the task at hand. To accomplish the solution described above, there will likely need to be several CNN models used in sequence or parallel, each of which dedicated to a single task. For example, there will likely need to be one model for detecting human faces, one model for detecting dog faces, and one model for the actual dog breed classification. Therefore, each of these models will need to be evaluated separately before implementing the final data pipeline. And not only that, but the detection models will need to be tested on both the dog and human datasets. In total, there will be 5 accuracy evaluation steps in this project (dog detection model tested on dog and human data, human detection model tested on dog and human data, and the dog breed classifier model tested on dog data).

After the models detailed above are trained and tested, the data pipeline solution will then be built out so that it can be deployed to a web app or some other interface. The diagram below is a basic and preliminary flowchart for how the solution might be laid out.


Raw image
Transformed image
CNN to detect human
"No human detected."
"This is an image of a human!"
CNN to detect dog
"This is an image of a dog!"
"Error: no human or dog detected!"
CNN to classify dog breeds
"This (human or dog) looks like a (dog breed)!"