Capstone Project Proposal

Guilherme Baldo Carlos, January 19th, 2022

Domain Background

This project deals with image classification which is one of the fundamental tasks of computer vision. In recent years this task has been successfully solved by the usage of CNNs. CNNs are neural networks that use filters and the convolution operation to automatically extract relevant features of images. Those features are fed to fully connected networks in order to perform classification. CNNs may be trained from scratch but its results may be boosted by using transfer learning. The idea behind transfer learning is to get a model previously trained in a large dataset and continue to train it on the dataset you want to perform classification. That works because computer vision tasks usually rely on extracting common features from images like detecting edges, lines, curves and so on. So a model that is able to do that for one dataset may be a good starting point to be trained for the target task. So this project explored image classification, building and training models from scratch, using prebuilt models and taking advantage of transfer learning and also building pipelines of models to perform a compound task.

Problem Statement

This project consists of using convolutional neural networks (CNNs) to produce an algorithm composed of a pipeline of CNNs. That algorithm identifies if an image provided by the user contains a person, a dog or neither. In the case of a dog, the algorithm responds with the dog breed. In the case of a person the algorithm responds with the dog breed the person in the picture most resembles. Finally in the case of neither a dog nor a person in the image the algorithm will respond with invalid input.

Datasets and Inputs

Two datasets are used to develop and evaluate the algorithm. First, to evaluate the face detector the LFW (Labeled Faces in the Wild) public dataset is used. This dataset contains 13 thousand images of people.
The second dataset is provided by Udacity and corresponds in a dataset with 133 dog breeds and 8351 dog images with varying image size. Those dog images are splitted in train set (6680 images), validation set (835 images) and test set (836 images).

Solution Statement

To solve the problem of human detection the face detector Haar Cascades provided in OpenCV is used.
To solve the problem of dog detection the VGG-16 model pretrained on ImageNet was used. This model is already able to detect a 1000 classes, so a function was built to binarize the output of this model from dog (class 151 to 278) and non dog (0 to 150 and 279 to 999).

To solve the dog breed classification per se two solutions are explored. The first solution consists of a model built and trained from scratch on the dog breed dataset following CNN design guidelines. That solution was built with inspiration in architectures from literature. The other solution was based on the selection of a state of the art model pretrained on ImageNet and fine tuned on the dog breed dataset.

Benchmark Model

The benchmark model that will be used to compare the solution built using transfer learning was the solution built from scratch. It is expected that the results achieved with transfer learning require fewer epochs for training and achieve better metrics.
There are imposed requirements for the accuracy of the models. The model built from scratch must achieve at least 10% accuracy and the model built with transfer learning must achieve at least 60% accuracy.

Evaluation Metrics

Since we are talking about classification, we want to measure how many of the human images the face detector is able to detect. How many dog images the dog detector is able to correctly predict. And finally how many dog breeds the classification model is able to predict correctly.
That metric is called accuracy and for classification purposes it is calculated as below:

$$accuracy = \#\ of\ correct\ predictions\ /\ \#\ of\ predictions$$

Project Design

The solution of the problem stated will be developed using Jupyter Notebook and the Pytorch framework. A notebook is already provided stating the guidelines for the problem development.
The diagram below illustrates the project design.

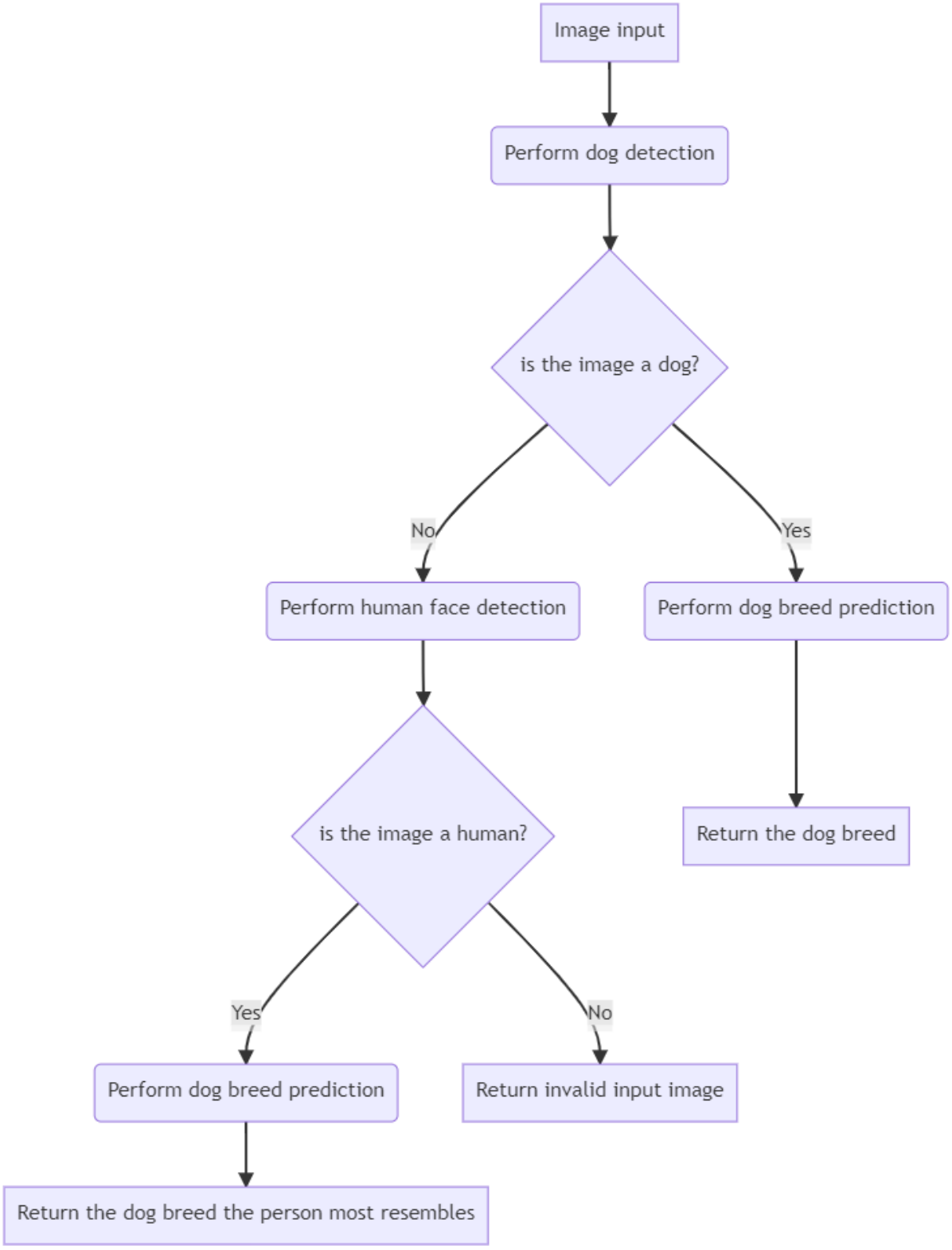
Image input
Perform dog detection
is the image a dog?
No
Yes
Perform human face detection
Perform dog breed prediction
is the image a human?
Return the dog breed
Yes
No
Perform dog breed prediction
Return invalid input image
Return the dog breed the person most resembles