Machine Learning Engineer Nanodegree
Capstone Project

BY

Khadija Mahmoud Taha

February 2021

1. Definition

Project Overview:

The research area focuses on providing the computer with the ability to distinguish between things as if it has an eye like human beings.

In the social media field, there is a try to generate an automatic description of images. This generation is done by collecting each description of each image, then to use these descriptions to predict the description of further images.

Here we are combining these two aspects in this project, so this is a computer vision problem, and we are solving a dog breed classification problem

Problem Statement:

The problem is a multi-label classification problem. The input data should be an image of a human or a dog.

The model is to predict which is the image of. If the image contains a dog, then predict its breed, and if it contains a human, then the model is to assign the most similar dog breed.

All the solution is to be done by the CNN algorithm with transfer learning. The reason why we chose this algorithm is this algorithm is flexible for data dimensionality increase as long as the user provide the model with images.

Metrics:

We will evaluate the model according to accuracy as the data set has balance in training data sets in dog and human categories.

$$accuracy = \frac{correct}{correct + incorrect}$$

The data set contains almost 60% of human images and almost 40% of dog images, and having a huge number of dog breeds in comparison with total data set size makes us expect low accuracy in predicting the true breed.

II. Analysis

Data Exploration:

Dog images dataset:

The dog image dataset has 8351 total images which are sorted into train (6,680 Images), test (836 Images) and valid (835 Images) directories. Each of this directory (train, test, valid) have 133 folders corresponding to dog breeds. The images are of different sizes and different backgrounds, some images are not full-sized. The data is not balanced because the number of images provided for each breed varies. Few have 4 images while some have 8 images.

Human data set :

The human dataset contains 13233 total human images which are sorted by names of human (5750 folders). All images are of size 250x250. Images have different background and different angles. The data is not balanced because we have 1 image for some people and many images for some.

Combing between these two data sets we have almost equally divided into human and dogs categories.

1) The training data set contains 6680 images.
2) The validation data set contains 835 images.
3) The test data set contain 836 images.
4) Dog breed classes of 133 classes.
Each image is colored with RGB color system.
The data set is almost equally divided into human and dogs categories.

Exploratory Visualization:

It’s tricky to a human to distinguish between similar breeds as followed:

In addition to lack of balance in images assigned to each breed.

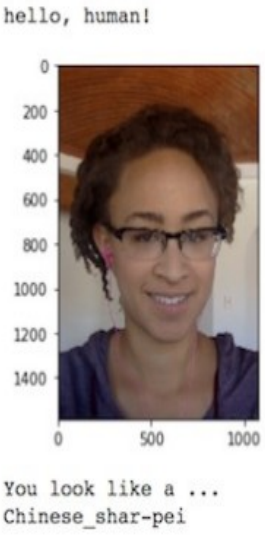
There is another challenge of having same breed but with different colors as followed:

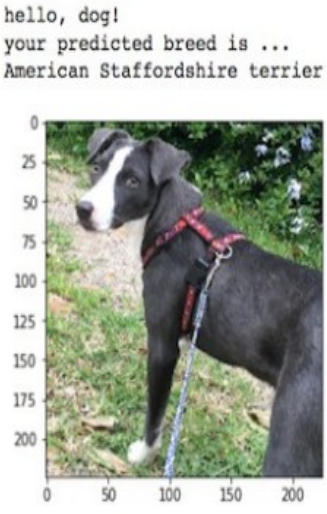
Yellow Labrador

Chocolate Labrador

Black Labrador

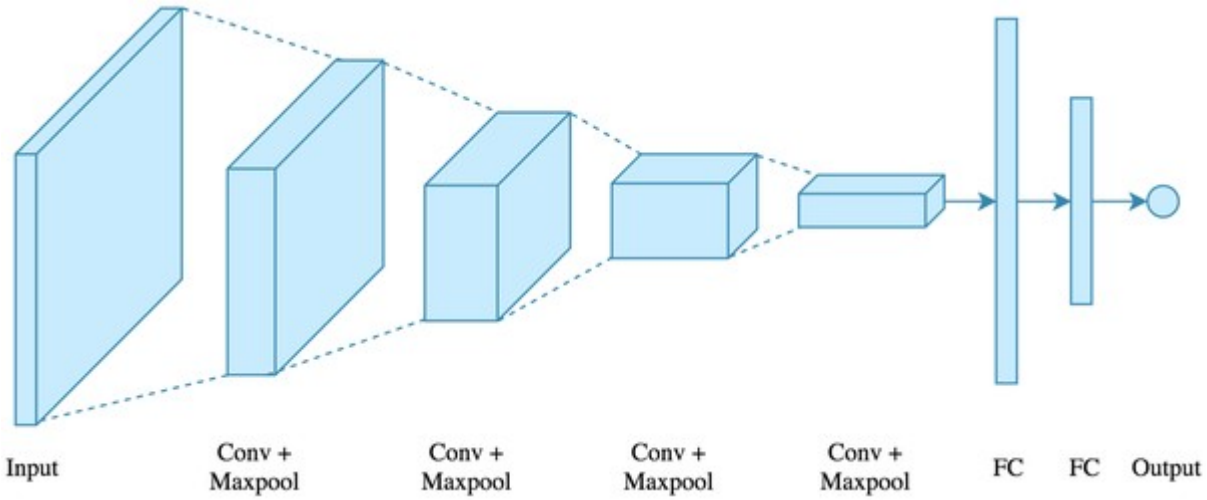
Data samples and Expected output:



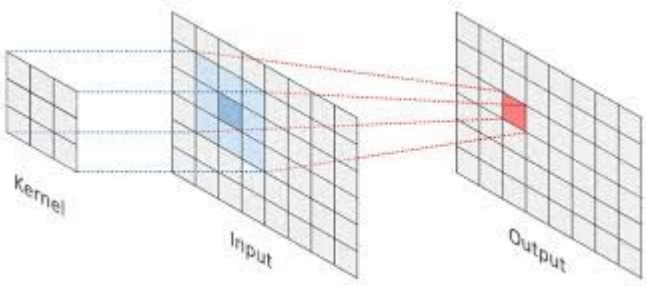


Algorithms and Techniques:

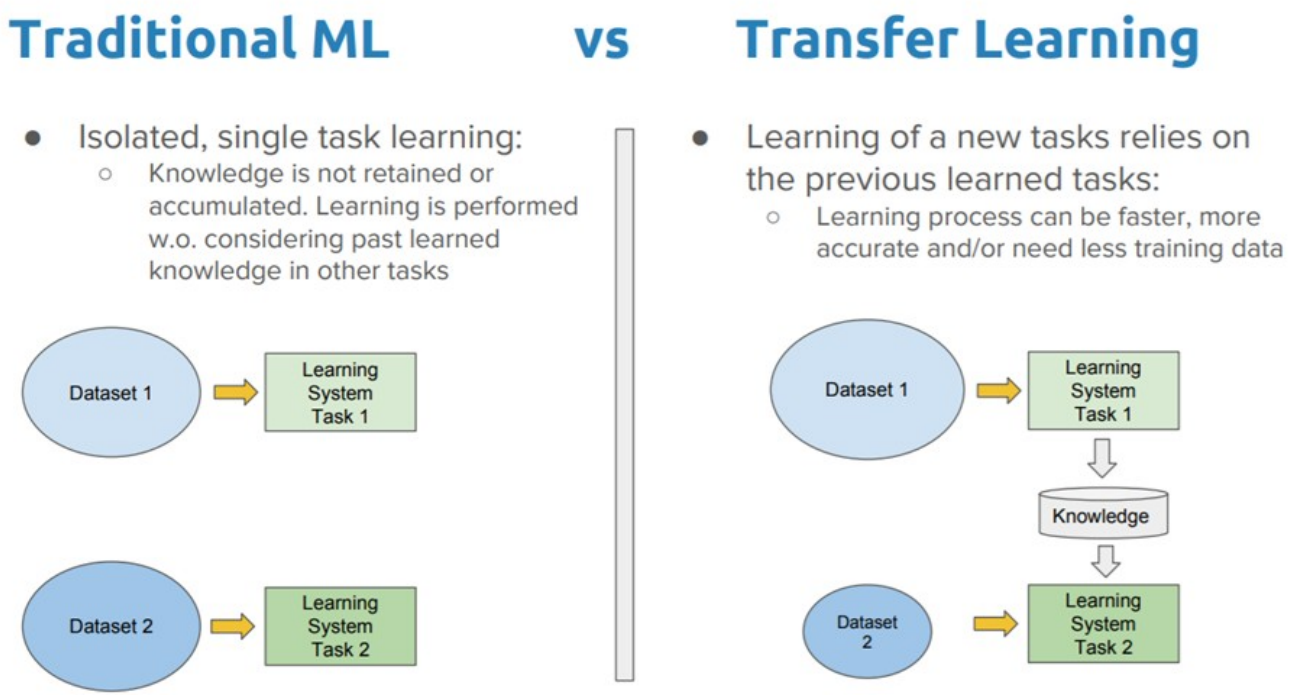
It’s a multi-label classification problem to be solved by the CNN algorithm.
This is how CNN works.



The figure below explains how convolution reduces the image dimensions



In addition to this algorithm is adaptable for high data dimensionality as it increases when the user inputs several images.
The transfer learnind did a great job rather than the CNN scratch model, here is the difference:



Steps of algorithm:

Firstly, to classify the image to a human or a dog by a pre-trained model.
In the case of a dog, the model is to classify it to its correct breed.
In the case of a human, the model is to find the closest breed similar to him as follow:

- if a dog is detected in the image, return the predicted breed.
- Else if a human s detected in the image, return the resembling dog breed.
- Else provide output that indicates an error.

Benchmark:

We chose a CNN model implemented from scratch.
Model architecture and data pre-processing is to:

- Convert the input image to a gray scale
- the gray scaled image is to be fed to the model with 224X224 dimensionality.
- This image passes to 3 convolution layers.
- The 3 hidden layers (with relu)are to be followed by max pooling layer.
- The image is to be flattened.
- pass the image with 2 fully connected layers(with relu).
- Last step is to drop out the neural network with 25% to avoid overfitting

III. Methodology

Data Preprocessing:

- Convert the input image to a gray scale
- the gray scaled image is to be fed to the model with 224X224 dimensionality.
- The image is to croped in the center.
- Last to assign the image to normalization stage.

Implementation:

- The Pre-trained VGG16 model is to extract features.
- The VGG16 model to be loaded as transfer learning.
- Retrain the last fully connected layer by defining a Linear layer with apropriate values for the number of input layers (224) with 133 final nodes as our output.

Refinement:

- Increase data set
- Increase number of epoches.

  Extra image processing stages and compare if it results in improvement.

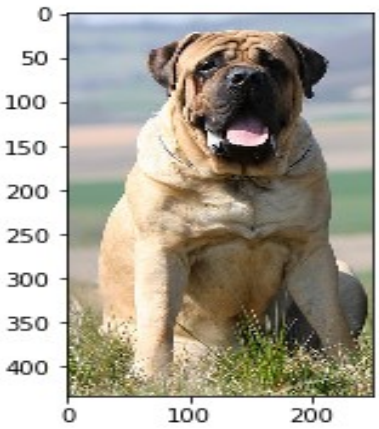
IV. Results

Model Evaluation and Validation:

The model built with transfer learning results is 86%.

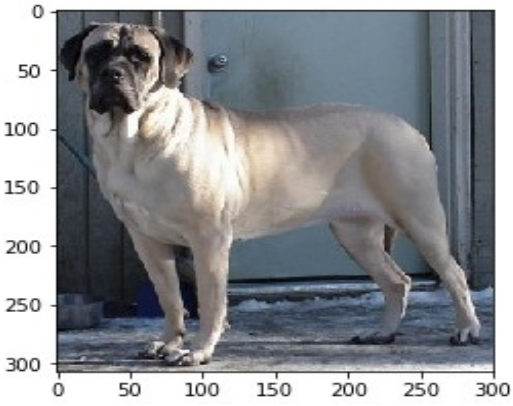
Dog input result:

hello, it is a dog!



Its breed is: Bullmastiff

hello, it is a dog!



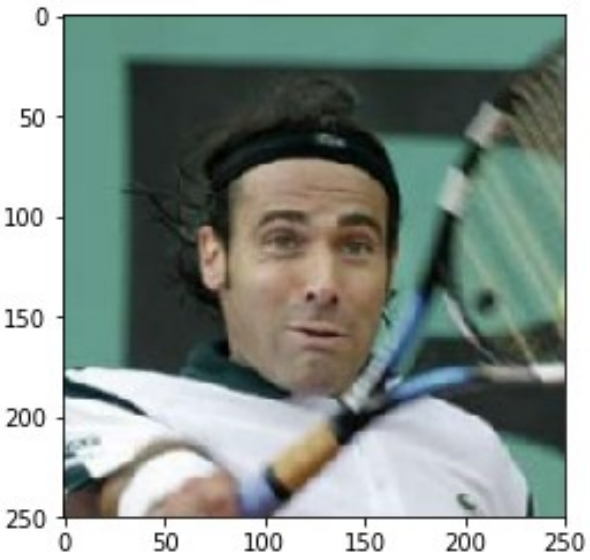
Its breed is: Bullmastiff

hello, it is a dog!

Its breed is: Bullmastiff

Human input result:

```
hello, human!
```




```
You look like a  Dachshund
```

```
hello, human!
```


```
You look like a  Beagle
```

Justification:

Comparing to the CNN model from scratch result which is only 10% , we are satisfied with the model which is with model transfer.