Macine Learning Engineere Nanodegree Capstone Project Proposal: Dog Breed Classifier

Ioannis Polymenis

23/12/2020

1. Domain Background

The concept of Machine Learning originated in the 1950s when the Computer Science pioneer Alan Turing proposed the question "Can machines think?" in the paper "Computing Machinery and Intelligence" (Turing 1950) and introduced the famous "Turing test" which explains how an intelligent machine should perform and other concepts that give life to modern Artificial Intelligence (AI) and Machine Learning (ML). The general concept that Turing proposes is how a general purpose computer can learn to perform a specific task on its own. In classical computer programming, the human gives instructions (the program) and data to be processed to the computer and the output is the answer to the problem. In contrast, with machine learning humans give the data and the answers to the computer and the outputs are the rules that the computer developed during the process, and these rules can be further applied to new data to generate novel answers.

The developments in machine learning as well as in deep learning is contineusly evolving espetially the during the last decade in areas of image recognition, face recognition, speech recognition, natural language processing, real-time translation and even the first intelligent system, with Google's Brain project of AlphaGo, which beat a human professional player in the game of Go without human intervention, in 2017 (Silver et al. 2017). Furthermore, in the image classification competition organised by ImageNet, the ILSVRC (ImageNet Large Scale Visual Recognition Challenge), the top five image classification error has been decreasing since its inception in 2009 (Jia Deng et al. 2009) and reached an astonishing 2.3% which is well below the human performance of around 5% error rate (Hu, Shen, and Sun 2018).

For a human image classification and object classification is part of our nature, we can just classify and recognise objects with ease. For computers on the other hand is not so straight forward this process. Thus, machine learning models, and particularly deep learning models, can be used to solve the problem of how a computer can see.Image classification is one of the main domains of deep learning where the deep neural networks play the most noticable role. In a broad sence, image clssification is the ability of a computer to analyze a given image and identify the objects of interest which is refered as a **"class"**. Therefore, a class is simple reffered to the object's label which for instance can be a "car", "annimal", a "person", and so on. In the current project, a "Dog Breed Classifier" will be developed since is an exelent oportunity to create classifier which will solve a such challenging task, because even for humans it is difficult to distinct one breed from another.

The complex task of dog breed classification will be accomplished using Convolutional Neural Networks (CNNs), which have been used since the 1990s in applications such as character recognition and are inspired by the brain's visual system (LeCun et al. 1998; Voulodimos et al. 2018), but their real advantages and benefits for computer vision started with the AlexNet in 2012 in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC).

2. Problem Statement

The aim of the current project is to create a pipeline image classification for a variety of dog breeds. A deep learning model will be used to distinguish the different breeds when a picture of a dog is given by the user. The first step will be to identify whether the given image is a dog or not, and at the next step should identify the correct dog breed. Furthermore, the model should be able to exhibit a reasonable performance on predicting mixed dog breeds. The final step in the model will be to detect whether a dog or human is in the given image, classify the dog breed and classify the dog breed that the human resembles. This is a supervised learning problem, and the learning method is based on labeled training examples and makes predictions on unseen data. Specifically, the training dataset has been appropriately labeled, for instance, in classification between different dog breeds and humans, and the algorithm is training on that dataset.

3. Datasets and Inputs

The datasets that will be used in this project are provided by Udacity through the GitHub repository. Firstly, the dog dataset contains images of 133 different breeds and is splitted to training, test, and validation sets totaling of 8351 dog images. The training set constitutes the largest portion of the entire dataset which is 6680 images, the test set has 836 images and the rest 835 images are in the validation set. Additionally, there is a human dataset which contains images of a variety of people and the total size of the dataset is 13233 images, and these images are used to the performance of the Human Face Detector.

4. Solution Statement

The project aims to estimate as accurate as possible the breed of a dog for any given image using a CNN model. Therefore, the first step to the solution is to detect humans using the implementation of Haar feature-based cascade classifier of the OpenCV framework. The next step will use a ResNet-50 pre-trained model to detect dogs in any given image using the weights that have been trained o ImageNet dataset. The pre-trained dog detector should be able to check the indexes between 151 to 268 and include all categories from “Chihuahua” to “Mexican hairless”. The final step will be to built a CNN model from scratch to classify dog breeds and must achieve a test accuracy of at least 10%, and in a later step of the project transfer learning will be used to improve the accuracy. At the last stage the model should achieve at least 60% of test accuracy.

5. Benchmark Model

For the model benchmark the pre-trained VGG-16 model with the trained weights of ImageNet will be used. This model uses transfer learning which reduce the training time without sacrificing the accuracy of the model. The pre-trained VGG-16 model is typically a saved network that has previously trained on a large scale image classification task, which allows the original network to effectively learn generic features. These features are used in the pre-trained models which have been customized to a given problem, and in this project is the dog breed classifier. In the present work the pre-trained model should be able to check the indexes between 151 and 268 and return the categories from “Chihuahua” to “Mexican hairless” as mentioned in the solution statement.

6. Evaluation Metrics

The current project has as an objective to compere the performance of the model created from scratch with the one used for benchmarking. Thus, the accuracy of the CNN model from scratch can be compared with the pre-trained VGG-16 model for the evaluation metrics.

7. Project Design

The workflow to approach the solution for the Dog Breed classification problem is as follows:

Step 0: Import Datasets

The datasets, as described in section 3, containing the different dog breeds and the humans, as well as the necessary libraries will be imported.

Step 1: Detect Humans

In this step the implementation of the Haar feature-based cascade classifier of the OpenCV framework will be used to detect human images from the dataset.

Step 2: Detect Dogs

The second step will use the pre-trained ResNet-50 model to detect dogs from the given dataset.

Step 3: Create a CNN model to classify dog breeds from scratch

The step 3 includes the designing of a CNN model from scratch. At this stage the model should achieve accuracy pf 10% or more, using the train, validation and test sets.

Step 4: Create a CNN model to classify dog breeds using transfer learning

At this step the CNN model will be used to classify the dog breeds with transfer learning using one of the following architectures:

- VGG-19
- ResNet-50
- Inception
- Xception

Step 5: Write the algorithm

The step 5 the algorithm of the model will be developed and will return the predicted dog breed when a dog is detected. In the case that a human is detected, then should resemble the dog breed. Lastly, in the case where neither a human or dog is detected an error message will occur.

Step 6: Test the algorithm

The last step includes the testing of the algorithm with custom images (my own images).

References

Hu, Jie, Li Shen, and Gang Sun. 2018. "Squeeze-and-Excitation Networks." In *Proceedings of the Ieee Computer Society Conference on Computer Vision and Pattern Recognition*, 7132–41. https://doi.org/10.1109/CVPR.2018.00745.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. "ImageNet: A large-scale hierarchical image database." In, 248–55. https://doi.org/10.1109/cvprw.2009.5206848.

LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE* 86 (11): 2278–2323. https://doi.org/10.1109/5.726791.

Silver, David, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, et al. 2017. "Mastering the game of Go without human knowledge." *Nature* 550 (7676): 354–59. https://doi.org/10.1038/nature24270.

Turing, A. M. 1950. "Computing machinery and intelligence." In *Machine Intelligence: Perspectives on the Computational Model*, 1–28. Taylor; Francis. https://doi.org/10.1093/mind/lix.236.433.

Voulodimos, Athanasios, Nikolaos Doulamis, Anastasios Doulamis, and Eftychios Protopapadakis. 2018. "Deep Learning for Computer Vision: A Brief Review." *Computational Intelligence and Neuroscience* 2018. https://doi.org/10.1155/2018/7068349.