Capstone Project

Dog Breed Classifier

By Gaurav Pal

I. Definition

Project Overview

The origins of dogs date back thousands of years, having evolved as domesticated descendants of the wolf, whereas modern dog breeds date back to the late 19th century. Prior to the Victorian era, there were different types of dogs that were defined by their function. Many different terms were used to describe dogs, such as breed, strain, type, kind, and variety. Dogs are the most variable mammal on earth, with artificial selection producing around 450 globally recognized dog breeds. Breed classification remains a relevant problem for the dog owners who searches for a show-class puppy or for a working dog with specific traits. Deep learning/Computer vision algorithms are particularly useful in identifying the image pattern and thus can be used to classify the dog breeds from a picture after training the model with substantial amount of training images. This is a multi-class classification problem where we can use supervised machine learning to solve this problem.

Problem Statement

The goal of this project is to build a CNN model that can be used to identify dog breeds from use supplied images. This project covers following tasks:

Dog Face Detector – Identify the dog breed from supplied dog image.

Human Face Detector – Identify the resembling dog breed from the human image.

Metrics

Accuracy metric is used as an evaluation metrics to gauge the performance of both models. Accuracy is calculated as number of correct predictions divided by total number of predictions and return a percentage value bounded between 0 and 100 and provides a good measure of the model performance. The data is split into train, test and valid dataset. The model is trained using the train dataset. Test data is used to predict the model outcome and accuracy is calculated for unseen data.

II. Analysis

Data Exploration

Classifier uses images as an input. Dataset used in this project is provided by Udacity and it contains images of dogs and humans.

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.

Human images dataset: The human dataset contains 13233 total human images which are sorted by names of human (5750 folders). All images are of size 250x250 with different background and angles.

The dog dataset is used for training, validation and testing of the CNN classifier. The human

Dataset is used for testing of the dog breed classifier in such case that it predicts the resembling breed from the human image.

Algorithms and techniques

Convolutional Neural Network (CNN) is used to address this multiclass image classification problem.

Convolutional Neural Network (CNN) is a type of deep learning algorithm which can take an image as input(tensors) and then differentiate the images from one another by assigning weights and biases to various aspects or objects in the image.

The solution has following steps to classify the provided images:

Step 1 - Detect human images, we can use existing algorithm like OpenCV's implementation of HAAR feature based cascade classifiers.

Step 2 - Detect dog-images we will use a pretrained VGG16 model.

Step 3 - Once the image is identified as dog/human, we can pass this image to an CNN which will process the image and predict the breed that matches the best out of 133 breeds.

Benchmark

The CNN model created from scratch must have accuracy of at least 10%. This can confirm that the model is working because a random guess will provide a correct answer roughly 1 in 133 times, so achieving a high accuracy using bare-CNN is not inherently possible.

The pre-trained model should have an accuracy of at least 60% so that it can be successfully used in a dog breed classifier app.

III. Methodology

Data Preprocessing

All the images are resized to 224*224, then normalization is applied to all images (train, valid and test datasets). Image augmentation is done to reduce overfitting. The train data, valid data and test images are randomly rotated and random horizontal flip is applied and then converted to tensor in order to pass to the model.

Implementation

CNN model with 3 layers is used to address the use case. All the convolutional layers have kernel size of 3 and stride 1. The input for first convolution layer is an image of 224 * 224 size with final layer producing output size of 128. Pooling layer of 2,2 is used to reduce the input size. Final layer gives us the output (one of 133 breeds based on the input data). A dropout of 0.25 is added to avoid over overfitting.

Refinement

The CNN created from scratch have accuracy of 11%. This meets the benchmarking, but the model performance can be further improved using a pre-trained model using transfer learning.

To implement this, VGG19 architecture which is pre-trained on ImageNet dataset, the architecture had top-1 error and top -5 error as 25.76 and 8.15, respectively. The model created as part of this project is added as the last layer of VGG19 model. The resulting model produces 133-dimension output to produce result of each category of dog breed as the dataset we used has 133 breed classes. This model performance in significantly better than the model that was created from scratch resulting in accuracy of 86 % satisfying the benchmarking criteria of 60%.

IV. Results

Model Evaluation and Validation

Human Face detector: The human face detector function was created using OpenCV's implementation of Haar cascade classifiers. 98% of human faces were detected in first 100 images of human face dataset and 17% of human faces detected in first 100 images of dog dataset.

Dog Face detector: The dog detector function was created using pre-trained VGG16 model. 100% of dog faces were detected in first 100 images of dog dataset and 0% of dog faces detected in first 100 images of human dataset.

CNN Model from scratch: Log loss was used to measure performance over different iterations. Benchmark model was trained over 25 epochs using stochastic gradient descent optimizer with learning rate of 0.05. On the test set, the model achieved 11% of accuracy.

CNN model with Vgg19: The implemented model was trained over 15 epochs, optimized by SGD with learning rate of 0.04 to reduce cross entropy loss function. On the test set, we achieved accuracy of 86% passing the benchmarking criteria.

Justification

The model created using transfer learning have an accuracy of 86% compared to the CNN model created from scratch which had only 11% accuracy.

Improvement

The model can be improved by adding more training and test data, currently the model is created using only 133 breeds. Performance can be further improved using a more complex neural network and varying the learning rates and epoch.

References

1. https://en.wikipedia.org/wiki/Dog_breed
2. Imagenet Documentation - https://github.com/pytorch/vision/blob/master/torchvision/models/vgg.py
3. PyTorch Documentation – https://pytorch.org/docs/stable/index.html
4. VGG19 documentation – https://pytorch.org/hub/pytorch_vision_vgg/