Udacity MLE project proposal: Dog Breed Classifier

Changlin Li
lcltopismine3@gmail.com

April 26, 2020

Abstract

In recent years, as the research of deep learning receives increasing attention, lots of image-based classification challenges have been resolved at higher accuracy. Although achieving satisfying performance, few literature shows how good can be made for customized dataset. To measure the performance of well-known models on customized dataset, I conduct an experiment to train a dog breed classifier on udacity provided dataset. Detailed report with a summary of performance evaluation will be provided after the proposal being approved by reviewers.

Keywords: *machine learning; deep learning; dog classification.*

1 Introduction

In the introduction I will provide a description for problem as well as current design to solve the problem I described above. I will start by offering a discussion for problem statement.

1.1 Problem Statement

Dog breed classification is a quite interesting issue where you are about to distinguish dogs from possible different breeds. As a supervised classification problem, you need to classify a sample to its corresponding category as accurate as possible. Usually, accuracy is a good measurement for binary classification problem.

And lots of general machine learning methodologies are applicable to classify images. Such as support vector machine with edge detector and random forest classifier. In this project, I will introduce deep learning classification network to see if it helps further improves classification performance.

1.2 Project design

To start, I will firstly present a start-from-stretch design as a benchmark model. It is a convolutional neural network with three blocks, where each block includes two convolutional layers and one maxpooling layer. To compare, I will also use transfer learning to train another model (VGG19 proposed) and fine tune it. The result of those two algorithms can be compared with each other. I will select the best performed one as the final model. Performance is measured in term of accuracy.

In the meanwhile, due to the limited computational resource available, I will transfer my code to Amazon web service (AWS) sagemaker instance. They have GPUs available to use, which is a great help to speed up training progress. After training progress finished on AWS, I will create endpoints and inference on the instance to evaluate performance of proposed model.

Since we are working on a customer dataset, it is necessary to conduct exploration on dataset to grab some insights. I'm planning to conduct research on dataset and primarily for 1). distribution of images. 2). number of class for dataset and how many "dog" categories are in the dataset.

2 Domain Background

Image classification is an important topic with already many improvements in the relevant field. Traditional computer vision methods provide some initial ideas, such as edge detectors that combined with support vector machine. Recently, as the advance of deep learning methods, convolutional neural network based framework provides much better results as their stronger ability to extract feature.

In Simonyan and Zisserman (2014), authors explore the possibly of adding more layers for better performance, which results in a deeper neural network (16 layers or 19 layers respectively). As observed from paper, a larger network depth is able to capture more context information, which in turn helps better classification performance.

In He, Zhang, Ren, and Sun (2015), authors push forward the idea of VGG Simonyan and Zisserman (2014) model by adding even more layers. However, direct summation of layers does not work better due to gradient vanishing issues. To counter, a residual connection is proposed. ResNetHe et al. (2015) family achieves even better results compared with VGG Simonyan and Zisserman (2014) models.

In this project, I select VGG19 as the transfer learning model for its strong ability of feature extraction. Hopefully I'm able to achieve a good performance.

3 Methodology

In this section we discuss the methodology I will implement for dog breed classifier. At the beginner, I will present the baseline method which is a simple convolutional neural network structure. Then I will introduce VGG19 model which is strong for its feature extraction ability and has a good potential for transfer learning.

3.1 Solution Statement

3.1.1 Benchmark Model

At the beginning, I will quickly discuss the baseline method for dog breed classifier. It is a simple convolutional neural network(CNN) structure with 3 blocks. For each block, two convolution layers with one maxpooling layers work together to extract feature from each image. To improve optimization efficiency, I also use a batchnorm layer after each maxpooling layer to adjust output distribution. For optimzer I'm planning to use SGD with learning rate of 0.001.

Here is a network structure for proposed simple CNN network.

3.1.2 Proposed solution: Transfer learning with VGG 19

As is often the case, it is hard to train a deep neural network without GPUs and good learning strategies. The whole process always takes tons of efforts. Hopefully, there is another way to go. For image classification, it is actually possible to introduce some pre-train weights and fine tune weights on our own dataset, which is called transfer learning.

As a comparison for baseline method, I pick VGG19 model and change the last layer to match the number of class of our dataset. I train for 20 epochs with SGD optimizer with the learning rate of 0.001. Since VGG19 model is pre-trained, it is expected to get a better performance as compared with proposed baseline method.

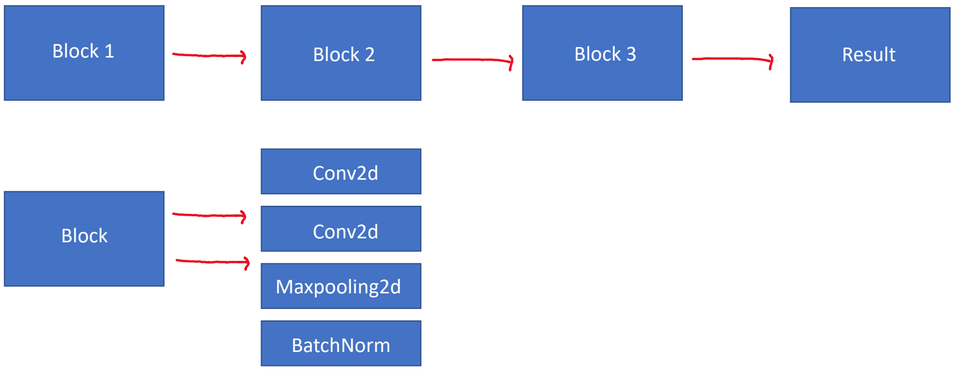

Figure 1: Visualization of simple CNN structure.

4 Dataset

4.1 Datasets and Inputs

In this project, I will use two datasets. Both of them are from Udacity project requirements. The first one is a dog dataset and the second one is a human face dataset. The project objective is to make a application where it is possible to predict if a given image contains a dog or human face, or something else.

Here is a simple summary for statistic of dataset. For human dataset, there are 13233 image in total. Majority of them have the same resolution. For dog dataset, there are in total 8351 images. Those image are divided into three (train/val/test) subsets. There are 6680 training image, 835 validation images and 836 test images respectively. The links for both dataset have been included in the Readme file.

I primarily use dog dataset as the input data to train a Convolutional neural network. They are pre-processed by a pytorch transformer with resize and normalization operation, together with different data augmentations. Then, a dataloader will serve data to pytorch deep learning framework.

4.2 Evaluation Metrics

For evaluation metrics, accuracy is the primary choice due to the topic of the project. The whole project is not multi-classification problem and thus there is no need to involve recall or precision, which is more suitable for multi-class classification problem.

5 Conclusions

In the proposal, I primarily discuss the proposed methods to implement a dog breed classifier. Literature for image classification has been reviewed as well as the introduction for baseline method and proposed method. The descriptions of dataset has been discussed in detail. The next step is to implement the proposed method to see how good it will be.

References

He, K., Zhang, X., Ren, S., & Sun, J. (2015, December). Deep Residual Learning for Image Recognition. *arXiv e-prints*, arXiv:1512.03385.

Simonyan, K., & Zisserman, A. (2014, September). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv e-prints*, arXiv:1409.1556.