

 capstone_proposal.md

Machine Learning Engineer Nanodegree

Capstone Proposal

Jong Lee
07/21/2020

Proposal

Domain Background

Image identification and classification has been one of the most prominent and rising Machine Learning research areas. Examples include self-driving cars recognizing cross streets and people¹, as well as medical radiology to analyze MRI and CT scans². Among many cases, animal recognition has been one of the foundational use cases which set the groundwork for other image recognition use cases. While there are a range of Machine Learning algorithms that can solve this image recognition problem, neural networks have been one of the most prominent algorithms with the rise in computing power and GPUs³. Neural networks perform well not only in binary classification but also in multi-classification problems, especially with large amounts of data and features, which is typical of many image recognition problems. However, beyond image recognition we wish to test the effect of transfer learning in training other image recognition models as well.

Problem Statement

Largely, there are two problems we are aiming to solve:

- **Dog breed classifier:** Given an image of a dog, identify the closest canine breed to that dog image.
- **Human resembler/classifier:** Given an image of a human, identify the most closely resembling canine breed.

Datasets and Inputs

The data we will use to train the models are images, as we are developing image recognition models. There are two types of images we are using: 1) dog images, and 2) human images. The images dataset was provided by Udacity.

- **Dog images:** There are a total of 8351 dog images, which will be divided into train, validation, and testing sets during the training of the model. Furthermore, there are 133 folders each corresponding to a different dog breed. The images are taken from a variety of angles, backgrounds, and sizes. Furthermore, one aspect to note is that some dog breeds have more images than others, resulting in not a uniformly balanced dataset.
- **Human images:** There are a total of 13233 human images. Similar to the dog images dataset, the human images have different backgrounds and angles; however, they are all sized the same. Again like the dog images dataset, the human images dataset is not uniformly balanced, as some humans have more training images than others.

Solution Statement

One solution we will be using to solve this image recognition problem is **Convolutional Neural Networks (CNN)**. A CNN is a deep neural network model which uses a combination of features, weights, and bias to output - in our case - a dog breed classification. Specifically in our project, our overall solution will involve multiple models, each developed in the steps below:

1. Create a dog/human detector (whether an image is of a dog/human) using pre-trained models
2. Create a dog breed classifier from scratch
3. Use transfer learning to develop a new model (not from scratch)

These three models will output the different dog breeds, from which we can check how accurate our models were using evaluative metrics.

Benchmark Model

Our benchmark generally would be a random guess. For our specific example, we have 133 dog breeds; a simple random guess would achieve 1/133, or 0.75% accuracy. However, since we are using a CNN (which should be better than a random guess), our test set accuracy should achieve at least 10% accuracy.

In the case of our transfer-learned model (not built from scratch), we would probably hope to achieve a better accuracy as we already have significant learnings from a previous model. Therefore, as a benchmark we would expect a test set accuracy of 60% or higher on our transfer-learned model.

Evaluation Metrics

The initial, starting evaluative metric will be accuracy: that is, did we correctly predict/classify the true dog breed? However, one downside to note is that accuracy doesn't fully reflect the imbalance in our datasets, with more than 100 different dog breeds. Therefore, at least in terms of training the model, we will be using the widely recognized multi-class log loss function to train and evaluate the model as well.

Project Design

The project will largely follow the detailed steps below:

1. Import the necessary datasets (dog and human images)
2. Use OpenCV - a previously implemented human face detector - to detect where in the image a human face lies (if at all).
3. Similarly, use VGG-16 - a previously implemented/trained dog detector - to detect dogs.
4. Now that we know which images have humans and dogs, create a CNN from scratch that identifies the *breed* of dogs (i.e. Chihuahua)
5. Validate, test, and evaluate the CNN model.
6. Develop another CNN model, but using transfer learning from the previously developed CNN model.
7. Validate, test, and evaluate the transfer-learned CNN model.

1: D. Cheng, Y. Gong, S. Zhou, J. Wang, N. Zheng. "Person re-identification by multi-channel parts-based cnn with improved triplet loss function". Proc. of IEEE Conference on Computer Vision and Pattern Recognition (27-30 June 2016), 10.1109/CVPR.2016.149

2: McBee, Morgan P., et al. "Deep learning in radiology." Academic radiology 25.11 (2018): 1472-1480.

3: Chen, Hongming, et al. "The rise of deep learning in drug discovery." Drug discovery today 23.6 (2018): 1241-1250.