Capstone Proposal: Machine Learning Convolutional Neural Network For Dog Breed Classification

---

Domain background

Classification of objects into categories, groups, or types is essential in technology, business, and biology. It allows us to generalize knowledge for quick understanding or application. As individuals, we are constantly assessing our own environment for safety or danger, or perhaps, clues to success or failure. Though we don't actively classify our assessments, the concept and its importance remain the same. Classification is essential to nearly all domains of life. It only makes sense we seek technology that aids us to do this.

Thirty years ago, these concepts inspired science fiction. In Terminator 2, Arnold Schwarzenegger had famously quoted "My CPU is a neural net processor… a learning computer." The real breakthrough in computer vision happened in 2012 and arguably brought forth a wave of confidence with it. The AlexNet neural network outpaced competition by greater than 10% in the ILSVRC (ImageNet Large-Scale Visual Recognition Challenge). The challenge was to accurately classify a set of 1000 classes of images [1,2]. Since this time, neural networks have gained more prominence and acceptance beyond the "black box" it was described as. However, success breeds success, and a slew of deeper and non-sequential neural networks have continued the advancement, led by large multinational companies like Google and Microsoft [3].

Innovative applications have appeared in various domains, such as in healthcare. In 2017, deep convolutional neural networks (CNN) were used for detection of skin cancer with dermatologist-level accuracy [4]. The application wasn't restricted to CNN application. Neural networks have since been used in healthcare diagnosis and administrative decision making [5].

Since that healthcare problem is "solved" barring the obstacles of regulation, we do have one that is quite similar, yet more fun: dog breed classification! We seek to classify pictures of dogs into one of 133 dog breeds in our data. This problem is less abstract than classifying skin lesions, but can have practical application at a veterinary office, provide useful learnings, or a fun application (which is our intent). This paper is not one to be nominated for a Nobel prize, but can be applied to thousands of similar problems in different domains, that perhaps collectively aid mankind.

Problem Statement

The aim is to accurately detect a dog in an image and predict the breed of that dog, based on 133 classes. This is accomplished in three parts – separate algorithms that determine whether (1) a dog is in an image (2) or a human is in an image and (3) what dog breed is predicted for the image. Our main focus will be on (3).

Datasets and Inputs

The following lists the data and inputs for model creation (note this does not list the data science and software tools):

- Dog Breed Image Dataset
  - This is a dataset of 133 classes/breeds of dogs, a total of 8,351 images. Each of the 133 dog breeds will be divided into a train, validation and test data set.
- Human Face Dataset
  - This is a dataset of 13233 images of humans.

Both datasets are comprised of images of various size and perspective, but all color (RGB, 3 channel). The datasets are used in the project in order to train, validate, and test predictive models.

Proposed Solution

The anticipated solution will use a machine learning framework based on a convolutional neural network (or multiple) to take an input image, detect if is a dog, and then classify that image based on the available classes. In particular:

- Convolutional neural networks, using transfer learning with well-known architectures, freezing the early layers, and fine tuning the last few layers to classify to dog breeds.
- The recommended solution is intended to be operational, and fast, in order to be used in an application. In particular a MobileNet, VGG, or SqueezeNet model will be trained, tested and selected.

The aim is to measure performance via appropriate metrics, in a replicable and quantifiable manner. As a result, datasets are provided in the GitHub repository, as well as code to replicate the model creation as selected in the final application.

Benchmark Model

We need a model, preferably parsimonious or historical, to compare the proposed solution with. As requirements for the project, a CNN from scratch is to be created. Random chance indicates that 0.75% of the time the correct class will be selected (based on 133 classes). However, our expectation is to have a benchmark model built and trained from scratch, that achieves greater than 10% accuracy. This will be our benchmark model.

The benchmark model will employ random initialization of weights, approximately five convolutional units, max pooling, batch normalization, dropout, and a final dense layer for classification. Multiple epochs and a smaller batch size will be used. The exact architecture will be built to exceed 10% accuracy based on cross categorical entropy.

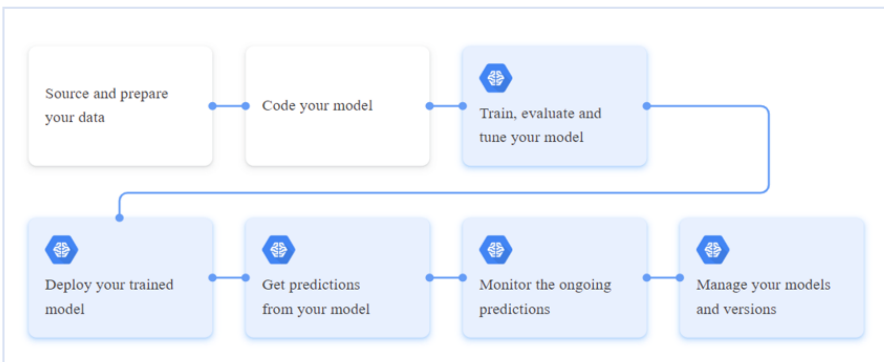
Evaluation Metrics

Both the benchmark and solution model(s) will be evaluated for accuracy (select correct dog breed class or not) using cross categorical entropy. Compared to MAE and other metrics, cross categorical entropy has been noted to allow better training with more complex datasets, and more generalizable results. [6]

Project Design

The general framework of the project follows these steps, and iteratively through multiple models as defined in the problem statement. These steps are as follows:

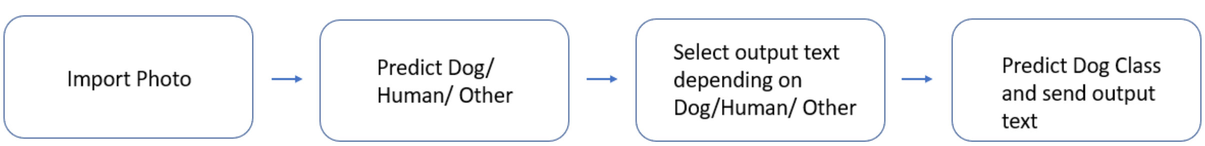
1. Retrieve Input Data
2. Clean and Explore
3. Prepare/Transform
4. Develop/Train Model
5. Validate/Evaluate Model
6. Deploy to Production
7. Monitor and Update Model & Date [7]



ML workflow

The lion share of the time in the project is expected to be spent in stages 4-5 or stage 6. Multiple iterations of the benchmark and proposed solutions will be developed, reflecting the need for development and evaluation. However, deploying to production while meeting operational requirements (avoiding lag time) is challenging as well, and likely to pose unexpected software and versioning challenges.

The application will make use of Heroku, an open source PaaS, with a Bootstrap framework. The CSS and HTML will be customized for simplicity but maintain the connectivity between the various components (algorithms). This will be our application's "scaffolding", with all documents maintained on GitHub. The Heroku framework was chosen due to ease of development and low cost. At last, the combination of models (1), (2), and (3) will be used to evaluate an image and predict the dog class as follows.



Citations

[1] https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[2] http://www.image-net.org/challenges/LSVRC/

[3] https://adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html

[4] https://www.nature.com/articles/nature21056

[5] https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0212356

[6] https://papers.nips.cc/paper/8094-generalized-cross-entropy-loss-for-training-deep-neural-networks-with-noisy-labels.pdf

[7] https://Udacity.com

[8] https://cloud.google.com/ml-engine/docs/ml-solutions-overview

[9] https://gombru.github.io/2018/05/23/cross_entropy_loss/

[10] https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/CNN20Whitepaper.pdf