Machine Learning Capstone Project Proposal
Dog Breed Classifier

Stefan Nae

June 2020

1 Domain background

In the field of Computer Vision (CV) the dog breed classification task fits into the fine-grained image classification (FGIC) research area. FGIC covers a wide range of CV problems including but not limited to the classification of vehicle models (e.g. the FGVC-Aircraft dataset [1]), animal species (e.g. the Caltech-UCSD Birds-200-2011 dataset [2]), and plant diseases (e.g. the Plant Pathology 2020 challenge dataset [3]) where the intra-class variance is high while inter-class variance can be both high and low. FGIC-related datasets are relatively small with a limited number of examples per class leading to extra emphasis on model generalization techniques.

2 Problem statement

This work is targeted at the dog breed classification task as posed by the Columbia Dog Breed dataset [4] characterized by high variance both intra- and inter-class. The task is to identify different dog breads.

3 Datasets and inputs

The Columbia Dog Breed dataset [4], containing labels for 133 dog breeds across 8351 images accompanied by eight part labels, together with the Labeled Faces in the Wild dataset [5, 6], packaging images of 5749 people (1680 people with two or more images) for a total of 13233 images of 250 by 250 pixel mostly color with a few grayscale, both with Udacity-provided access through the dedicated Dog Project Workspace constitute the data for the project. The first dataset is used for training and testing the neural network solution for dog breed identification while the second dataset is used in testing the human face detection and input extraction prior to inference using the dog breed classifier on arbitrary inputs for a customized user experience able to distinguish between humans and non-humans.

4 Solution statement

The solution will explore the use of convolutional neural networks (CNNs) within both a custom solution and by transfer learning using a pre-trained model, like ResNet, DenseNet, or Inception, as input feature extractor (encoder) for training a fully-connected classification layer. Furthermore, the solution will be able to distinguish between humans and dogs and will output the top-1 dog breed class with the associated probability, as a similarity score within intra-class dog breeds, for a human input.

5 Benchmark models

For reference the performance of the solution will be compared to the initial research work result on the employed dataset, a top-1 breed inference accuracy of 67% [4].

6 Evaluation metrics

The model performance will be assessed using the accuracy metric on a test dataset from a train/test split.

$$Acccuracy = \frac{Total\ number\ of\ correct\ predictions}{Total\ number\ of\ tested\ cases} \quad (1)$$

7 Project design

The solution will follow the machine learning life-cycle:

1. explore and process data (load and explore the datasets, identify the dog-breed image transforms useful in augmenting the limited dataset class-sizes – e.g. use random rotation, crops, and horizontal flips –, and create a train/test split on the initial data);

2. modeling (train two CNN models – a fully-custom CNN based architecture and partially-custom CNN architecture based off of a pre-trained model – using a train/validation data split for hyper-parameter tuning and early stopping and prepare for generic model evaluation with both dog and human inputs to first identify if there is a human in the picture (using a Haar cascade for both identifying a face and extracting it for model inference)), then submit the image for inference, and provide a customized display of the breed classification result correlated with the human of non-human character of the input;

3. deployment (create a REST API using Flask).

References

[1] Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft, 2013.

[2] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011.

[3] Ranjita Thapa, Noah Snavely, Serge Belongie, and Awais Khan. The plant pathology 2020 challenge dataset to classify foliar disease of apples, 2020.

[4] Jiongxin Liu, Angjoo Kanazawa, David Jacobs, and Peter Belhumeur. Dog breed classification using part localization. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *Computer Vision – ECCV 2012*, pages 172–185, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[5] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[6] Gary B. Huang Erik Learned-Miller. Labeled faces in the wild: Updates and new reporting procedures. Technical Report UM-CS-2014-003, University of Massachusetts, Amherst, May 2014.