Dog Breed Classification

Problem statement

Identification of dog breeds is a difficult task. Even experts and people who work with dogs for a living are not able to identify the breed of a dog by simply looking at it **olson2015inconsistent** Adding to the problem is that fact that most dogs are actually mixed-breed; that is, they are a combination of two or more breeds. Determining the breeds in a mixed-breed dog by visual inspection is nearly impossible for the human eye. For example Figure 1 shows a pure-bred border collie and a pure-bred beagle alongside a border collie-beagle mix. The mixed breed looks nothing like a beagle.

(a) Border Collie

(b) Beagle

(c) Border Collie - Beagle Mix

Figure 1: An example of how a mixed breed dog can look very different than its constituent breeds.

Dog-breed identification is not simply an academic problem. Correctly identifying dog breeds has implications for public policies and adoption strategies for dog shelters **king2012breeding** Adoption strategies focus on how to optimally match shelter dogs with prospective families. Prospective adopters usually want to know as much information about the dog as possible. For example, an active family may want to adopt a dog who will join them on hikes and other outdoor activites. Another family may just want a dog who will cuddle with them on the couch. Although every dog has its own personality, a dog's breed can give an indication of its broad personality traits. Moreover, the breed of a dog is also a great indicator of how big the dog will be when it is fully grown. The size of a dog can be a determining factor in adoption. It turns out that humans are very poor at determining what breeds are in a dog based solely on visual inspection. Even veterinary experts and shelter staff struggle to identify the component breeds of mixed-breed dogs. Multiple studies have demonstrated the shortcomings of human visual identification of dog breeds. In Dog

Breed Identification: What kind of dog is that?, the authors show that experts identified the prominent breed correctly only 27% of the time. In another work **olson2015inconsistent** the authors show that shelter staff consistently misidentifies pit-bull type dogs. Better breed identification could help city shelters implement better adoption policies, which could in turn contribute to greater shelter throughput. More dogs would therefore be off the street and placed in loveable households.

The breed identification problem also has a bearing on breed specific legislation (BSL). According to **weiss2001breed** BSL is "a statute or regulation that is directed toward one or more specific breeds of dogs." BSL often involves banning a particular breed because it is deemed to be dangerous. The breeds can be banned from countries, cities, neighborhoods, or even individual apartment complexes (see e.g. Wikipedia: BSL for a generic overview). There are numerous stories of people being forced out of their home because they share their home with a banned breed (see Denver BSL Article). Some of these stories have additional twists in which their dog was actually not even a banned breed, but was misidentified as such because the control officer was unable to identify the breed correctly. Moreover, there is mounting evidence that BSL is not an effective policy to reduce dog bites or attacks **mora2018fatal**, **ott2008there**, **cornelissen2010dog** Other work has discussed negative aspects of BSL, including misconceptions of breeds contributing to biased reports of dog attacks **creedon2017dog** which can lead to poor shelter adoption rates.

The dog-breed classification problem is important and challenging. In this project, you will take a step towards classifying dog breeds from a dataset consisting of over 20K images involving 120 different breeds.

Project goal: Propose, build and evaluate models for classifying pure dog breeds. You will not consider mixed-breeds in this project.

Data resources

The data will come from the **Stanford Dogs Dataset**: `http://vision.stanford.edu/aditya86/ImageNetDogs/main.html`.

High-level project goals

- Build a few models to classify dog breeds.
- Evaluate the predictive quality of the models.
- Compare the results from each model.
- Discuss the relative merits of each model.
- Use your model to determine "outliers".

This project has a several challenges:

- Significant pre-processing of the images must be done. Not all images are the same size. Most are color but a few are in black and white. The truth labels and directory structure will need to be cleaned up and modified for your model to easily process the data.
- Although the database consists of 20,580 images, it also includes 120 breeds. This means that there is actually not a lot of data for learning. There are only about $150 - 200$ images per dog breed. You may want to consider data augmentation techniques. You can also think about working with super-classes rather than each individual dog breed. A superclass may be defined as a broad class of dog breeds. For example, border collies, blue heelers, and Australian Shephards are examples of herding breeds. The American Kennel Club (AKC) has grouped dog breeds into seven distict groups (AKC Dog Breed Groups). A nice description of each group can be found at the National Dog Show website: The Seven Groups. You may want to consider using the seven groups as superclasses.
- You will need to think about a good way of doing outlier detection. We consider a dog breed an outlier if it is significantly different than all other breeds. For example, a pomeranian dog looks completely

different than border collies and Australian shephards. We would therefore consider a pomeranian to be an outlier. One possibility of evaluating the outlier detection is to hide some breeds from the training and check if the model identifies them as outliers.