Dog Breed Image Classification

Nicole Aguilera
DSC 680-T302 – Summer 2021
https://github.com/ntiana55/Portfolio_NicoleAguilera/tree/main/AppliedDS/Project2

Data Domain

The domain of this data is breeds of dogs. This project will classify images via deep learning methods. While the data set being used for my project is a popular data set on Kaggle, it can be used in many different ways, as seen in some of the sources below. The reports will only be used for guidance and will not be copied, as models can vary with each construction. Below is the annotated bibliography for this project.

- Dog Breed Classification. Kaggle. https://www.kaggle.com/abhinavkrjha/dog-breed-classification
  This is the source of the data for this project. It includes nearly 700 images of dogs of different breeds for the model to classify.
- Dog Breeds. American Kennel Club. https://www.akc.org/dog-breeds/
  The American Kennel Club recognizes 195 breeds of dogs, with many more breeds that are working towards being recognized officially. With this project, we will only be recognizing a small portion of that total 195 breeds.
- New App DoggZam! Can Identify Dog Breed With Just a Photo. (2018). PetMD. https://www.petmd.com/news/lifestyle-entertainment/new-app-doggzam-can-identify-dog-breed-just-photo-37329
  An application exists that can identify the breed of your dog just by taking a photo and uploading it to their software. This uses image classification software in the real world.
- Dog Breed Identification. (2018). Kaggle. https://www.kaggle.com/c/dog-breed-identification/overview
  Kaggle hosted a competition for using a classifier for identifying dog breeds. Participants were allowed to include pre-trained models in order to complete this task. The competition concluded 3 years ago.
- GitHub. https://github.com/prateeksawhney97/Dog-Breed-Classification-Project-Using-Flask
  The user, Udacity, of GitHub was able to produce a web application through flask to classify dogs by their images. If a picture of a human was uploaded to the application, the program would return the dog breed that the human looks the most like.
- Dog Breed Identification. Stanford. https://web.stanford.edu/class/cs231a/prev_projects_2016/output%20(1).pdf
  This project was created by students to construct a model to classify dog breeds. Facial keypoints are essential to identify when creating a classification model such as this.
- Build Dog Breeds Classifier. KD Nuggets. https://www.kdnuggets.com/2020/06/build-dog-breeds-classifier-aws-sagemaker.html
  This dog breeds classifier is accomplished through the AWS SageMaker console. This provides a different approach to classifying images.
- Write an Algorithm for a Dog Identification App. Datascience-Enthusiast. https://datascience-enthusiast.com/DL/dog_app.html

This site provides a guide on how to construct an application for the image classification of dog breeds. The process of CNN was used in this project.

- Dog-breed. GitHub. https://github.com/topics/dog-breed
  This page of topics provides repositories for reference of dog breed classifiers. Everyone has a different approach and method to the model that they construct.
- Dog Identification using Soft Biometrics and Neural Networks. https://arxiv.org/pdf/2007.11986.pdf
  This project used soft biometrics rather than just the use of neural networks. The soft biometrics used were the breed, height, and gender of the dog. This supported in identifying which breed the dog was.

Data

The dataset that is being examined is Dog Breed Classification, uploaded by a Kaggle user. There are 14 breeds of dogs in this data set to be classified. The data set can be found at the following link: https://www.kaggle.com/abhinavkrjha/dog-breed-classification

Research Questions and Benefits

Some research questions for this classifier are, but not limited to, the following:

- What percentage of dog breeds will the classifier get correct?
- Will the color of the dog impact the model's ability to classify?

The benefits to running this classifier can be translated over to any other image recognition software. Even today, some classifiers have a hard time distinguishing between images because it has not been trained off the correct images. Running more classifications models like this can help work out errors found in other models.

Methods

The main method being used is classification through either TensorFlow or Keras. This will be creating a neural network in order to classify these images.

Potential Issues

The obvious potential issue is that the model will not work because there is not enough variation in the data set. This would cause the model to not be able to classify the images correctly.

Another issue, that might be unavoidable, is that the image uploaded is of a dog of a breed that has not yet been distinguished as a breed or they are a mutt, which is a dog of a mix of multiple breeds. The breed identified will be the closest resembling breed, which could potentially be one of the breeds that the dog is mixed with.

Concluding Remarks

The goal of this model is to classify what type of breed the dog is in the image through a classification model. There are around 700 images being run for this project. Although there are only 14 breeds of dogs being run against the model, this is a good baseline for images of other dogs to see what type of group the breed would belong to based off its physical features. The data set being used can be found on Kaggle: https://www.kaggle.com/abhinavkrjha/dog-breed-classification