Younghoon Choi
00433997
ychoi2@conncoll.edu

Final Project - Write Up
Machine Learning, COM307

2022-11-27

1 Overview

This program was developed to classify different dog breeds from an image. I chose python as my language because of all the library access I knew I had such as sklearn, numpy, and matpilot.

I was interested in this topic because I only knew about template matching for classification in terms of computer vision. I wanted to use what I learned in this class to dig in more into my interest of image processing.

2 Data Break Down

The data was imported from Kaggle which consist of images of dogs. Images had 5 different classes meaning 5 different breed including beagle, greyhound, bulldog, pomeranian and golden retriever.

Training Data had 275 images in total.
59 Beagles, 53 Bulldogs, 63 Golden Retriever, 56 Greyhound, 44 Pomeranian

Testing Data had 50 images in total
10 images of each breed

3 Algorithm

3.1 Pre-Precessing

Go through all the images in the training folder.
1. Process the images from the folder.
2. Resize the image, 150 by 150.
3. Flattened image to an array.
Goes into Training Mode

3.2 Training

Support Vector Machine
- Learning methods used for classification and regression.
- Effective in high dimensional spaces.

Grid Search CV
- Technique for finding the optimal parameter values from a given set of parameters in a grid.
- Cross-validation technique.
- After extracting the best parameter values, predictions are made.

Model
- Create a model that will be used for predictions.

3.3 Testing

Go through all the images in the testing folder or image file.
1. Resize the testing image to match testing, 150 by 150.
2. Flatten the images.
3. Predict each image to the model.

4 Analysis

4.1 Result

The total accuracy for all testing images is 88 percent.

Pomeranian had 100 percent of accuracy as I believe its unique hair helped to classify.

Beagle had 90 percent of accuracy as the spots helped.

Greyhound had 90 percent as its lengthy mouth and body helped to classify.

Bulldogs had 80 percent of accuracy because of certain bulldog's spots from certain angles make them look like beagles.

Golden Retriever had 80 percent of accuracy. It had issues with classifying younger golden retrievers as pomeranians.

4.2 Discovery

I was limited due to my status and CPU at the moment in terms of access to data sets. I would like to work with a bigger data set both for training and testing data. Not only more breeds and images for each of them but different categories with in that category. For example, grown gold retriever vs young retriever. I think that will fix the issues I am having of bulldogs being different color, hair quality of dogs in same breed... etc.

I would also like to try training and testing the data in different size. I worked with 150 by 150 for reduced training time but curious to see the result in bigger size.