Capstone Project 2: Dog Breed Classification

◆ Problem and data set

This project tries to identify the dog breeds by their images. It will be using the Stanford Dogs dataset, which contains 120 different breeds (categories) of dogs, added into 20580 total images. Labels (dog breeds) are provided.

◆ Project Motivations

Technical Motivation

Dog classification is one of the fine grained image recognition problems, where the differences between categories are relatively subtle. In this case, for example, a Labrador retriever and a golden retriever are much more alike than a Labrador and a boat, hence it is quite a challenging problem in machine learning.

Also, if an algorithm can learn to tell the subtle differences between dogs, it may very well likely to be able to learn subtle differences between other similar objects, such as breeds of other animals, styles of paintings, and so on.

Potential Products

Besides the technical challenge, it is also a practical problem to solve. If certain accuracy is achieved, it can be built into a dog lovers' app, which would be able to identify a dog's breed by a picture of it, and from there on the app can search from internet and finds out all the other information about that dog, such as its life span, disposition, and where it can be found/purchased etc.

◆ Data set and Pre-handling

There are 120 sub-folders, each of which contains the pictures of one specific dog breed. The original data set looks like this:

Image Folder:

- Dog breed 1: Images
- Dog breed 2: Images
- ……

In order to validate the training, I divided the data set into two parts, the training set and the test set, the test set contains 20% randomly selected images from the original data set, and the rest is the training set. And the altered data set looks like this:

Image Folder:

- Training set (80%):
  - Dog breed 1: Images
  - Dog breed 2: Images
  - ……
- Test set (20%):
  - Dog breed 1: Images
  - Dog breed 2: Images

……

The pictures' resolution varies from under 10 KB to over 1 MB. All of the images would be unified to 128*128 right before training, and no other image handing is performed at this point.

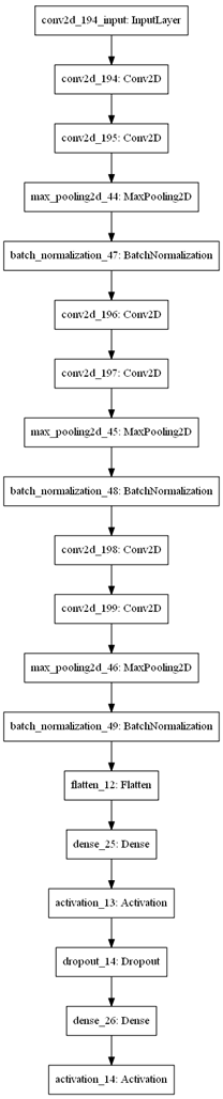
◆ Potential Data sets

Although there doesn't seem to exist a similar data set, pictures of certain breeds of dogs can be gained from search engine, and add into the dataset.

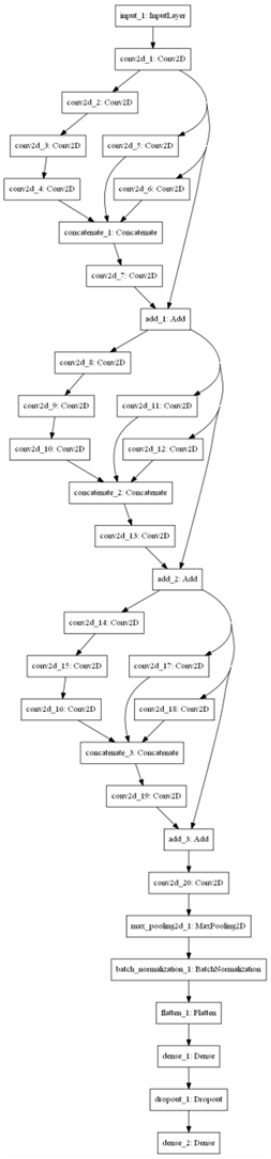
◆ **Models & Performances**

Three of neuro network models are built and trained from scratch. The architects of each model are as follow:

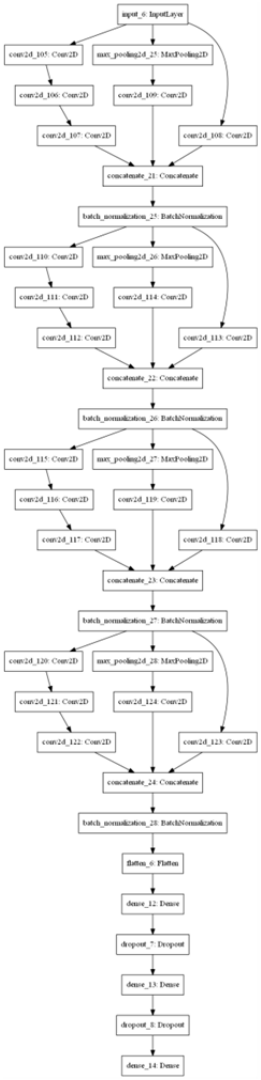
VGG Model:



Inception-Residual Model:



Inception Model:



The whole training process is divided into three steps:

Initially, in order to gain a fast training result, images are resized to a low resolution (32*32 pixels) before fed to the models. After an early estimation of the algorithm performances, models are trained with moderate resolution (128*128) images. After these steps, image resolution is increased again (200*200) to see if an optimum accuracy can be gained. The table below indicates the converged validation accuracy for each resolution/model:

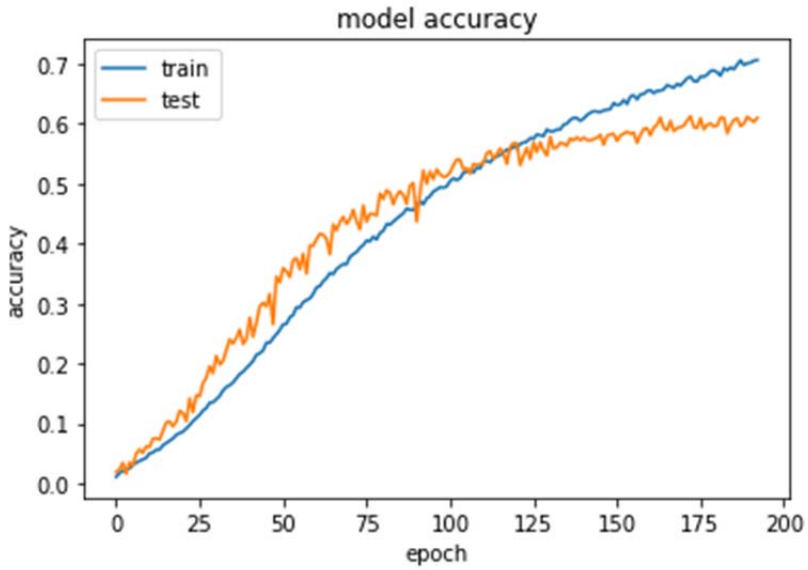
Resolution/Model	VGG	Inception-Residual	Inception
32*32	28%	26%	24%
128*128	58%	52%	40%
200*200	61%	56%	* N/A

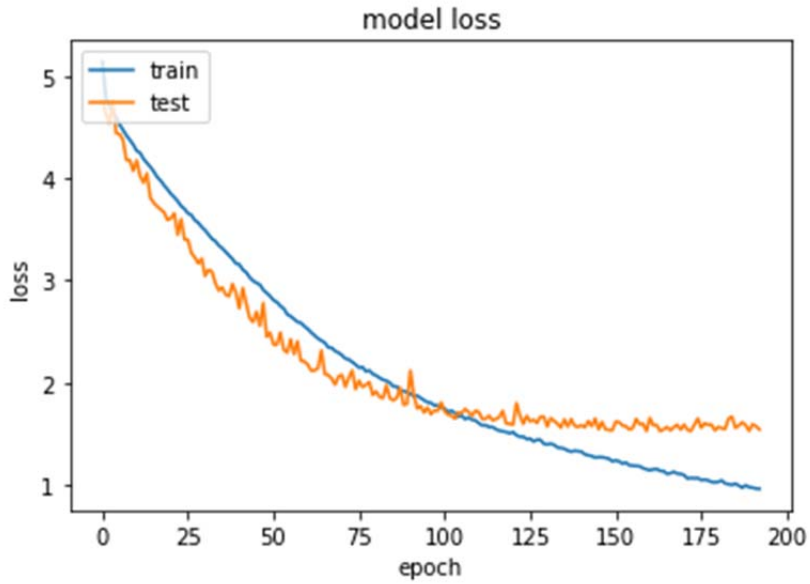
* By the first two training steps, It is clear that the inception model does not perform as well as the other two models, hence it was not trained with the 200*200 images.

◆ Results of The Third Training Steps

VGG model:

The overall training process for the VGG is shown in the following two figures:

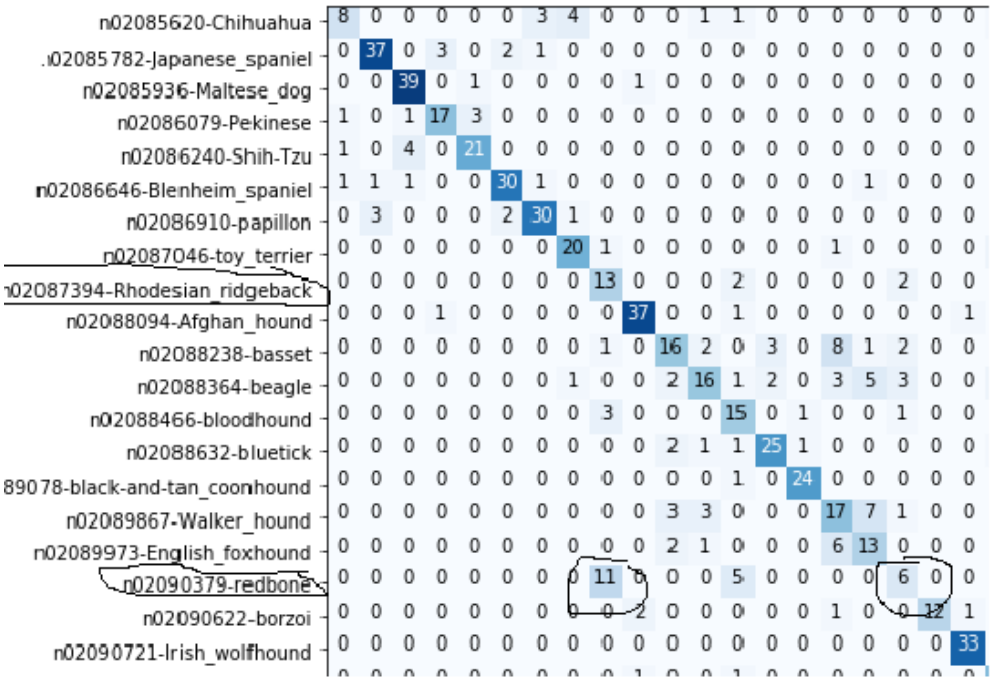




As the figures indicate, after around 100 epochs, the model starts to over-fit, the training accuracy keeps improving, but the validation accuracy seems to converge. At the end, the validation accuracy remains at the 59%-61% range.

For the VGG model, some of the dogs are tend to be confused with a certain another kind, for instance, below are three pairs of dogs extracted from the Confusion Heat Map (In the confusion heat map, the diagonal line indicates the number of images that correctly recognized, elsewhere are the numbers that confused with other kinds):

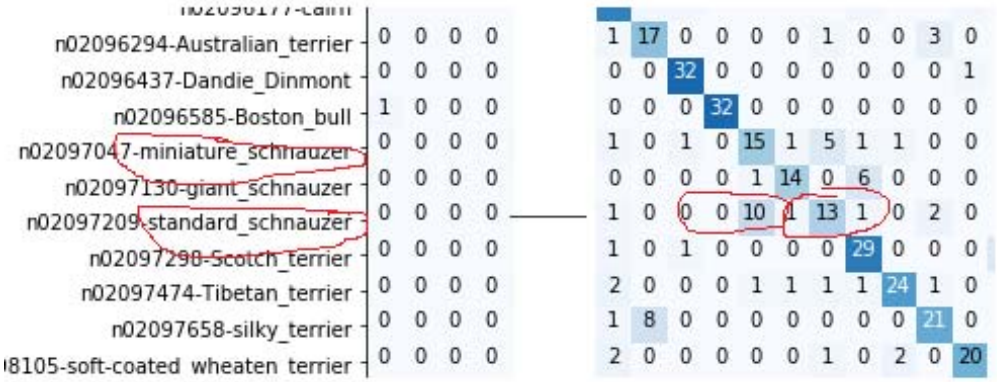
First Pair: Redbone and Rhodesian Ridgeback



The confusion heat map show that only 6 images of the Redbone is correctly recognized by the model, and 11 of them are taken as Rhodesian Ridgeback. And here is what the two kinds of dogs looks like:


Left: Rhodesian-Ridgeback Right: Redbone

Second Pair: Stand Schnauzer and Miniature Schnauzer:



13 images of Standard Schnauzer are recognized, 10 of them are taken as Miniature Schnauzer, and here are the two dogs:


Left: Miniature Schnauzer Right: Standard Schnauzer

Other than the sizes, they look almost exactly the same.

The Third Pair: Miniature Pinscher and Greater Swiss Mountain Dog

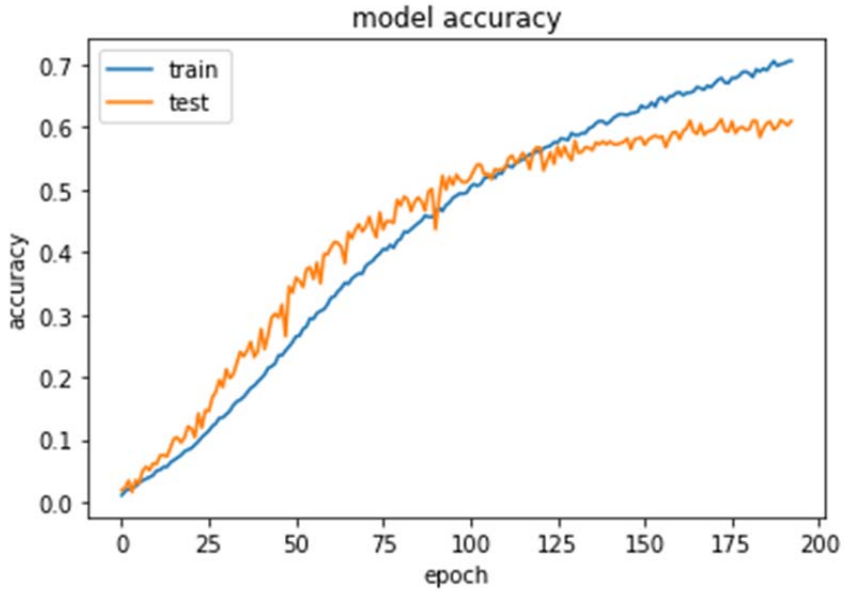
6 of the Miniature Pincher images are recognized, 12 of them are taken as Greater Swiss Mountain Dog, below are the images of the two dogs:

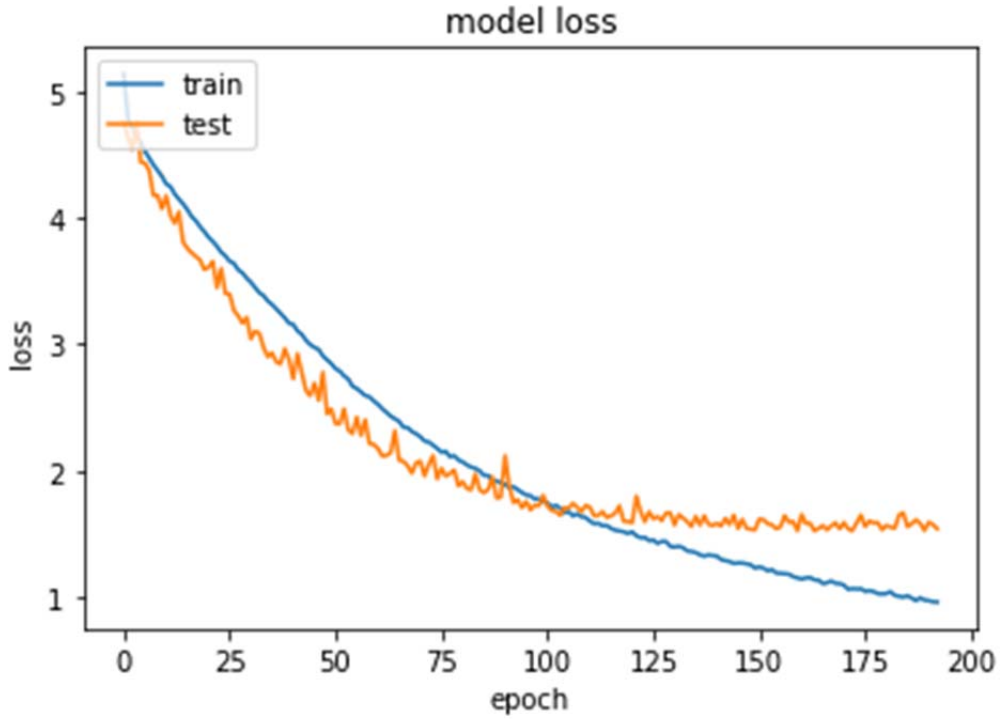
Left: Miniature Pincher Right: Greater Swiss

This pair looks quite different in human eyes, but somehow the model confuses them a lot.

Inception-Residual Model:

The overall training process for the VGG is shown in the following two figures:





After around 110 epochs, the model starts to over-fit, the training accuracy keeps improving, but the validation accuracy seems to converge. At the end, the validation accuracy mainly remains at the 55%-56% range.

The Inception-Model does not make the same confusion as the VGG model does, it can recognize the *Redbone and Rhodesian Ridgeback* pair better (9 right, 4 wrong); the *Stand Schnauzer and Miniature Schnauzer* pair about the same (7 right, 7 wrong); and worse for the *Miniature Pinscher and Greater Swiss Mountain Dog* pair (3 right, 14 wrong).

Below is the table of performances of the two models:

	Test Accuracy	Avg. Precision	Avg. Recall	Avg. F1-score

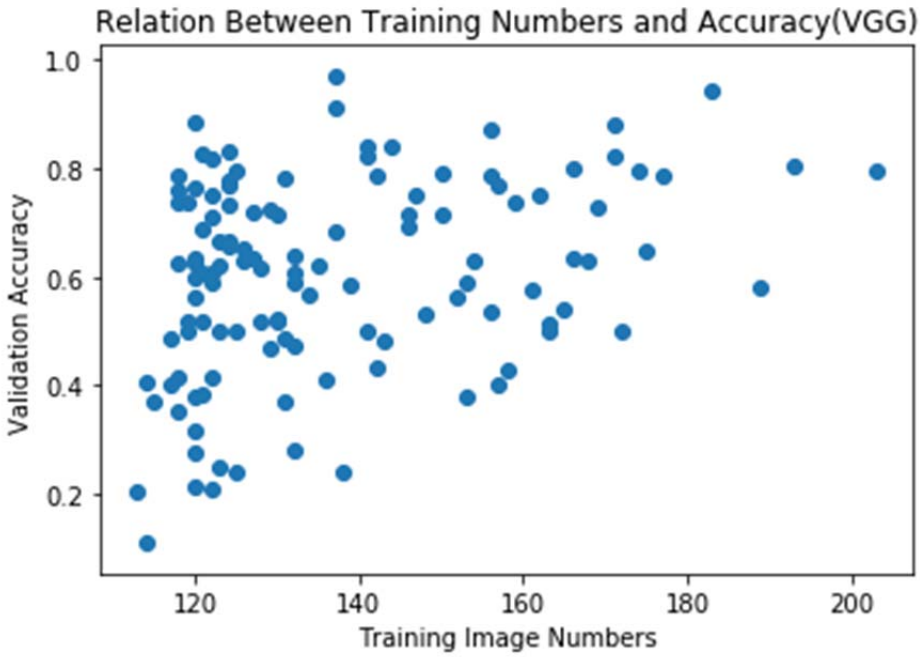
VGG	61%	62%	61%	60%
Inception-Residual	56%	54%	51%	51%

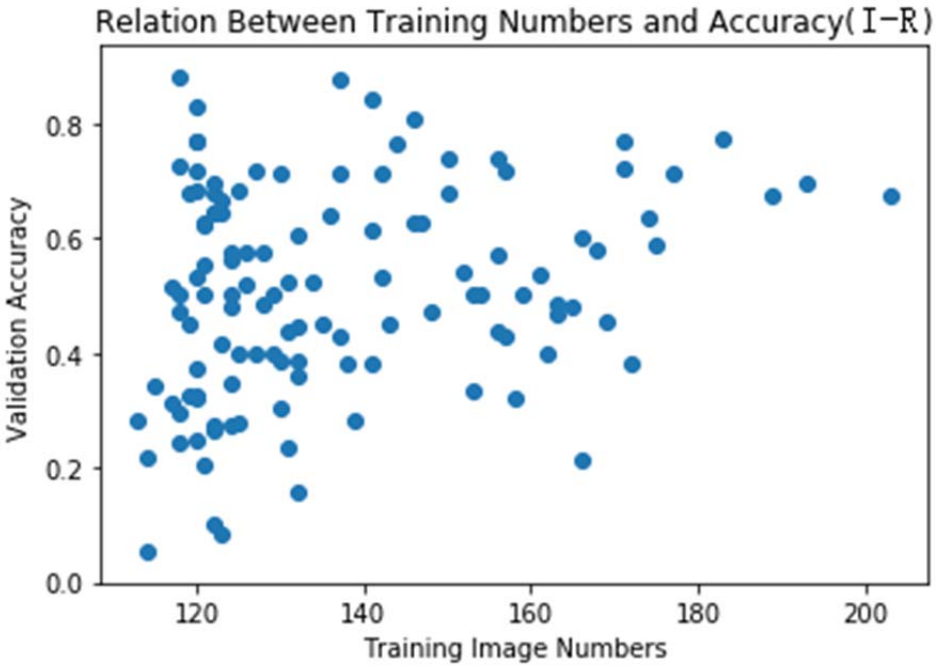
- Conclusion:

  With a fine grained image recognition problem such as dog breed classification, it is essential to feed the model with relatively high resolution images, while other conditions are the same, higher resolution can more than double the accuracy (from 28% to 61%).

  The right algorithm is fundamental; in this case, the difference of accuracy between models is higher than 20%.

  Also, sufficient amount of data can increase the accuracy vastly, below are the two figures indicating the relationship between the number of Training data and the validation accuracy:





Clearly, for both models, there is a strong positive correlation between the training data amount and the validation accuracy.