Create test data

Date : 2015.10.18
Tags :
Made by : Asutosh SATAPATHY

Goal : To a create test data (smaller).

 The training data is large. It takes a lot of time to run the code. It is not advisable to go over the whole training data for smoke tests. It will be better to work with a smaller subset of the training data to perform initial tests before finally running the system.

 It is improtant to include all the classes in the test data. This will ensure the application will perform properly when it is using the actual training data.

All the classes should have at least 10 images and less than 100 images, instead of hundreds or thousands of images.

Procedure :

One of the observations about all the test files was the name of the files. They all had numbers as their filenames and ranged from 100 - 999999. And every folder a mix of different length of filenames.

1. Make a copy of the training data.

2. Delete all files which have 5 or more characters in their filename.

Results :

1. Reduced number of images for testing from 30460 to 4652.

2. Runtime for smoke-test was reduced to under ~2mins as compared ~5-7 mins.