機器學習作業1

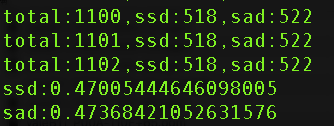
404410107蔡東霖

§ 實做內容

首先分出training set 跟testing set，training set共有38個label，分別為每個資料夾中排序過後的前35張照片，而testing set 則是所有剩下的照片，這邊import os中listdir的方式找path，再依照文件夾名稱做字串串接處理去訪問子目錄，讀取檔案則用PTL的Image.open，存到nparray的型態採用float，最後再設一個陣列test\_label存這些testing set原本的在文件夾的位置(即為label)，以便後續比較正確性。

接著是將testing set 中的圖拿來比對各個label中每張圖片，這邊我用了多重for迴圈一張一張做SSD及SAD，找出最小值後，看看label是否相同，相同則表示成功，找完之後成功張數除以總測試張數即可換算成功率。

§ 執行結果



SSD : 47%

SAD : 47.3%

§ 困難

一開始區分training 及testing時是採取存path的方式，所以多做了些字串處理的程序，也因此花了一些時間。之後發現，若是一開始在資料預處理時存path，到for迴圈中再讀檔的話，執行時間會非常長，但是改成直接在預處理時存成nparray的話，執行時間就會快上許多，應該是for迴圈中同一個檔讀太多次造成的。

由於測資部分並非全部都65張，所以一律設定為35張做為training 其他做為testing，資料夾也少了’14’這個人，所以在for迴圈中有特定處理。

原本存nparray的型態為int，但是發現成功率有點低，因此改成使用float的型態，可增加SAD SSD的精確性，並提高成功率。

不太熟悉python，在debug時有時會因為空格的不習慣而沒發現錯誤。