

Hw2 孫渝鈞 409510049 資工四

1. Method description:

what are your reference codes?

Color histogram: <https://blog.gtwang.org/programming/python-opencv-matplotlib-plot-histogram-tutorial/>

LBP: <https://chtseng.wordpress.com/2017/04/07/透過葉片紋理判斷植物種類-使用 lbp/>

<https://pyimagesearch.com/2021/05/03/face-recognition-with-local-binary-patterns-lbps-and-opencv/>

<https://www.geeksforgeeks.org/create-local-binary-pattern-of-an-image-using-opencv-python/>





Gabor filter: <https://stackoverflow.com/questions/28723916/how-to-do-feature-extraction-using-gabor-filter>

另外還有一些 Chatgpt 的協助

How to run your test?

將./train 資料集每張圖片去計算 feature 後以 columns=['filename', 'label', 'histogram']的格式去存，其中 histogram 為計算完後 feature 值的分類(有些方法沒有再 histogram)，接著在將./test 資料集每張圖片去計算 feature 後以 columns=['filename', 'histogram']的格式去存，其中 histogram 為計算完後的 feature 值。然後就是利用 NN 分類方式來進行分類，我寫了兩種分類方式 SAD, SSD，主要都是以 SAD 為主，除了 raw image 是以 SSD 為主。NN 就是去計算與每個 train feature 的距離並且以距離最短的那一個 label 當作分類結果，最後再將所有結果寫到.csv。

2. Experimental results

| | | | |
|--|----------------|----------------|--------------------------|
|  gabor_filters_sad.csv Complete (after deadline) · now | 0.22921 | 0.22921 | <input type="checkbox"/> |
|  LBP_sad.csv Complete (after deadline) · 22s ago | 0.21158 | 0.21158 | <input type="checkbox"/> |
|  color_histogram_sad.csv Complete (after deadline) · 1m ago | 0.7544 | 0.7544 | <input type="checkbox"/> |
|  raw_image_ssd.csv Complete (after deadline) · 2d ago | 0.33186 | 0.33186 | <input type="checkbox"/> |

3. Discussion

除了 color histogram 我的其他三種分類方式準確率都蠻低的，我也嘗試優化很久，甚至在 LBP 不套用 openCV function 而是用上網找的直接刻出來的方式準確率還是很低，那我想應該不會是程式有什麼問題，而是這種方式畢竟比較簡單，相對於利用深度學習方式，準確率確實會比較低一點。而 color histogram 會有 0.7 左右我想應該是它看了 R, G, B 三個通道的資訊，所以準確率會高一

些。在 raw image 一開始我的準確率只有 0.1 左右，但因為 raw image 本身更簡單一點，所以我才想說採用 SSD，取平方看看能不能提高準確率，結果實測，真的可以提升，不過也才提升到 0.3 左右而已。最後在 Gabor filter 一開始我的準確率也只有 0.1 左右，那我也想辦法去調 function 參數，最終也只能提升到 0.2 而已。

4. Problem and difficulties

在這次的作業中，我遇到最大的困難在於 color histogram 程式執行到一半就會被 killed 掉，我覺得是記憶體不足，所以一開始嘗試用 batch 的方式，想說用拆分成好幾個部分算應該就不會有問題，但是還是一樣解決不了，也有去監測 CPU 與 memory 用量，但是都顯示正常。



最後發現原來我的 color histogram bin 開到 256 所以每張圖會存 $256 \times 256 \times 256$ 個元素，確實是記憶體不足的問題。但會有電腦監測不出來記憶體問題，是因為有限制只能給程式一部分的記憶體，但其實整個電腦還有很多空間。所以之後把 bin 改成 8，就沒有程式被 killed 掉的問題了。

```
Last login: Sat Nov 18 06:08:26 on ttys000
(base) sunyujun@sunyujundeMacBook-Pro ~ % ulimit -a
-t: cpu time (seconds)          unlimited
-f: file size (blocks)          unlimited
-d: data seg size (kbytes)      unlimited
-s: stack size (kbytes)        8192
-c: core file size (blocks)     0
-v: address space (kbytes)      unlimited
-l: locked-in-memory size (kbytes) unlimited
-u: processes                   2784
-n: file descriptors           256
(base) sunyujun@sunyujundeMacBook-Pro ~ %
```

最後，真的非常感謝助教的協助幫忙。