

GoGoFruit! 利用深度學習判斷水果並計算營養量

國立台北大學 資訊工程學系 專題報告書

專題編號：PRJ-NTPUCSIE-108-008

專題成員：鄭景鴻、莊承翰、詹沁昀、林柏均

一、簡介

「營養攝取」為國人健康最重要的議題，而外食族(如上班族、學生)因為餐飲選擇的較少，往往不能兼顧營養均衡，為了補足日常所需的維生素和礦物質，最方便取得的食物為水果，為確保每日攝取的維生素足夠，市面上許多 APP 提供與營養師互動的功能，但大多數都須付費或等待營養師回覆，且營養師透過傳送的照片或使用者描述的情況不同，也會造成營養量預估與實際的落差。

GoGoFruit!是一款兼具「便利」、「準確」、「快速」的 APP，只要拿起手機對準水果按下快門，就能立即得知您欲食用水果的營養含量，**不需付費!不需費時等待!且準確度極高!**

二、模型選擇

在本計畫中，使用的核心為利用深度學習來判斷水果種類，並以其結果得知水果在照片上涵蓋的面積，依據同種水果密度大致固定的特性，即可將面積轉為重量，進而推估其營養量。

A. 神經網路選擇

深度學習神經網路的種類繁多，一開始考慮到快速與模型輕量的特性，我們採用 YOLO-v3，但發現其對於水果判斷，輸出的結果為水果位置的 Bounding Box (Figure 1)，無法得知水果準確涵蓋的面積，因此我們需要選擇

有 Mask 功能的神經網路。

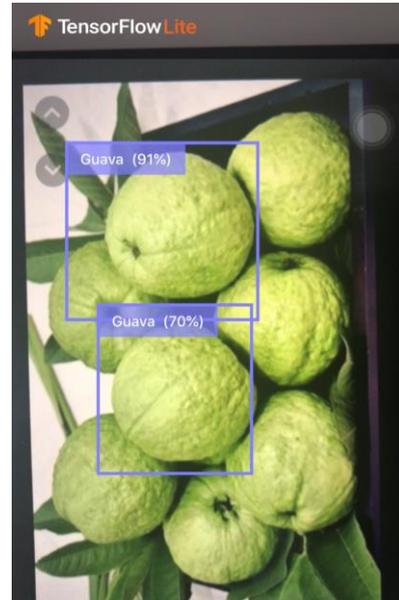


Figure 1 YOLO 示意圖

B. Mask 模型選擇

具有 Mask 功能的神經網路有許多種，像是 YOLACK、Mobile Mask RCNN、Mask RCNN 等，前兩者為放在手機端運行的模型，速度快但準確度較差，基於計算面積上較為重要的「準確性」，我們選擇了使用電腦端運行的 **Mask RCNN** 作為我們的核心模型。



Figure 2 Mask RCNN 結果示意圖

三、核心技術

A. Mask RCNN

Mask RCNN (Figure 3) 是由 Faster RCNN (Figure 2) 衍伸出的模型，兩者底層同樣都是先經 CNN 神經網路產生 feature map，再由 RCN 網路選出適當的特徵區域，最後產生 Bounding Box 和 Classification 資訊。

Mask RCNN 不同於 Faster RCNN 的是 Mask RCNN 會另外產生 Mask Branch 資訊，且 Mask RCNN 是採用 ROI Align 也就是雙線性內插法，解決了 Faster RCNN 用 ROI Pooling 在做 Max pooling 時，使用最近插值法會遺失部分 pixel 的問題，能有效提高面積預測準確度。

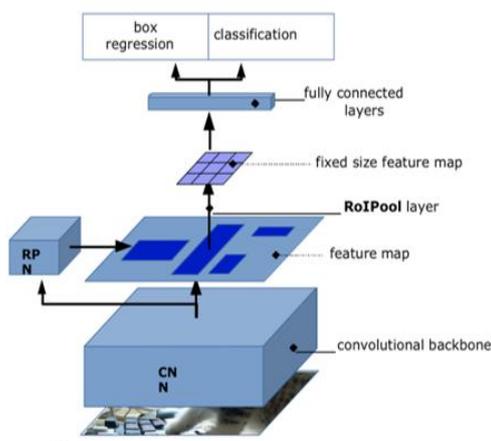


Figure 2 Faster RCNN 示意圖

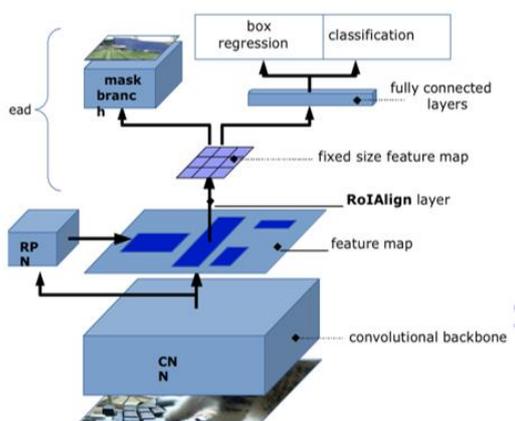


Figure 3 Mask RCNN 架構圖

B. 訓練資料準備

使用 Labelme 程式手動產生大部分訓練圖片資料集 (Figure 4)，並使用 Matlab 輔助產生少部分資料集 (Figure 5)，由於 Matlab 二值化只能偵測邊緣，不能標示多顆水果，所以只適用於背景單純、單一水果的圖片，效果有限，所以大多資料都是由我們手動 label 產生，耗時許久。

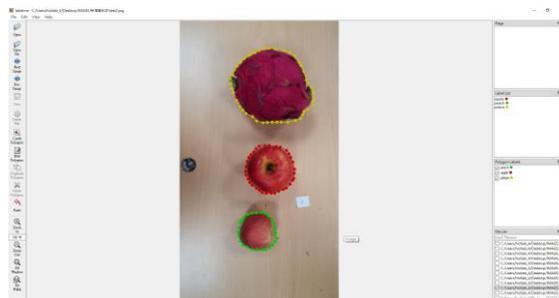


Figure 4 使用 Labelme 進行 label 示意圖

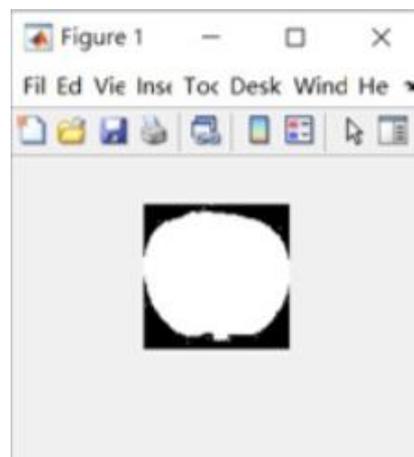


Figure 5 使用 Matlab 二值化取邊緣 示意圖

C. 訓練過程

將產生的 json 檔、原圖、Mask 圖片餵入模型訓練，一次跑一百個 epoch 並產生一個紀錄檔 (.h5 檔案)。

```

nctlab-group2@nctlab-ml:~/Workspaces/maskrcnn/Mask_RCNN [16x1]
訓練(C) 選擇(V) 拖拽(O) 確定(Y) 處理(C) 取消(N)
image_id 548
image_id 548
31/100 [=====] - ETA: 5:40 - loss: 3.0468image_id 548
image_id 548
33/100 [=====] - ETA: 5:12 - loss: 3.0271image_id 548
image_id 548
image_id 548
image_id 548
image_id 548
image_id 429
34/100 [=====] - ETA: 6:10 - loss: 2.9823image_id 429
image_id 429
image_id 429
36/100 [=====] - ETA: 5:42 - loss: 2.9015image_id 429
image_id 429
38/100 [=====] - ETA: 5:17 - loss: 2.7996image_id 429
40/100 [=====] - ETA: 4:53 - loss: 2.7005image_id 429
42/100 [=====] - ETA: 4:30 - loss: 2.5997image_id 429
43/100 [=====] - ETA: 4:20 - loss: 2.5490image_id 429
45/100 [=====] - ETA: 4:19 - loss: 2.5565image_id 1449
46/100 [=====] - ETA: 4:11 - loss: 2.6004image_id 1449
image_id 1449
image_id 1449
49/100 [=====] - ETA: 3:45 - loss: 2.6595image_id 1449
image_id 1449
image_id 1449
50/100 [=====] - ETA: 3:37 - loss: 2.6541image_id 1449
image_id 1449
53/100 [=====] - ETA: 2:59 - loss: 2.5449image_id 1449
image_id 1449
image_id 1449
56/100 [=====] - ETA: 2:53 - loss: 2.5211

```

Figure 6 訓練過程 示意圖

D. 營養量換算

有了 Mask RCNN 的結果，就能以下步驟換算營養量。

1. 算出 Mask 區域的 pixel 數：

不同類水果的 Mask 會在不同的 Channel 上，就能利用迴圈算出不同每個水果的 pixel 數。

2. 以 pixel 對照水果真實重量：

同種水果的密度大致相同，可以用 pixel 數去對映真實重量，求出之間的換算比例。

3. 重量換算營養量：

衛服部食品藥物管理署有提供食品營養成分查詢的功能，只要打上欲知的水果種類，就能知道其每 100g 的各種營養成分組成。

四、問題與解決方式

- 語意分割 VS 實例分割

語意分割 (Figure 7)：會將一張圖片中同一種水果判斷為一個個體。

實例分割 (Figure 8)：會將一張圖片中的每個水果都判斷為一個獨立的個體。

當時我們產生訓練資料時做成語意分割，導致訓練結果不理想，後來才發現要使用實例分割來產生資料集，解決方式就是全部重新 label，後來結果就有得到改善。

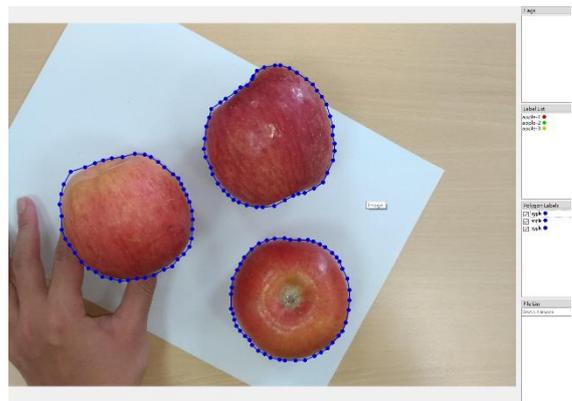


Figure 7 語意分割



Figure 8 實例分割

- 拍照高度影響結果

使用手機拍攝水果時，若是手機鏡頭跟水果的高度不同會使判斷出的 pixel 數有很大的差距。



Figure 8 同顆水果不同拍攝高度

解決方法：

1. 測距儀

使用手機內建的測距儀算鏡頭與水果間距，但需移動手機才能得知兩點距離，這與我們想要只拍一張照就知道結果的想法相違，所以不適用。

2. A4 紙

用 A4 紙當拍照基準，讓手機拍照畫面對準 A4 紙，讓他剛好填滿螢幕，以固定拍攝距離，但 A4 紙太小，能放入的水果太少，且便利性不足，很少人會隨身攜帶 A4 紙，所以不適用。

3. 十元硬幣

設計一個圓框在拍攝畫面上，讓使用者將圓框對準十元硬幣再拍照，如此就能固定拍攝距離，而且利用十元硬幣可以讓拍攝範圍增大，便利性也足夠，所以我們採用此方法。



Figure 9 利用十元硬幣對準圓框

五、實作結果



Figure 10 開始畫面

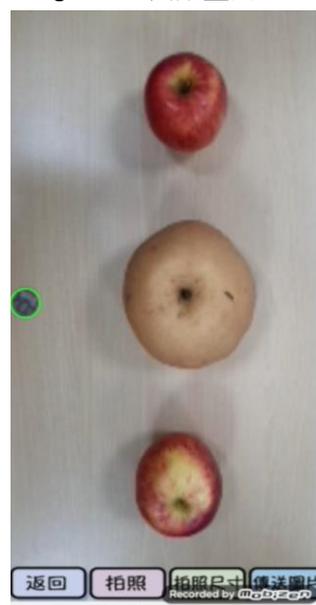


Figure 11 拍照畫面

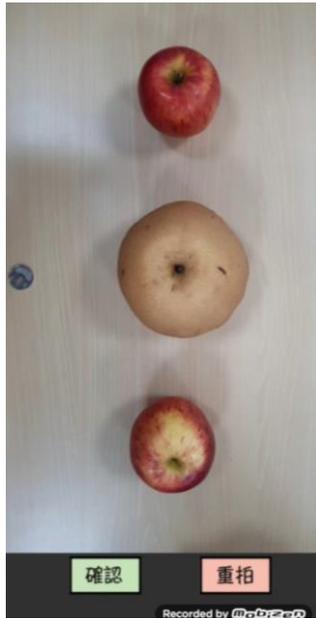


Figure 12 按下拍照後



Figure 13 server 回傳結果

GoGoFruit! APP 使用方法非常簡單，只要在畫面中放入一枚十元硬幣以及想要知道營養量的水果，將綠色圓框對準十元硬幣後按下拍照，就能得知水果的各種營養量了!

六、準確度分析

將拿 GoGoFruit!測量得之水果營養量，與秤重水果然後換算得到的真實營養量比對，發誤差基本上不超過5%，以下有一些對照圖。



The object is: apple
The total pixel: 18150
245.025



Figure 14 以蘋果比對準確度



The object is: kiwi
The total pixel: 11819
124.56162289999999



Figure 15 以奇異果比對準確度

可以看到 (Figure 14)、(Figure 15) 中以 GoGoFruit! 測量得到的誤差分別為 0.8 % 及 1.5 %，差距非常小。

七、未來展望

未來希望能再加強的有以下幾點：

1. 增加可辨識之水果種類：

目前可辨識的水果為 9 種，分別為 Apple、Banana、Guava、Kiwi、Papaya、Peach、Pear、Pineapple、Pitaya，希望之後能增加到至少 15 種以上。

2. 將硬幣也加入辨識：

增加硬幣的訓練資料到訓練集中，若是能辨識硬幣並算出其 pixel 數就能對照真實硬幣面積，也就不再需要對齊小圓框拍照了，在任何高度拍照都可以，讓操作更加直觀、便利。

八、參考文獻

- [1] 衛生福利部食品藥物管理署-食品藥物消費者專區-整合查詢服務：
<https://consumer.fda.gov.tw/Pages/List.aspx?nodeID=6>
- [2] Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick. Mask R-CNN. In *CVPR*, 2017. 3
- [3] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. In *CVPR*, 2015. 6