

火龍果與荔枝航照影像判釋－運用卷積神經網路 影像辨識技術與作物特徵萃取分類演算法

陳偉文^{1*} 卓柏漢² 林莉珊²

摘要

本研究透過深度學習的七種影像分類方法 (2D-CNN、AlexNet、VGG16、ResNet50、Inception-v1、Inception-v3、InceptionResNet-v2) 及兩種深度學習影像分割方法 (FCN 及 Mask R-CNN)，進行農作物判釋分類，藉以協助政府機關統計作物產量，解決產銷失衡等問題。在初期研究，以火龍果及荔枝作物的航空影像作為訓練母體樣本，除了探討深度自動學習影像特徵的方法外，也實作了傳統透過預先提取作物特徵，透過作物的紋理、形狀、色彩分布及 11 項植物植生指標特徵搭配隨機森林影像分類方法進行火龍果、荔枝及其他作物分類判釋。實驗結果顯示，以 CNN 為基礎架構設計的 InceptionResNet-v2 演算法有較佳準確度 (92.97%) 優於作物特徵萃取分類方法 (91.35%)。

關鍵詞：影像分類、影像分割、特徵萃取

1. 簡介

農業長期以來面臨到因氣候變遷造成產銷失衡、嚴重災損問題，另外，農地上常見非法使用之農地工廠等情況。這些問題一直牽絆政府機關農業部門，其中一個重要的原因，就是無法即時有效的掌握農地現況資訊。傳統透過現地調查的方式，進行農地普查以取得農地資訊，需要投入大量人力和物力，且缺乏資訊的即時性。對於短期作物的調查非常不利。透過大數據的資料收集與分析，需要研判資料的來源及數據的正確性，以及需建立許多資料篩選機制，以提高資料正確性。所需付出的成本相對也較高，也未必有好的分析結果。遙測影像技術已發展多年，從衛星影像、航空影像到無人機 (UAV) 影像，從低解析度到高解析度分析，可以完整呈現土地的地形、地貌。且拍攝距離越高，資料取得的時間越短，即時性越好。因此，如果善加利用遙測影像進行農地現況資訊的調查，會是一條穩健且可行的途徑。現今遙測影像的分析，多侷限在拍測距離較

近的無人機應用，主要原因是因為影像解析度較佳，較容易辨識和應用。例如：農作物災損面積調查和作物疾病防治等，屬於小區域特定用途應用。然而，對於全國性農業問題，如產銷失衡問題，便不容易透過無人機 (UAV) 的方式處理。因為所需調查的範圍較廣，且需要在短時間內取得所有的農地影像，以便進行後續判釋。本研究選用航空影像作為農作物判釋來源，主要由於航空影像的拍攝範圍較無人機 (UAV) 寬廣，可以短時間內取得大範圍的作物生長資訊，以輔助農業生產分析，平衡農業生產與銷售。不過，由於航空照片拍攝距離較高，使得影像解析度品質較差，加上還有雲遮，霧霾等外在因素影響，在進行物件判釋時，困難度較高。因此，希望透過一些傳統分析方法的加強以及人工智慧深度學習技術來輔助進行農作物航攝影像判釋。目前，國內進行農作物航空影像判釋主要還是依賴農業專家以人工檢視判釋居多，也有一定的準確率，但相對所需付出的人力、經費成本也較高，並且

¹ 農業科技研究院農業政策研究中心 副研究員

² 農業科技研究院農業政策研究中心 研究助理

* 通訊作者, 電話: 02-23681718 轉 505, E-mail: wwchen@mail.atri.org.tw

收到日期：民國 109 年 02 月 21 日

修改日期：民國 109 年 03 月 06 日

接受日期：民國 109 年 03 月 18 日

還有技術傳承的問題。在過去的研究中，也有專家學者提出以作物的植生指標進行農作物判釋，或是透過作物的紋理特徵進行判釋，但鮮少有研究以多種不同的作物特徵組合進行機器學習分類判釋，因此本研究特別針對作物的多種不同特徵進行特徵重要性分析以找出較佳特徵組合，藉以提升判釋率。此外，近年人工智慧深度學習技術發展快速，也有部分研究以 CNN 模型進行航照影像判釋分析，如：Kussul *et al.* (2017) 的研究中曾經提出以 2D-CNN 方法進行作物及土地分類。但也僅限於早期發展的 CNN 模型，為了研究後續發展出來的中大型網路模型及衍生出來的物件偵測、影像分割技術的判釋效益。

本研究另外實作出七種深度學習影像分類模型，包括 2D-CNN、AlexNet (Krizhevsky *et al.*, 2012)、Inception-v1 (Szegedy *et al.*, 2015)、Inception-v3 (Szegedy *et al.*, 2015)、VGG16 (Simonyan & Zisserman, 2015)、ResNet50 (He *et al.*, 2016)、InceptionResNet-v2 (Szegedy *et al.*, 2017)。此外，本研究也開發兩種影像分割、物件偵測技術的分類方法，分別為全卷積網路 (Fully Convolutional Networks, FCN) (Long *et al.*, 2015) 及遮罩區域卷積類神經網路 (Mask Region based Convolution Neural Networks, Mask R-CNN) (He *et al.*, 2017)，進行農作物語意分割 (Semantic Segmentation) 判釋。最後本研究精進傳統以作物特性的分類方法，嘗試找出最佳特徵分類組合，以提升判釋準確率，並與深度學習技術進行效能及優劣比較。

2. 研究材料

本研究使用農林航空測量所 (農航所) 所提供之 2016 年 11 月拍攝屏東、臺中、嘉義、南投、雲林、臺南、彰化、高雄等地的火龍果及荔枝作物航照影像 (表 1) 依地籍範圍切分每類作物 925 筆影像作為訓練及測試樣本，以 8:2 的比例取其中 740 筆影像做為訓練樣本，另外 185 筆影像做為測試及驗證樣本。

表 1 火龍果及荔枝航照影像資訊

項目	資訊
航照影像畫素 (尺寸)	11460 × 12260
水平解析度	96 dpi
重直解析度	96 dpi
地面解析度	25 cm
波段數	4
航照影像數量	125
涵蓋面積	1050 km ²

2.1 影像前處理

如上文提到，本研究採用兩種深度學習影像辨識技術，進行作物航照影像判釋。首先是影像分類 (Image Classification) 方法，另一項為影像分割 (Image Segmentation) 方法。深度學習影像分類方法以 CNN 卷積類神經網路架構為主，包含 2D-CNN、AlexNet、Inception-v1、Inception-v3、VGG16、ResNet50、InceptionResNet-v2。而深度學習影像分割方法主要採用 FCN 及 Mask R-CNN 技術。深度學習影像分類中除了 Inception v3 使用 299×299 的樣本影像大小，其餘方法皆使用 224×224 的影像大小樣本作為訓練及測試樣本，主要是參照訓練模型的初始設計建議。至於影像分割方法的運作原理是在一張 1024×1024 的航照影像圖中，先標示訓練物件範圍 (如：荔枝作物土地範圍)，建立訓練模型後，再以物件偵測 (Object Detection) 方式，從所有作物航照圖中，描繪出目標物件範圍。圖 1 為臺南市後壁區前菁寮火龍果種植地的航照影像，將原圖切分成兩類子圖，第一類按照地籍範圍切分給影像分類 (Image Classification) 方法如：2D-CNN、AlexNet 等演算法使用。第二類則依 1024×1024 像素大小切分給影像分割 (Image Segmentation) 方法如：FCN 以及 Mask R-CNN 使用。

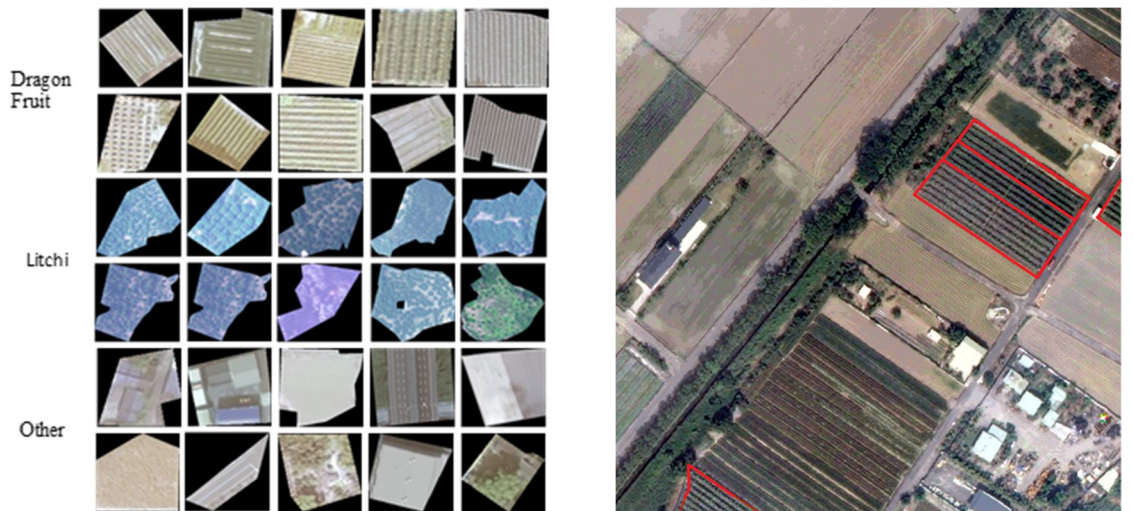
圖 2(a) 為依地籍切分的火龍果、荔枝及其他作物航照影像。圖 2(b) 中紅框的部份為標示火龍果作物 (物件) 區域。為避免影像轉換固定訓練樣本大小 (224×224 或 299×299)，造成變形、扭曲現象。透過 Keras 的 `preprocessing.image` 函式庫

ImageDataGenerator 去處理影像重塑 (reshape) 轉換。如圖 3 所示，按照地籍大小切分出原始作物影像 (499×367)，由於採用標準建議訓練影像大小 (考量影像品質及訓練時間)，需將影像重塑 (reshape) 或重製 (resize) 至兩種影像尺寸

(224×224、299×299)，需要先將影像進行填補的動作，再放大或縮小至建議訓練影像大小，進行後續模型訓練。ImageDataGenerator 函式庫提供原始/訓練影像轉換功能，免去額外處理 reshape 或 resize 動作。



圖 1 依地籍大小及 1024×1024 範圍切分影像分類及影像分割樣本



(a) 作物航照影像(按地籍切分) (b) 作物航照影像(依 1024×1024 大小切分)

圖 2 火龍果、荔枝及其他作物航照影像

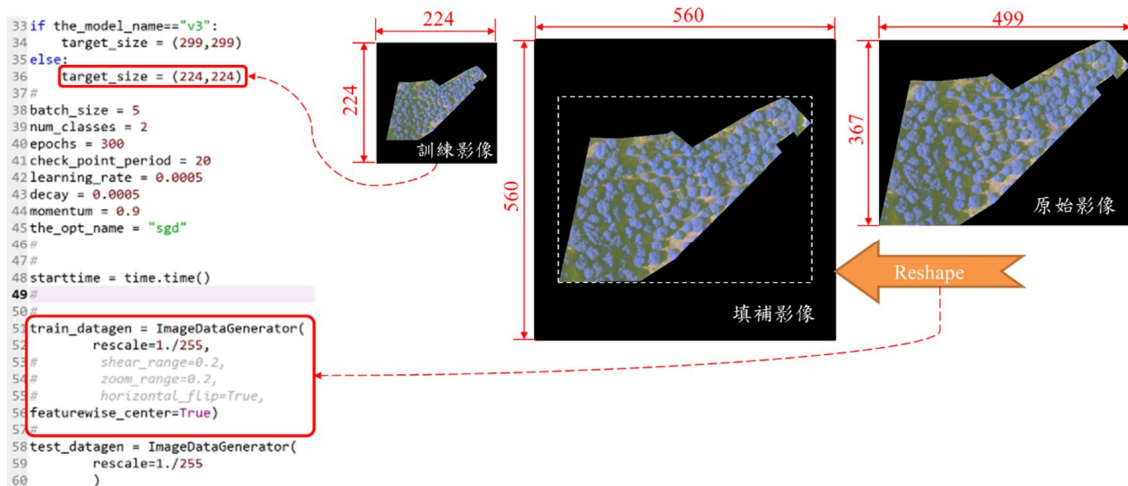


圖 3 原始影像 (地籍大小) 重塑至標準訓練影像大小 (224×224)

3. 研究方法

為了提升農作物航空影像的判釋準確率，研究採用兩種深度學習影像辨識技術及傳統透過作物特性的機器學習方法，進行火龍果及荔枝航照影像判釋。第一是透過深度學習影像分類方法，第二是使用深度學習影像分割方法，第三則是藉由作物的紋理、形狀大小、色彩分布及 11 種植物植生指標特徵搭配隨機森林 (Breiman, 2001) 分類器分類。

3.1 基本架構說明

圖 4 為基礎的 CNN 架構模式示意圖，用以說明以影像分類判釋法進行火龍果、荔枝及其他作物影像判釋架構。以不同層數的卷積層 (Convolution Layer)、池化層 (Pooling Layer) 及全連接層 (Fully Connected layer) 的組合，建構多層不同大小的網路進行影像訓練。這些網路模型，請參考引用文獻。首先將原始影像依地籍大小切分當作訓練、測試及驗證樣本，再透過 CNN

深度學習架構的七種演算法 (2D-CNN、AlexNet、VGG16、ResNet50、Inception-v1、Inception-v3、InceptionResNet-v2) 進行訓練和判釋。透過多層的卷積層、池化層網路連接全連接層進行分類，最後輸出各類別作物的判釋機率值。

圖 5 則是說明影像分割判釋方法的基本架構，首先，先將原始影像依照固定大小 (1024x1024) 分割成數張圖像，作為輸入訓練、測試及驗證樣本，提供給深度學習影像分割演算法 FCN 和 Mask R-CNN 使用。此類模型再依據作物的地籍範圍分割訓練物件，進行物件標註 (Labeling / Annotation)。將標註後產出的訓練物件 (json 格式)，進行模型訓練，最後在目標航照圖上，找出類似目標作物的位置。

最後的作物特性判釋方法(圖 6)，則是預先提取作物特徵的判釋分類方法，利用作物的紋理、大小形狀、色彩及 11 項植生指標向量空間建構訓練指標，再比較特徵值的重要性，選擇最佳特徵和特徵組合，提供給隨機森林分類器分類，提升判釋準確率。

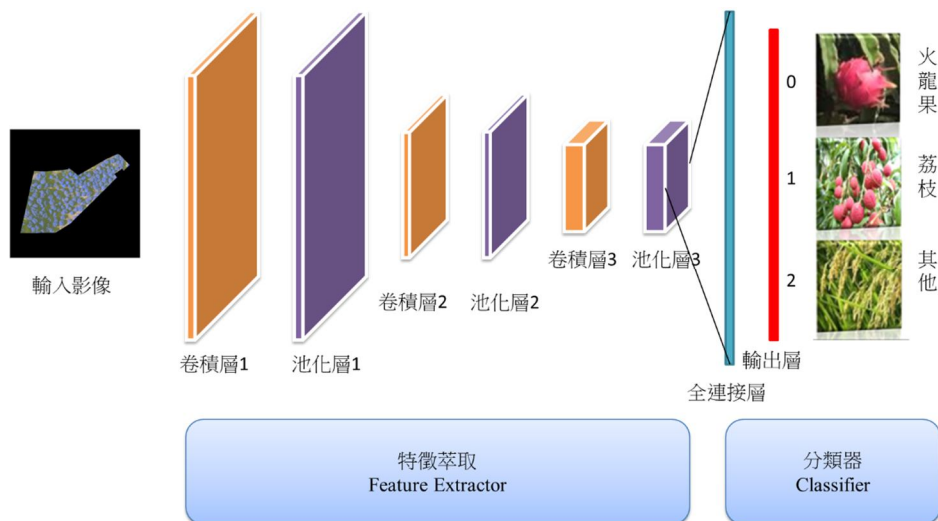


圖 4 農作物航照影像判釋架構示意圖 (影像分類判釋法)

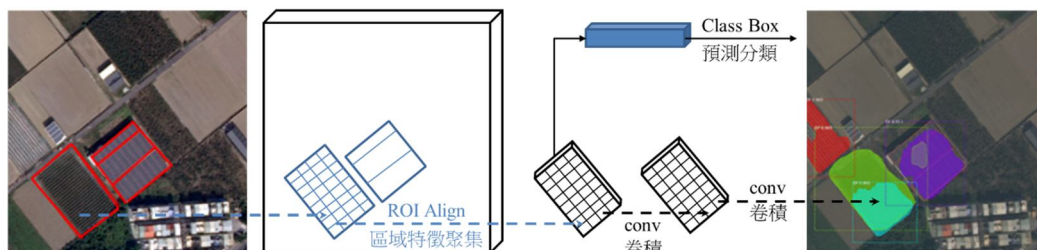


圖 5 農作物航照影像判釋架構示意圖 (影像分割判釋法)

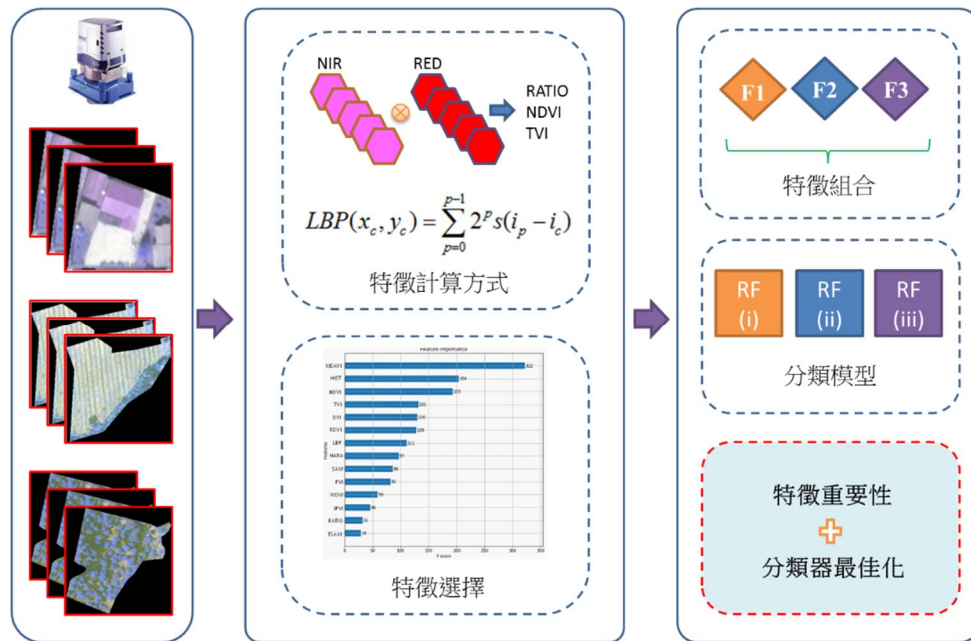


圖 6 農作物航照影像判釋架構示意圖 (作物特性判釋法)

3.2 深度學習 (卷積神經網路判釋法)

卷積神經網路 (Convolutional Neural Network) 在視覺識別上的表現令人稱奇。因為這些神經網路往往是帶有上百萬參數和許多隱含層的「龐然怪物」，其訓練深度可以從數層到數百層之多。像是 AlexNet, VGG, Inception 和 ResNet 等是近幾年所提出的 CNN 網路。影像分類傳統的作法分成兩個模塊，一個是特徵提取，另一個則是分類。特徵提取的方式有很多，諸如，前述中提到的紋理特徵、色彩特徵等。常用的分類器則有隨機森林 (Random Forest, RF) 和支持向量機 (Support vector machine, SVM)。這類作法有一個問題需要被討論，就是特徵提取不能根據圖像和其標籤進行調整。如果選擇的特徵缺乏一定的代表性來區分各個類別，模型的準確性就大打折扣，無論採用什麼樣的分類策略。

透過 CNN 深度學習訓練網路提取特徵的好處就是可以經由許多參數的訓練來識別圖像的特徵，再根據標籤數據進行分類。透過反覆性的參數調整，找出最佳的特徵提取演算法。唯一需要注意的是訓練過度擬合 (Overfitting) 的問題，不過這類問題已有許多檢測和防止方法可以預

防。以本研究為例，採用 Early Stopping 方法，在每一個 Epoch 結束時計算驗證集 (validation data) 的準確率，當準確率不再提高就停止訓練。

本研究透過使用 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 競賽中幾種具權威性的 CNN 架構模型作為判釋模型。從模型規模較小 (層數少) 的 AlexNet 到規模較大 (層數多) 的 ResNet 和 InceptionResNet。不同的作物在不同規模的訓練模型下，表現仍有差異。

本研究使用的七種深度學習方法主要差別在於訓練深度/層數的不同，其主要架構還是以 Convolution、Max Pooling 去建立 Feature Map 特徵分類的基礎。最後透過全連接層分類器的激活函式 (activation function) 算出機率最大的類別，作為判釋結果。

3.3 預先提取農作物特徵判釋方法

有別於傳統以人力判釋農作物的方式，本研究透過統計方法進行農作物航照影像特徵提取 (分述如下)，包含 11 項植物植被指數 (NDVI、TVI、IPVI……)、色彩分布 (Color Histogram)、紋理 (Haralick、Local Binary Pattern) 及形狀大小 (Hu moments)。最後，再透過隨機森林分類器進行判釋。

3.3.1 色彩特徵

本研究透過色彩直方圖 (Color Histograms) (Chapelle *et al.*, 1999) 統計各像素強度頻率的資訊來呈現整張影像的色彩分布情形。因此可以透過此資訊分析該圖像的物件特性。色彩直方圖可以針對任意色彩空間使用，例如：RGB 和 HSV 的三維色彩空間。而針對灰度圖像通常使用亮度直方圖。圖 7 為本研究針對火龍果影像所產生的色彩直方圖，將三個色彩空間影像作線性串接，合併成一整個大型色彩直方圖特徵以表示影像色彩資訊。每個顏色通道為 0 到 255 之間，以色彩空間域來表示時，同一個色彩空間的直方圖規模就會有 $256 \times 3 = 768$ 個值。因為不需要將顏色特徵取的太細小，所以本研究設定區塊數量參數 (Bins=8)，表示在 0 到 255 的值中，將範圍內切 8 等分，並且將影像轉換至 HSV 色彩空間 (Color space)。最後利用線性串接的方式，把所有的色彩直方圖串起來，將每個 bin 內的值視為單獨一維的特徵，可以得到 24 維度的色彩特徵。若要取得更多特徵，也可加入更多色彩空間 (RGB、HSV、HSL、YCbCr、LUV……等)，並將 Bin 值設高一點，可以得到更多維度的特徵資訊 (色彩空間數 color space \times 區塊數量 bins \times 通道數 channels)。

3.3.2 紋理特徵

紋理特徵用來描述物件表面的材質，量化比較材質表面的觸感、紋理、花紋。本研究採用了 Haralick *et al.* (1973) 和局部二值模式 (Local Binary Pattern, LBP) (Ojala *et al.*, 2002) 兩

種方式計算作物的紋理特徵。

Haralick *et al.* (1973) 從灰度共生矩陣 GLCM (grey-level co-occurrence matrix) 計算影像紋理在相對位置下像素間的灰階值關係，排成一機率分佈矩陣，由此矩陣可得知在某相對位置下像素對的灰階值關係為 i 和 j 出現的機率大小。再由 GLCM 算出以下 14 個特徵值 (表 2)。

$p(i, j | d, \theta)$ ：為在正規化的灰度共生矩陣中，灰階度值級數分別為 i 與 j 所對應到的元素值。其值的計算是由 $p(i, j | d, \theta) = p(i, j | d, \theta) / N$ 得來，其中 $p(i, j | d, \theta)$ 為正規化前的灰度共生矩陣中的元素值， N 為兩灰度值共同發生的總次數。

$p_x(i)$ ：為在正規化的灰度共生矩陣中的第 x 行的所有機率值總和，即 $p_x(i) = \sum_{j=1}^{N_g} p(i, j | d, \theta)$ ，其中 x 為矩陣中橫座標軸的值。

$p_y(j)$ ：為在正規化的灰度共生矩陣中的第 y 列的所有機率值總和，即 $p_y(j) = \sum_{i=1}^{N_g} p(i, j | d, \theta)$ ，其中 y 為矩陣中縱座標軸的值。

N_g ：為此影像的灰階度值級數；

$N_g = (\text{最大灰階度值} - \text{最小灰階度值}) + 1$

$p_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j | d, \theta)$ ，其中 $k = 2, 3, \dots, 2N_g$

$p_{x-y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j | d, \theta)$ ，其中 $k = 0, 1, \dots, (N_g - 1)$

```
def fd_haralick(image):
    # convert the image to grayscale
    gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
    # Compute the haralick texture feature vector
    haralick = mahotas.features.haralick(gray).mean(axis=0)
    return haralick

def fd_histogram(image, mask=None):
    bins = 8
    # convert the image to HSV color-space
    image = cv2.cvtColor(image, cv2.COLOR_BGR2HSV)
    # COMPUTE THE COLOR HISTOGRAM
    hist = cv2.calcHist([image], [0, 1, 2], None, [bins, bins, bins], [0, 256, 0, 256, 0, 256])
    # normalize the histogram
    cv2.normalize(hist, hist)
    # return the histog...
    return hist.flatten()

def fd_hu_moments(image):
    image = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
    feature = cv2.HuMoments(cv2.moments(image)).flatten()
    return feature
```

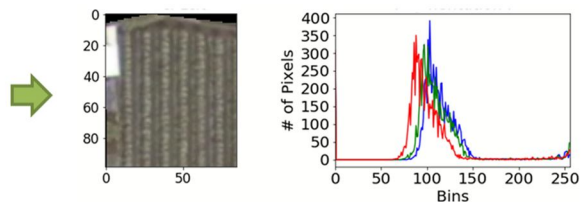


圖 7 火龍果色彩特徵

表 2 14 項 Haralick 紋理特徵計算公式

紋理特徵	計算公式	備註
二階度動量 (Angular Moment) Second	$F_1 = \sum_i \sum_j \{p(i, j d, \theta)\}^2$	
對比度 (Contrast)	$F_2 = \sum_{n=0}^{N_g-1} n^2 \times \left\{ \sum_{\substack{i=1 \\ i+j =k}}^{N_g} \sum_{j=1}^{N_g} p(i, j d, \theta) \right\}$	
相關性 (Correlation)	$F_3 = \frac{\sum_i \sum_j (i \times j) \times [p(i, j d, \theta) - \mu_x \mu_y]}{\sigma_x \sigma_y}$	其中， μ_x 及 μ_y 為 $p_x(i)$ 及 $p_y(j)$ 的平均數，而 σ_x 及 σ_y 為 $p_x(i)$ 及 $p_y(j)$ 的標準差。
平方和 (Sum of Squares)	$F_4 = \sum_i \sum_j (i - \mu)^2 \times p(i, j d, \theta)$	
逆差異動量 (Inverse Moment) Different	$F_5 = \sum_i \sum_j \frac{1}{1+(i-j)^2} \times p(i, j d, \theta)$	
總和平均 (Sum Average)	$F_6 = \sum_{i=2}^{2N_g} i \times p_{x+y}(i)$	
總和變異 (Sum Variance)	$F_7 = \sum_{i=2}^{2N_g} (i - F_6)^2 \times p_{x+y}(i)$	
和熵 (Sum Entropy)	$F_8 = -\sum_{i=2}^{2N_g} p_{x+y}(i) \times \log \{P_{x+y}(i)\}$	
熵值 (Entropy)	$F_9 = -\sum_i \sum_j p(i, j d, \theta) \times \log \{p(i, j d, \theta)\}$	
變異數差 (Difference Variance)	$F_{10} = \text{variance of } p_{x-y}$	
差熵值 (Difference Entropy)	$F_{11} = -\sum_{i=0}^{N_g-1} p_{x-y}(i) \times \log \{P_{x-y}(i)\}$	
相關性的資訊量測 (Information Measures of Correlation)	$F_{12} = \frac{HXY - HXY1}{\max \{HX, HY\}}$	HX 和 HY 分別是 P_x 與 P_y 的熵值。 $HXY = -\sum_i \sum_j p(i, j d, \theta) \times \log \{p(i, j d, \theta)\}$
	$F_{13} = (1 - \exp[-2.0 \times (HXY2 - HXY)])^{1/2}$	$HXY1 = -\sum_i \sum_j p(i, j d, \theta) \times \log \{p_x(i) \times p_y(j)\}$ $HXY2 = -\sum_i \sum_j p_x(i) \times p_y(j) \times \log \{p_x(i) \times p_y(j)\}$
最大相關係數 (Maximal Correlation Coefficient)	$F_{14} = (\text{Second largest eigenvalue of } Q)^{1/2}$	$Q(i, j) = \sum_k \frac{p(i, k) \times p(j, k)}{p_x(i) \times p_y(k)}$

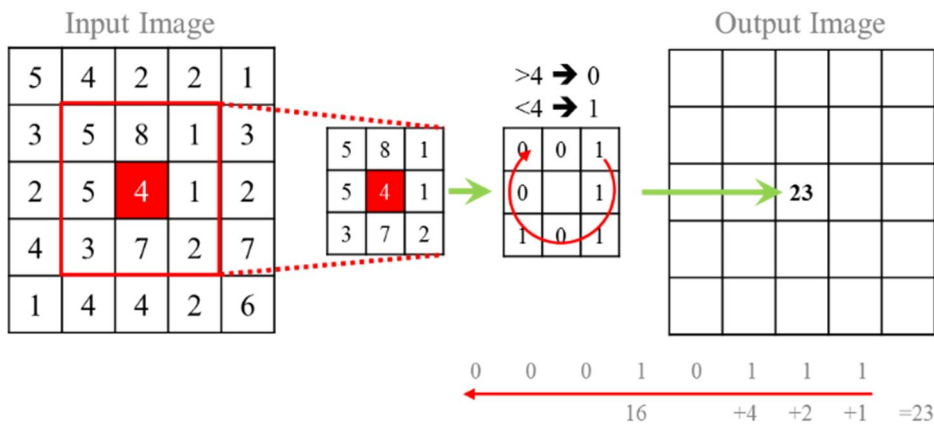


圖 8 LBP 特徵值計算方式

而 *LBP* 則是透過灰度不變和旋轉不變的特性，在物體的材質、樣式、甚至人臉辨識分析有突出的表現。使用方法是，先將圖像轉換為灰度圖，接着計算 *LBP* 特徵圖，最後計算其直方圖作為特徵向量。如圖 8 所示，對於某像素的 8 個鄰域，大於該像素值為 0，小於該像素值為 1，最後組成八位二進制碼。

LBP 可以被定義為式(1)：

$$LBP(x_c, y_c) = \sum_{p=0}^{p-1} 2^p s(i_p - i_c) \dots\dots\dots(1)$$

其中 (x_c, y_c) 是中心像素，亮度是 i_c ，而 i_p 則是相臨像素的亮度。

$$s(x) = \begin{cases} 0 & \text{if } x \geq 0 \\ 1 & \text{else} \end{cases} \dots\dots\dots(2)$$

共同加減數值其實並不會影響到整體計算的結果，這個特性形成了所謂的「灰度不變性」。例如，當光線偏亮或偏暗時，各像素的強度值會因光強度而共同增減一定數值，但對於計算結果不會有影響。

3.3.3 形狀特徵

Hu-moments (Hu, 1962) 可用來產生作物的形狀特徵，依據物件中心點計算七個 moments 特徵值。透過影像在經過旋轉、移動、縮放和鏡像等變形後這些數值仍保持不變的特性，來比較物件的相似度。Hu-moments 計算七個 moments 特徵值方式如下：

$$\begin{aligned} M_1 &= \mu_{20} + \mu_{02} \\ M_2 &= (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \\ M_3 &= (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{30})^2 \\ M_4 &= (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \\ M_5 &= (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})((\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2) + \\ &\quad (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})(3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2) \\ M_6 &= (\mu_{20} - \mu_{02})((\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2) + 4\mu_{11}(\mu_{30} + 3\mu_{12})(\mu_{21} + \mu_{03}) \\ M_7 &= (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})((\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2) - \\ &\quad (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})(3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2) \end{aligned}$$

為抵消尺度變化對中心矩的影響，用零階中心矩 η_{00} 對其餘各階中心矩進行歸一化處理，得到歸一化中心矩為式(3)：

$$\mu_{pq} = \frac{\eta_{pq}}{\eta_{00}^r} (r = \frac{q+p}{2}) \dots\dots\dots(3)$$

歸一化後具有尺度不變性。為使中心矩具有平移不變性，利用質心坐標建構中心矩

$$\eta_{pq} = \sum_{x=1}^C \sum_{y=1}^R (x - x_0)^p (y - y_0)^q f(x, y) \quad p, q = 0, 1, 2, \dots\dots$$

由零階原點矩和一階原點矩，可以求得目標區域的質心坐標為式(4)、式(5)：

$$x_0 = \frac{m_{10}}{m_{00}} \dots\dots\dots(4)$$

$$y_0 = \frac{m_{01}}{m_{00}} \dots\dots\dots(5)$$

0 階矩 (m_{00})：目標區域的質量

1 階矩 (m_{01}, m_{10})：目標區域的質心

2 階矩 (m_{02}, m_{11}, m_{20})：目標區域的旋轉半徑

3 階矩 ($m_{03}, m_{12}, m_{21}, m_{30}$)：目標區域的方位和斜度，反應目標的扭曲

$$m_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x, y) dx dy \quad p, q = 0, 1, 2, \dots\dots$$

影像概率密度分佈函數為 $f(x, y)$ 的二維連續隨機函數的 $p + q$ 階原點矩 m_{pq} 。所以影像的 hu 矩是一種具有平移、旋轉和尺度不變性的影像特徵。

3.3.4 植生指標特徵

利用作物的紅光和近紅外光反射的組合運算，可以顯示植物生長、生態系的活力與生產力等資訊。數值愈大表示植物生長愈多。常用的植生指標有 NDVI 常態化差值植生指標 (Normalized Difference Vegetation Index) (Payero *et al.*, 2004)，本研究中應用表 3 中 11 項表現植物的植被指數。

NIR 表示近紅外光波段，*RED* 表示紅外光波段，*L* 為常數通常給定 0.5， a_0 和 a_1 通常給定 0.0255 和 1.1351。*DVI* 為差值植被指數，適用於早 - 中期植被生長覆蓋度檢測，當植被覆蓋濃密度

(≥80%) 時，它對植被的靈敏度會下降。

4. 研究結果

4.1 深度學習 (卷積神經網路判釋法)

從表 4 可以明顯觀察到整體判釋率 (OA)，以 InceptionResNet-v2 的 92.97%最佳，但相對所需花費的訓練時間成本也較高。若單獨檢視荔枝

判釋率，小網路架構的 AlexNet (97.3%) 及 2D-CNN (96.22%) 也有不錯的表現，甚至優於更深層的網路架構。本次訓練的 Epoch 為 300 回，取訓練結果最佳的 Epoch 模型做為測試預測模型。

圖 9(a)顯示 FCN 訓練模型在火龍果作物判釋的結果，判釋準確率約在 70%左右。而 Mask R-CNN(圖 9(b))則可以到達 85%以上的判釋準確率，且能夠標示出疑似目標作物之區域，可提供二次判釋檢核之用。

表 3 11 項作物植生指數計算公式

植生指標	計算公式	參考文獻
Band ratio (RATIO)	$\frac{NIR}{RED}$	Tucker, 1979
Normalized difference vegetation index (NDVI)	$\frac{NIR - RED}{NIR + RED}$	Tucker, 1979
Transformed vegetation index (TVI)	$\sqrt{NDVI + 0.5}$	Tucker, 1979
Infrared percentage vegetation index (IPVI)	$\frac{NIR}{NIR + RED}$	Crippen, 1990
Perpendicular vegetation index (PVI)	$\frac{NIR - a_1 \times RED - a_0}{\sqrt{1 + (-a_1)^2}}$	Wiegand <i>et al.</i> , 1991
Soil adjusted vegetation index (SAVI)	$\frac{NIR - RED}{NIR + RED + L} \times (1 + L)$	Huete, 1988
Transformed soil adjusted vegetation index (TSAVI)	$\frac{a_1 (NIR - a_1 \times RED - a_0)}{RED + a_1 \times NIR - a_1 \times a_0}$	Wiegand <i>et al.</i> , 1991
Modified soil adjusted vegetation index (MSAVI)	$\frac{2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - RED)}}{2}$	Qi <i>et al.</i> , 1994
Difference vegetation index (DVI)	$NIR - RED$	Roujean & Breon, 1995
Renormalized difference vegetation index (RDVI)	$\sqrt{NDVI \times DVI} = \sqrt{\frac{(NIR - RED)^2}{NIR + RED}}$	Roujean & Breon, 1995
Weighted difference vegetation index (WDVI)	$NIR - a_1 \times RED$	Qi <i>et al.</i> , 1994

表 4 火龍果、荔枝及其他作物判釋準確率比較

訓練模型	判釋率 (%)				最佳時期 (Epoch)	訓練時間
	火龍果	荔枝	其他作物	整體準確率		
2D-CNN	82.16	96.22	81.62	86.67	287	4h51m
AlexNet	88.65	97.30	87.57	91.17	293	3h55m
VGG16	90.27	96.76	85.95	90.99	94	7h18m
ResNet50	85.95	92.97	89.73	89.55	260	5h33m
Inception-v1	86.49	96.76	90.81	91.35	171	4h25m
Inception-v3	89.19	96.22	85.95	90.45	281	9h27m
InceptionResNet-v2	94.05	94.59	90.27	92.97	128	10h58m

所有分析方法的訓練時間記錄在表 4，本研究使用的機器硬體規格為 CPU Intel Core i7

2.2GHz Coffee Lake 14nm, 32GB RAM, 雙 GPU, GPU0 是 Intel UHD Graphics 630, GPU1 是 4GB

NVIDIA GeForce RTX 2070 with Max-Q。InceptionResNet-v2 大型深層網路所需耗費的訓練成本最高。但這只是在一次的訓練過程所需的時間，一旦產出分析模型，將不需要太多的預測及判釋時間。

另外，針對個別類別進行二元混淆矩陣 (Confusion Matrix) 驗證，結果如表 5~7 所示，以二元混淆矩陣的 True Positive (TP)、False Negative (FN)、False Positive (FP) 及 True Negative (TN) 去分別計算 Positive predictive value (PPV / Precision)、True Positive Rate (TPR / Sensitivity / Recall)、False Negative Rate (FNR / Miss rate)、False Positive Rate (FPR / Fall-out)、True Negative Rate (TNR / Specificity)、False Discovery Rate (FDR)、False Omission Rate (FOR) 以及 Negative Predictive Value (NPV)。其中本研究特別關注

TPR 及 FNR 的結果，以火龍果為例，農業調查機關在進行火龍果農地判釋時，希望減少判釋的數量，但不能漏掉所有火龍果農地。本研究可以明顯看出 InceptionResNet-v2 對火龍果作物有較高的 TPR，而對於荔枝而言，則是以 AlexNet 表現的 TPR 較好。由此可知，雖然理論上深層大型網路模型的訓練判釋結果應該比較好，例如：火龍果的判釋結果以 InceptionResNet-v2 最佳。但這並非是不變的真理，因為當網路更多層時，更容易發生梯度消失或梯度爆炸的問題。梯度不穩定的現象常常會引起不穩定的表現，而且容易發生過度擬合 (overfitting)。對於較容易判釋的影像，過度複雜的函數，是極容易導致過度擬合的。這也表示為什麼對荔枝而言，小網路架構的 AlexNet 表現較大型更深層網路好。

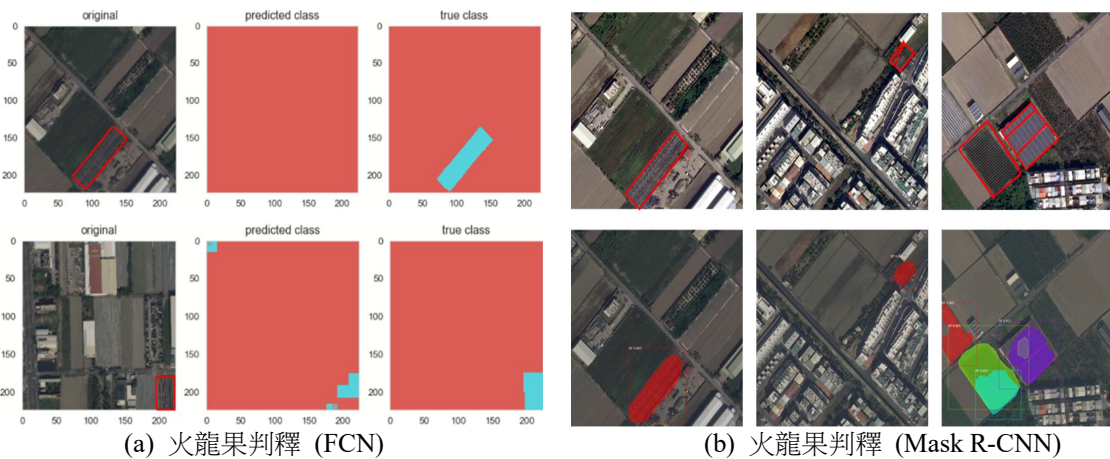


圖 9 火龍果判釋 (Image Segmentation)

表 5 火龍果混淆矩陣驗證指標

訓練模型	TP	FN	FP	TN	PPV (%)	TPR (%)	FNR (%)	FPR (%)	TNR (%)	FDR (%)	FOR (%)	NPV (%)
2D-CNN	152	33	33	337	82.16	82.16	17.84	8.92	91.08	17.84	8.92	91.08
AlexNet	164	21	19	351	89.62	88.65	11.35	5.14	94.86	10.38	5.65	94.35
VGG16	167	18	21	349	88.83	90.27	9.73	5.68	94.32	11.17	4.90	95.10
ResNet50	159	26	16	354	90.86	85.95	14.05	4.32	95.68	9.14	6.84	93.16
Inception-v1	160	25	14	356	91.95	86.49	13.51	3.78	96.22	8.05	6.56	93.44
Inception-v3	165	20	27	343	85.94	89.19	10.81	7.30	92.70	14.06	5.51	94.49
InceptionResNet-v2	174	11	20	350	89.69	94.05	5.95	5.41	94.59	10.31	3.05	96.95

粗體表示火龍果最佳訓練模型 TPR 數值

表 6 荔枝混淆矩陣驗證指標

訓練模型	TP	FN	FP	TN	PPV (%)	TPR (%)	FNR (%)	FPR (%)	TNR (%)	FDR (%)	FOR (%)	NPV (%)
2D-CNN	178	7	12	358	93.68	96.22	3.78	3.24	96.76	6.32	1.92	98.08
AlexNet	180	5	10	360	94.74	97.30	2.70	2.70	97.30	5.26	1.37	98.63
VGG16	179	6	12	358	93.72	96.76	3.24	3.24	96.76	6.28	1.65	98.35
ResNet50	172	13	14	356	92.47	92.97	7.03	3.78	96.22	7.53	3.52	96.48
Inception-v1	179	6	8	362	95.72	96.76	3.24	2.16	97.84	4.28	1.63	98.37
Inception-v3	178	7	4	366	97.80	96.22	3.78	1.08	98.92	2.20	1.88	98.12
InceptionResNet-v2	175	10	3	367	98.31	94.59	5.41	0.81	99.19	1.69	2.65	97.35

粗體表示荔枝最佳訓練模型 TPR 數值

表 7 其他作物混淆矩陣驗證指標

訓練模型	TP	FN	FP	TN	PPV (%)	TPR (%)	FNR (%)	FPR (%)	TNR (%)	FDR (%)	FOR (%)	NPV (%)
2D-CNN	151	34	29	341	83.89	81.62	18.38	7.84	92.16	16.11	9.07	90.93
AlexNet	162	23	20	350	89.01	87.57	12.43	5.41	94.59	10.99	6.17	93.83
VGG16	159	26	17	353	90.34	85.95	14.05	4.59	95.41	9.66	6.86	93.14
ResNet50	166	19	28	342	85.57	89.73	10.27	7.57	92.43	14.43	5.26	94.74
Inception-v1	168	17	26	344	86.60	90.81	9.19	7.03	92.97	13.40	4.71	95.29
Inception-v3	159	26	22	348	87.85	85.95	14.05	5.95	94.05	12.15	6.95	93.05
InceptionResNet-v2	167	18	16	354	91.26	90.27	9.73	4.32	95.68	8.74	4.84	95.16

粗體表示其他作物最佳訓練模型 TPR 數值

表 8 15 項農作物特徵指標判釋準確度比較

特徵指標 \ 作物	火龍果(%)	荔枝(%)	其他作物(%)	整體準確率(%)
MSAVI	92.43	97.84	78.92	89.73
HIST	85.41	96.22	84.86	88.83
NDVI	91.35	92.97	78.92	87.75
TVI	89.73	92.43	78.92	87.03
DVI	83.78	95.14	81.08	86.67
RDVI	87.03	92.43	75.68	85.05
LBP	78.38	95.14	72.43	81.98
Hara	76.22	83.78	85.41	81.80
SAVI	84.32	90.27	67.57	80.72
PVI	81.08	78.92	81.08	80.36
WDVI	82.16	76.22	81.08	79.82
IPVI	74.05	83.24	80.00	79.10
RATIO	64.32	82.16	80.54	75.68
TSAVI	61.62	78.38	78.92	72.97
Hu	28.11	77.30	24.86	43.42

粗體表示 MSAVI 特徵指標有最佳整體判釋準確率

表 9 不同特徵組合指標判釋準確度比較

特徵指標 \ 作物	火龍果(%)	荔枝(%)	其他作物(%)	整體準確率(%)
MSAVI + Hi	94.05	99.46	80.00	91.17
MSAVI + Hi + NDVI	92.97	99.46	81.62	91.35
MSAVI + Hi + NDVI + TVI	92.97	99.46	79.46	90.63

粗體表示 MSAVI、Histogram 及 NDVI 的特徵指標組合有最佳整體判釋準確率

4.2 預先提取農作物特徵判釋方法

為了檢視 15 項農作物特徵指標對判釋影響的強弱程度，我們個別進行隨機森林分類訓練，得到如表 8 的結果。以整體判釋準確度而言，MSAVI 植生指標的表現較好，判釋準確度 89.73%。其次是 Color Histogram 的 88.83%。MSAVI 是 Qi *et al.* (1994) 提出的概念，是 SAVI 的改良版。其優點是增加植被信號的動態範圍，同時進一步減少土壤背景的影響，提供更好的植被靈敏度。其次表現較好的色彩直方圖 (Color Histogram) 代表 RGB 色彩的強弱分佈情況，一般而言，這是最直接、最基本有用的特徵表現。為了提升作物的判釋準確率，本研究嘗試透過表現較佳的前幾個重要特徵指標組合 (MSAVI, HIST)、(MSAVI, HIST, NDVI) 及 (MSAVI, HIST, NDVI, TVI)，進行隨機森林分類器訓練，實驗結果如表 9 所示。整體準確率往上提升至 91.35%。

5. 結論

本研究以火龍果及荔枝作物為例，透過 CNN 架構的深度學習影像分類及影像分割方法進行航照影像判釋，並且設計精進傳統以作物特性為特徵預先萃取分類法，進行比較。結果顯示以 CNN 深度學習分類方法中的 InceptionResNet-v2 模型表現最佳，整體判釋準確率 (Overall Accuracy) 為 92.97%。另外，透過本研究設計的以作物特徵預先萃取分類方法，也有不錯的判釋率 (91.35%)。透過深度學習的訓練方法，有一項優點就是不需額外設計特徵擷取機制，反觀，透過預先設計好的特徵值提取分類方法，則需要花費較多時間研究特徵值的表現，要找到好的特徵萃取方法並不容易。而透過影像分割的物件偵測方法，Mask R-CNN 也有不錯的表現，對於初判時漏判的物件偵測或二次判讀檢核也有相當的貢獻。本研究的成果，未來可應用推廣到全國土地分類判釋，將土地以階層式分類方式，從農、林、

漁、畜牧、公共建設的主要分類，往下細分到單一作物，甚至到不同生長時期的作物產量評估。

誌謝

本研究由農委會計畫 COA 1082101011919-200101-i5 補助支持，特此誌謝。

參考文獻

- Breiman, L., 2001. Random forests, *Machine Learning*, 45(1): 5-32.
- Chapelle, O., Haffner, P., and Vapnik, V., 1999. Support vector machines for histogram-based image classification, *IEEE Transactions on Neural Networks*, 10(5): 1055-1064.
- Crippen, R.E., 1990. Calculating the vegetation index faster, *Remote Sensing of Environment*, 34(1): 71-73.
- Haralick, R.M., Shanmugam, K., and Dinstein, I., 1973. Textural features for image classification, *IEEE Transactions on Systems, Man, and Cybernetics*, SMC3(6): 610-621.
- He, K., Gkioxari, G., Dollár, P., and Girshick, R., 2017. Mask R-CNN, *The 2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 2980-2988.
- He, K., Zhang, X., Ren, S., and Sun, J., 2016. Deep residual learning for image recognition, *The 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp.770-778.
- Hu, M., 1962. Visual-pattern recognition by moment invariants, *IRE Transactions on Information Theory*, 8(2): 179-187.
- Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI), *Remote Sensing of Environment*, 25(3): 295-309.

- Krizhevsky, A., Sutskever, I., and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, The 25th International Conference on Neural Information Processing Systems(NIPS), Lake Tahoe, NV, USA, pp.1106-1114.
- Kussul, N., Lavreniuk, M., Skakun, S., and Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data, IEEE Geoscience and Remote Sensing Letters, 14(5): 778-782.
- Long, J., Shelhamer, E., and Darrell, T., 2015. Fully convolutional networks for semantic segmentation, The 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp. 3431-3440.
- Ojala, T., Pietikäinen, M., and Mäenpää, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions Pattern Analysis and Machine Intelligence, 24(7): 971-987.
- Payero, J.O., Neale, C.M.U., and Wright, J.L., 2004. Comparison of eleven vegetation indices for estimating plant height of alfalfa and grass, Applied Engineering in Agriculture, 20(3): 385-393.
- Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., and Sorooshian, S., 1994. A modified soil adjusted vegetation index, Remote Sensing of Environment, 48(2): 119-126.
- Roujean, J.L., and Breon, F.M., 1995. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements, Remote Sensing of Environment, 51(3): 375-384.
- Simonyan, K., and Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition, The 2015 International Conference on Learning Representations (ICLR), San Diego, CA, USA.
- Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A.A., 2017. Inception-v4, inception-Resnet and the impact of residual connections on learning, The 31th Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, pp.4278-4284.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., 2015. Going deeper with convolutions, The 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp.1-9.
- Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation, Remote Sensing of Environment, 8(2): 127-150.
- Wiegand, C.L., Richardson, A.J., Escobar, D.E., and Gerbermann, A.H., 1991. Vegetation indexes in crop assessments, Remote Sensing of Environment, 35(2-3): 105-119.

Dragon Fruit and Litchi Interpretation from Aerial Photographs - Using CNN Network and Crop Feature Extraction Classification Algorithms

Wei-Wen Chen ¹ Bo-Han Cho ² Li-Shan Lin ²

Abstract

The research undergoes crop recognition and classification through seven CNN architecture models: 2D-CNN, AlexNet, VGG16, ResNet50, Inception-v1, Inception-v3, and InceptionResNet-v2 to help the government agency in tabulating the crop production and solve the imbalance between production and sales. Besides, our research also designed the imagery classification of the machine learning method by pre-extracting crop's texture, shape and size, color, and 11 vegetation indices features. We compared the advantages and disadvantages between the two. The result of the experiment shows that the training model with CNN as the basic architecture is better than the traditional imagery classification method.

Keywords: Image classification, Image segmentation, Feature extraction

¹ Associate Research Fellow, Agricultural Policy Research Center, Agricultural Technology Research Institute

² Research Assistant, Agricultural Policy Research Center, Agricultural Technology Research Institute

* Corresponding Author, Tel: 886-2-23681718 ext.505, E-mail: wwchen@mail.atri.org.tw

Received Date: Feb. 21, 2020

Revised Date: Mar. 06, 2020

Accepted Date: Mar. 18, 2020