

國立雲林科技大學工業工程與管理系

碩士論文

Department of Industrial Engineering and Management

National Yunlin University of Science & Technology

Master Thesis

應用集成式深度學習模型進行芒果分類辨識

Recognize the Classes of Mango by Using Ensemble Deep
Learning Models

許峻棠

Jun-Tang Xu

指導教授：蘇純增 博士

Advisor: Chwen-Tzeng Su, Ph.D.

中華民國 110 年 6 月

June 2021

摘要

愛文芒果是台灣農業外銷的一大主力，每年為台灣地區帶來龐大的利益。然而，隨著各國農業技術的進步，各國都開始生產高品質的愛文芒果，並展開外銷的價格戰，最終對台灣的外銷產業造成極大威脅。現階段台灣芒果農主要是以降低生產、包裝與運送成本來應戰。但在整個成本削減過程中，果農目前最苦惱的是芒果等級辨識的成本，因為這個工作不僅繁複也十分精細，過往都必須靠大量的人力才能完成，並因此產生大量的人力成本。為了克服這個問題，本論文提出一套新的人工智慧辨識框架來讓機器自動化的辨識芒果等級。該框架包含幾大步驟，包含(1)將原始圖像透過 Mask R-CNN 找出芒果位置並清除背景雜訊，(2)分別將第一步驟所獲取的圖像進行圖像模糊化、圖像平均化與轉換至 HSV 色彩空間，並得到三種類的芒果圖像資料集，(3)將三種資料集分別輸入三個不同的卷積神經網路訓練，並得到三個模型，以及(4)將第三步驟獲得的三個卷積神經網路結果與第一步驟的結果進行集成的動作。藉由上述的方式，我們預期能以高精準度的方式進行芒果等級的自動化辨識。最終，我們也使用了真實的芒果圖像資料集來驗證我們所提出方法的有效性。

關鍵字：Mask R-CNN、卷積神經網路、集成式學習、影像辨識、芒果

Abstract

The agricultural product export of Aiwen mango is Taiwan main force. Every year it brings huge benefits to Taiwan. However, with the advancement of agricultural technology in various countries, they began to produce high-quality Aiwen mangoes and started price reduction war between countries, which ultimately posed a huge threat to Taiwan's export industry. In order to gain a certain advantage at this stage, Taiwan mango farmers began to reduce production, packaging and transportation costs. But in the whole process of cost reduction, farmers are most worried about the cost of mango grading, because this work is not only complicated, but also requires skill. In the past, it required a lot of labor to complete, causing a lot of labor costs. To overcome this problem, in this paper we propose a new artificial intelligence identification framework to allow machines to automatically identify mango grades. The framework consists of several major steps, including (1) the background noise of original image removed and find the position of the mango through Mask R-CNN. (2) The image obtained in the first step divided into blur, image average, converted to HSV color space to obtain three different type of mango image datasets, (3) three datasets become input of three different convolutional neural network, training and obtain three different models, (4) Ensemble the results of the three different models with the results from first step. Finally, we expect to be able to perform automatic identification of mango grades with highly accurate .Beside that, we also used real mango image datasets to verify the effectiveness of our proposed method.

Keywords : **Mask RCNN** 、**Convolution Neural Network** 、**Ensemble Learning** 、
Image Recognition 、**Mango**

致謝

首先要特別感謝我的指導教授蘇純增老師，也感謝陳奕中老師，在我雲科大碩士這段期間的指導，由其是覺得自己遇到瓶頸，感覺人生跌到谷底並拉我一把的時候，若是當初在碩一升碩二時，沒有老師的支持與鼓勵，我覺得我走不到今天這一步，對於老師的關懷與耐心是當時支持著我前進的動力，也謝謝老師在碩士這段期間對於人工智慧領域、如何跟廠商溝通與如何解決目前遇到的問題的指導與教授，最後也謝謝老師給予的肯定。

再來要感謝教育部人工智慧競賽與標註資料蒐集中心給予芒果資料集的使用許可，讓我順利完成碩士論文。

特別感謝實驗室的同學，柏翔與吉晨，認識這兩位同學是我念碩士以來最大的收穫，在程式方面遇到困難時可以一直耐心的教導我寫程式，一直給予我幫助，並適時的鼓勵我與陪伴我，真的非常感謝他們，這份論文的完成若是沒有他們的教導與指教，可能也沒有辦法完整的寫完。感謝姿靜在這段期間一直互相扶持與鼓勵，謝謝你跟我一起進實驗室，在我遇到困難時，總覺得有你在跟我一起努力著，你也是功不可沒的一員。謝謝姿璇在我忙著寫論文的時候，可以幫我分擔管理學助教的工作，這兩年的助教妳真的辛苦了。謝謝詠緹，只要我有甚麼問題拿去問妳，妳都可以很詳細的告訴我怎麼做，真的很感謝有妳的幫忙。謝謝文成，在我最後衝刺論文的時候給予的幫忙，也讓我在碩論的完成上更順利。

也謝謝東錡學長、映均學姐、富正大哥、豪哥、學長姐們在這段期間的幫助與教導，暑期還空出時間來協助學弟妹，辛苦你們了，也謝謝實驗室所有同屆的同學在這兩年的討論與交流，讓我受益很多。

最後謝謝所有在碩士這兩年幫助過我的人，謝謝你們的扶持成就了現在這個我，對於你們的幫忙，我由衷感謝。

目錄

摘要	ii
Abstract.....	iii
目錄	v
表目錄	vii
圖目錄	viii
第一章 緒論	1
1.1 研究背景與動機	1
1.2 研究目的	3
1.3 研究架構流程圖	5
第二章 文獻探討	6
2.1 圖像前處理	6
2.1.1 傳統目標檢測的方法	6
2.1.2 深度學習目標檢測的方法	7
2.1.3 圖像增強處理	10
2.2 卷積類神經網路(Convolutional Neural Network , CNN)	11
2.3 集成式學習(Ensemble Learning)	13
第三章 研究方法	16
3.1 圖像資料集介紹與資料清洗	16
3.2 Mask R-CNN 應用於芒果辨識的作法	18
3.2.1 檢測實物區域	18
3.2.2 Mask R-CNN 模型之完整步驟	18
3.3 圖像增強處理應用於芒果辨識的作法	20
3.3.1 模糊化處理	20
3.3.2 平均色處理	21
3.3.3 轉換 HSV 處理	21
3.4 卷積神經網路應用於芒果辨識的作法	22
3.4.1 深度可分離卷積(Depthwise Separable Convolution)	23

3.5 集成式學習(Ensemble Learning)應用於芒果辨識的作法	26
3.5.1 Stacking	26
第四章 實驗模擬	29
4.1 參數設定與資料集介紹	29
4.2 前處理之模型有效性	29
4.2.1 以目標檢測模型為例	30
4.2.2 以圖像增強處理模型為例	32
4.3 集成學習之模型有效性	33
4.4 本研究實驗之混淆矩陣	34
4.5 實驗質化合理性說明	35
第五章 結論與未來研究	40
參考文獻	41

表目錄

表 2-1 傳統目標檢測方法文獻整理	7
表 2-2 深度學習目標檢測方法文獻整理	9
續表 2-2 深度學習目標檢測方法文獻整理	10
表 2-3 圖像增強處理文獻整理	11
表 2-4 卷積類神經網路文獻整理	12
續表 2-4 卷積類神經網路文獻整理	13
表 2-5 集成式學習文獻整理	15
表 3-1 芒果圖像資料欄位	17
表 4-1 目標檢測方法之績效比較	31
表 4-2 圖像增強模型績效	32
表 4-3 集成學習使用之分類器績效	34
表 4-4 目標檢測模型其混淆矩陣	34
表 4-5 圖像增強模型其混淆矩陣	35
表 4-6 集成深度學習模型其混淆矩陣	35

圖目錄

圖 1-1 好辨識之芒果圖像	3
圖 1-2 不好辨識之芒果圖像	3
圖 2-1 卷積類神經網路架構圖[65]	12
圖 3-1 研究流程圖	16
圖 3-2 A 類芒果圖像	17
圖 3-3 B 類芒果圖像	17
圖 3-4 C 類芒果圖像	17
圖 3-5 Mask R-CNN 檢測實物區域架構圖	18
圖 3-6 Mask R-CNN 架構圖	19
圖 3-7 本研究 Mask R-CNN 去除背景雜訊示意圖	20
圖 3-8 3×3 的權重矩陣 [21]	21
圖 3-9 Mask R-CNN 遮罩示意圖	21
圖 3-10 本研究 Xception 架構圖	23
圖 3-11 Depthwise 卷積架構	24
圖 3-12 Pointwise 卷積架構	24
圖 3-13 池化層 MaxPooling2D	25
圖 3-14 展平層	26
圖 3-15 基於 Stacking 之集成式學習概念圖[26]	26
圖 3-16 本研究集成式深度學習架構	27
圖 3-17 SVM 三維空間示意圖	27
圖 3-18 決策樹架構圖	28
圖 4-1(a)(b)(c)(d)為背景複雜之芒果圖像	30
圖 4-2 芒果圖像 Grad-CAM 示意圖	30
圖 4-3 原圖經由 Mask R-CNN 處理	31
圖 4-4 原圖經由 YOLO_V3 處理	31
圖 4-5 Mask R-CNN 之處理流程示意圖	32
圖 4-6 轉換至 HSV 色彩空間處理示意圖	33

圖 4-7 圖像模糊化處理示意圖	33
圖 4-8 圖像平均色處理示意圖	33
圖 4-9 Grad-CAM 以 00924.jpg 為例使用平均色模型，A 類被分為 A 類	36
圖 4-10 Grad-CAM 以 03064.jpg 為例使用 Mask R-CNN 模型，A 類被分為 B 類 .	37
圖 4-11 Grad-CAM 以 05154.jpg 為例使用 HSV，A 類被分為 C 類.....	37
圖 4-12 Grad-CAM 以 06330.jpg 為例使用 Mask R-CNN 模型，B 類被分為 A 類 .	37
圖 4-13 Grad-CAM 以 07755.jpg 為例使用模糊化模型，B 類被分為 B 類.....	38
圖 4-14 Grad-CAM 以 03102.jpg 為例使用平均色模型，B 類被分為 C 類.....	38
圖 4-15 Grad-CAM 以 02272.jpg 為例使用 HSV，C 類被分為 A 類.....	39
圖 4-16 Grad-CAM 以 01735.jpg 為例使用模糊模型，C 類被分為 B 類.....	39
圖 4-17 Grad-CAM 以 02844.jpg 為例使用 Mask R-CNN 模型，C 類被分為 C 類..	39

第一章 緒論

1.1 研究背景與動機

近年來隨著人工智慧的發展，將人工智慧精準化的技術導入於農業中，融合自動化技術、物聯網技術，為傳統農業賦予了新的技能，這些技能不僅可以提高生產效能，也能改善農業人口短缺所造成的問題以及農產品所得收入，加上愛文芒果目前為台灣重要出口年產品，不僅在近年躍升為三大外銷高經濟生鮮水果產品之一，外銷通路甚至拓寬到日本、中國、韓國以及香港等地，根據統計 100 年 1 月至 108 年 12 月止，生鮮芒果約占生鮮出口總值前三名，109 年生鮮芒果出口至各國甚至成長了 17%[67]。

目前國際上愛文芒果的銷量提升，是以削價競爭為主要手段，因此若想提高台灣愛文芒果在國際上的優勢就得善用科技輔助，經詢問業者後，得知目前最有可能大幅降低成本的是採收後芒果品質的辨識，目前台灣愛文芒果在品質控管的方式下分為 A、B、C 三類，A 類為國外出口用；B 類為國內銷售用；C 類為食品產業加工用，目前由於缺乏新的技術及自動化的設備，導致廠商的篩選水果的效率及品質都不盡理想，在篩果廠時，如果篩果不夠精準，太多的不良品跑到蒸熱場，蒸熱場的不良率就會過高，因此人力的部分就會卡在篩果，造成生產及包裝產線上的人力會不足，進而會影響到廠商整體的出貨進度，若以外銷金額來說，每年將損失 16000 萬台幣的收入[63]。

為了解決以上問題，業主開始尋求自動化概念的 AI 影像辨識技術，若是要以人工智慧之技術導入全自動化水果分級來說，在果園進行全天監控處理，經由即時影像判斷水果是否成熟，並且由機器手臂將水果摘取放置到輸送帶，再透過即時攝影將水果進行品質分級，目前台灣是以人工篩選的方式來進行芒果品質分級，因此本研究也將針對採收後處理技術的品質分級做後續研究。

根據過往傳統對於水果表面品質影像分級等研究大多以機器學習達成，舉例來說，Zarifneshat *et al.*[58]發現人工神經網路(ANN)的技術已被評估為預測蘋果瘀傷的替代方法，不需要使用人工來檢測；Moallem *et al.*[30]比較各分類器，將蘋果圖像分類為健康或缺陷，並且將蘋果分級，但在背景複雜的情況下，皆無法準確提取目標特徵。Fukuda *et al.*[8]學者證明了隨機森林(Random Forest)使用芒果果皮

顏色來評估芒果果肉品質，可以獲得較好精度與實用性。Yu *et al.*[56]提出一種基於多圖像特徵與 K-means 的演算法，來實現蘋果自動分級。Azarmdel *et al.*[1]提出使用人工神經網路(Artificial Neural Network，ANN)與支持向量機(Support Vector Machine，SVM)對水果進行分類，其證實此模型適合在實際應用上。但由於傳統機器學習對於複雜圖像其分類效果不好，機器學習較可處理單一問題，無法處理太複雜之問題為機器學習之缺點。

近年來深度學習模型為科學界中最受矚目的方法之一，大量的應用在各個領域，也包含了本論文所探討水果品質影像分級之研究，由於對於不同圖像會有不同之處理方式，自從 Simonyan *et al.*[45]學者使用卷積神經網絡(Convolution Neural Network，CNN)在 2014 年 ImageNet 大規模視覺識別挑戰賽(ILSVRC)上贏得比賽，在當時卷積神經網絡占據了圖像分類任務的巨頭，並且相關模型開始釋出，眾多學者開始專注於朝向可以主動提取圖像特徵的深度學習之影像分級開始做研究，大概分為兩個方向，包含(1)透過後選區域/框加上深度學習分類，通過提取候選區域(2)基於深度學習的迴歸方法。第一個方向對於影像處理在目標檢測之方法大多數學者會使用後選區域/框加上深度學習分類，通過提取候選區域，並對相應區域進行以深度學習方法為主的分類的方法，舉例來說 Girshick 學者[11]所提出的 Fast Region Convolution Neural Network、Ren *et al.* 學者[10]所提出的 Faster Region Convolution Neural Network，可以在同張圖片快速提取不同目標等方法，可以同時輸出個別目標框。第二個方向為基於深度學習的迴歸方法，提出物體分類和物體定位在一個步驟中完成，舉例來說 Redmon *et al.* [11]提出 You Only Look Once(YOLO)與 Liu *et al.* [12]提出 Single Shot MultiBox Detector(SSD)，大降低了目標檢測的耗時，提高了實時性，上述方法是從 2014 年至目前陸續被重視等目標檢測之方法，也各大領域中皆被證實深度學習主動提取特徵來代替人為設計，較大地提高了精度和效率。過去學者在應用於芒果品質檢測問題時，有遇到困難點，舉例來說，好辨識之芒果圖像如圖 1-1，可以看到圖像背景單調且無其他雜訊，模型在學習上不會受雜訊干擾；不好辨識之芒果圖像如圖 1-2，可以看到圖像背景有其他目標物與雜訊，會造成模型學習狀況不佳，學習到其他雜訊，若是針對不好辨識之芒果圖像進行分類辨識會遇到兩大問題(1)使用候選框來做目標檢測，方框內之目標的背景也同時會被候選框抓取，若是背景複雜或目標有重疊，因而

被背景雜訊所影響，造成精度不佳等問題。(2)將未經過處理的圖像直接輸入至模型中，使模型在訓練過程中消耗大量的運算成本，造成運算時間過長與精度不佳等不符實務上的應用。



圖 1-1 好辨識之芒果圖像



圖 1-2 不好辨識之芒果圖像

1.2 研究目的

本研究將提出幾個概念解決上述問題：對於問題(1)提出解決概念(a)透過 Mask R-CNN 提取目標輪廓。對於問題(2)提出兩點解決概念(b)使用多種前處理方式，提高模型強度與精度。(c)將多個模型透過集成式學習(Ensemble Learning)重新訓練結果，以下將詳細說明：(a)透過 Mask R-CNN 提取目標輪廓，針對上述第一點問題，因考慮到本研究所使用之資料集背景會有其他芒果，導致使用其他目標檢測之方法無法實際達到本研究所需，然而透過 Mask R-CNN 可以直接將輪廓提取，去掉不必要的背景讓模型不會學習到其他雜訊，因此本研究將會使用 Mask R-CNN 來清除背景雜訊。(b)使用多種前處理方式，提高模型強度與精度，針對上述第二點問題，由於將未經處理之圖像直接輸入模型，會因圖像缺陷特徵不明顯與圖像背景複雜，讓模型無法學習到更多的特徵資訊，因此本研究將引入

HSV(Hue,Saturation,Value)顏色空間並輸入簡易模型進行分級，調整圖像明暗程度、飽和度與色調來突顯芒果圖像的表面缺陷；使用圖像模糊化處理並輸入簡易模型進行分級，將邊緣雜訊模糊處理，降低雜訊對於模型之影響；透過圖像模糊平均色處理並輸入簡易模型進行分級，來突顯果實成熟度的平均顏色讓模型學習。(c)針對上述第二點問題，本研究使用多種方法來進行集成式學習(Ensemble Learning)並重新訓練分級結果，舉例來說，同一張圖片經由不同方法後有不同之結果，若將不同模型之結果輸入集成深度學習模型當中，讓模型進行重新訓練與分類，對於模型之精度與強度都會有所提升。如下圖 1-3 所示：

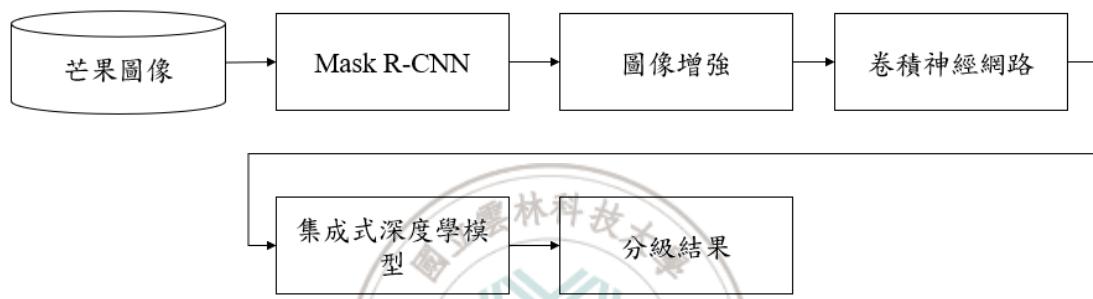


圖 1-3 集成式深度學習模型之架構圖

1.3 研究架構流程圖

由於要先觀察圖像資料集有甚麼問題，因此本研究會先將圖像輸入原始模型，確認原始模型分級的情況，再查看相關文獻，進行後續研究，在確定研究方法後開始進行圖像前處理部分，先使用 Mask R-CNN 去除背景雜訊以利後續建模，再來將圖像進行其他前處理，降低雜訊對於模型之影響，並建立簡易模型對輸入進行分級，最後透過集成學習重新訓練答案，得到最終模型與答案，如圖 1-4 所示。



圖 1-3 研究架構流程圖

第二章 文獻探討

由於水果在採收後作業是一項非常耗費勞力與時間的一項工作，因此需要透過實現自動化技術來提高效率。隨著農業智慧的發展，越來越多學者投入此議題，本章節將列舉出與本研究相關的幾個重要研究，共分成三個小節，包含了圖像前處理、卷積神經網路(Convolutional Neural Networks，CNN)、集成式學習(Ensemble Learning)。

2.1 圖像前處理

2.1.1 傳統目標檢測的方法

傳統目標檢測的方法大概分為三個階段：第一步在圖像上選擇候選區域，然後對候選區域提取特徵，最後使用訓練的分類器進行分類。下列將依序介紹。

(1)選擇候選區域：對目標位置進行定位。由於目標可能出現在圖像的任意位置，且目標的大小、長寬比例非常多變化，因此一開始要使用預測框對整個圖像進行滑動，而且需要設置不同的尺度，不同的長寬比。此舉雖然包含了目標所有可能的位置，但缺點為時間過長，並產生過多的預測框，這也嚴重影響後續特徵提取和分類的速度和性能。(2)特徵提取：由於目標的形態、光照與背景多樣性等因素使得設計一個有效的特徵並不是那麼容易。然而提取特徵的好壞直接影響到分類的準確性，常用的有 SIFT 與 HOG。舉例來說 Dalal 學者[6]提出 HOG 特徵提取工具來捕捉目標的輪廓特徵；Tan et al.[48]使用 HOG 來提取特徵，並使用 SVM 來快速檢測有水果狀的區域；Sengupta et al.[42]使用 SIFT 用來檢測或提取未成熟之柑橘類水果的局部特徵。(3)分類：提取特徵後對目標進行分類，分類器常用的有 SVM、Adaptive Boosting(AdaBoost)等。舉例來說 SVM 為監督式學習方法，針對二元分類問題所開發的分類工具，Nemade et al.[32]以 SVM 分類器來對葡萄與芒果圖像進行分類，並且可以有效提高水果標註結果、Li et al.[27]透過 HOG 提取特徵並分別對花生圖像進行分類，其結果發現 SVM 可以有效將花生圖像分級與 Zhang et al.[59]學者透過 SVM 對柑橘的光譜數據進行分類，其結果表明透過 SVM 可以得到較高的檢測準確度；AdaBoost 為將多個不同決策樹用一種非隨機方式組合，每一輪分類後會將分類錯誤之樣本進行增強，加權後再用於下一個分類器繼續使用，直到達到設定的最大疊代次數才會停止分類，Lin et al.[29]學者提出了一

種使用 AdaBoost 分類器為柑橘圖像進行分類、Zhao et al.[60]學者開發了可以在溫室內自動檢測番茄是否成熟的工具，使用灰階圖像先對圖像提取特徵，再使用 AdaBoost 分類器進行分類。

表 2-1 傳統目標檢測方法文獻整理

作者(年分)[參考文獻]	目標	方法
Dalal(2006)[6]	捕捉目標的輪廓特徵	HOG
Sengupta <i>et al.</i> (2014)[42]	提取未成熟之柑橘類水果的局部特徵	SIFT+SVM
Lin <i>et al.</i> (2016)[29]	為柑橘圖像進行分類	AdaBoost
Zhao <i>et al.</i> (2016)[60]	分類番茄在溫室中是否成熟	AdaBoost
Tan <i>et al.</i> (2018)[48]	解決提取輪廓特徵等問題	HOG+SVM
Li <i>et al.</i> (2018)[27]	對花生圖像進行分類	HOG+SVM
Zhang <i>et al.</i> (2018)[59]	SVM 對柑橘的光譜數據進行分類	SVM
Nemade <i>et al.</i> (2020)[32]	對葡萄與芒果圖像進行分類	SVM

2.1.2 深度學習目標檢測的方法

不同於與傳統的目標檢測方式，卷積神經網路(Convolution Neural Network，CNN)向網路中輸入一張圖片，接著將它傳遞到多個卷積層和池化層中，最後輸出目標所屬之類別。但由於使用 CNN 模型來辨識圖片中的目標只能用非常多不同大小的候選區域去做偵測，這非常耗費時間，因此就有學者 Girshick *et al.* [12]開發新的目標檢測工具 Regions with CNN(R-CNN)預先篩選出約 2000 個可能的區域，再將這 2000 個可能區域個別去作分類，再以支持向量機(Support Vector Machine, SVM)來區分是否為物體或者背景，最後經由一個線性迴歸模型來校正預選區域位置。但 R-CNN 存在某些問題，因為在一開始產生 2000 個可能區域，再由 CNN 來擷取特徵並且存放起來，每個區域都至少得跑 2000 次的 CNN 模型，造成一直重複提取特徵，很耗費設備與時間，因此提出 R-CNN 的作者之一 Girshick [11]在 2015 年又提出了一個改良方法 Fast Region Convolution Neural Network(Fast R-CNN)，不需要做 2000 次 CNN 模型，只需要做一次 CNN 模型，為有效的簡化訓練過程。由於 R-CNN 或 Fast R-CNN 都還是要先透過選擇性搜索算法預選候選區域，速度上還是較慢，因此在 2015 年時 Ren *et al.*[40]提出了 Faster

Region Convolution Neural Network(Faster R-CNN)，增加 Region Proposal Network(RPN)，經由 RPN 可以得到最有可能的預選區域，並且得到該預選區域可能包含目標物的機率。

上述方式為較常使用的兩階段目標檢測之工具，先獲得建議之候選區域，再來是對候選區域進行分類，看是否有目標在候選區域。再者是單一階段較常使用的目標檢測工具由 Liu *et.al*[12]所提出的 Single Shot MultiBox Detector(SSD)，Redmon *et al.* 學者相繼提出 YOLO、YOLO-V2、YOLO-V3[37][38][39]，對於 YOLO 各版本來說，與 R-CNN 相比，優點是單一網路設計，少了讓模型學會去看哪邊有物體，統一為一個迴歸問題，判斷的結果會包含預選區域位置，以及每個預選區域所屬的類別及機率。它容易訓練，而且速度快，但還是會裁切到複雜背景的部分。He *et al.*[15]所提出的 Mask R-CNN，不只是找到候選區域，還可以做到接近完美的遮罩。Mask R-CNN 是建構於 Faster R-CNN 之上，Faster R-CNN 就是沒辦法做到 Pixel 層級的遮罩，因此 Mask R-CNN 改採用雙線性插值法(Bilinear Interpolation)來改善 RoIPooling，新增 RoIAlign，會讓遮罩位置更精準，因此許多人將這項技術運用在背景複雜之圖像檢測上。Zuo *et al.*[62]對於在即時指針的自動讀取辨識(RPM)，此應用在軍事方面，由於通過機器視覺辨識儀表的指針其準確性會受到圖像照明不均、不同圖像的照明變化範圍很廣、複雜的背景、儀錶板傾斜、圖像模糊等問題，因此學者提出了改善 Mask R-CNN 中 RoiAlign，對指針類型進行分類，並且提出角度法計算儀表指針，其發現對於自然環境中此模型具有較強的計算效率與模型的強健性。Gang *et al.*[10]針對在航空業中，電纜支架之檢查方法，其目的是為了解決傳統檢查方法效率低等問題，學者訓練了基於 Mask R-CNN 的分類器，並且提出了半自動電纜支架檢查方法，其實驗結果表明，此方法是有效的，並且為了驗證此目的，最終開發了此框架系統。Qiao *et al.* [35]針對於智慧化之畜牧業，開發了基於 Mask R-CNN 之牛的實例分割與輪廓提取方法，解決牛實際在飼養場中的目標提取等問題，其結果表明，平均精度(MAP)落在 0.92，精度優於當時最先進的其他實例分割算法。

近年來由於目標檢測技術越來越成熟，逐漸有越來越多學者對於水果品質分類這個議題來做研究，舉例來說，Bargoti *et al.*[3]提出在果園中進行水果目標檢測，並且使用 Faster R-CNN 來當作目標檢測的工具，此研究對於在果園進行目標

檢測之貢獻相當大。Wan *et al.*[52]學者提出基於 Faster R-CNN 的目標檢測框架，應用於多類別水果檢測，結果表明與傳統目標檢測工具相比，具有高的準確度與短的處理時間。但因為識別低，速度較慢，不能滿足實時檢測場景。Valdez *et al.*[51]學者將水果缺陷視為目標檢測問題，訓練 YOLO-V3 模型，檢測哪些蘋果是健康的或有缺陷。Tian *et al.*[49]學者提出了 YOLO-V3 結合 DenseNet 結構解決深度學習方法運用農作物在果園環境下檢測蘋果與其學者所測試的模型相比，其 YOLO-V3 結合 DenseNet 具有更好準確性與實用性。但上述方法對於在複雜的背景下之圖片來做分類，還是會因為背景有雜訊，而造成分類精度較低。

依照本研究之資料集來看，因背景複雜多變，因此需要方法來解決背景雜訊過多的問題。Yu *et al.*[57]提出一種基於 Mask R-CNN 的水果檢測方法，讓傳統機器視覺的模型強度、精度、速度增強，平均檢測精度都落在 95.78% 左右，並且適用在重疊、隱藏與光照環境下的水果，也有助於提取完整輪廓。Jia *et al.*[18]為了能使用深度特徵與良好的目標檢測性能，使用 Mask R-CNN 模型來檢測重疊的蘋果，準確率有達到 97.31%，解決了水果被遮擋住導致無法提取的問題。

表 2-2 深度學習目標檢測方法文獻整理

作者(年分)[參考文獻]	目標	方法
Girshick <i>et al.</i> (2014)[12]	解決傳統目標檢測問題	R-CNN
Liu <i>et al.</i> (2014)[12]	解決過去目標檢測速度與精度不佳等問題	SSD
Girshick (2015)[11]	簡化訓練過程，提升目標檢測速度	Fast R-CNN
Ren <i>et al.</i> (2015)[40]	提出 RPN，解決過去精度不佳等問題	Faster R-CNN
Redmon <i>et al.</i> (2016)[39]	以單一網路設計，統一迴歸問題	YOLO
He <i>et al.</i> (2017)[15]	解決背景複雜等問題	Mask R-CNN
Bargoti <i>et al.</i> (2017)[3]	在果園中進行目標檢測	Faster R-CNN
Qiao <i>et al.</i> (2019)[35]	解決牛實際在飼養場中的目標檢測問題	Mask R-CNN
Tian <i>et al.</i> (2019)[49]	在果園中對蘋果進行目標檢測	YOLO-V3

續表 2-2 深度學習目標檢測方法文獻整理

作者(年分)[參考文獻]	目標	方法
Yu <i>et al.</i> (2019)[57]	在重疊、隱藏與光照的環境下 對水果進行輪廓提取	Mask R-CNN
Zuo <i>et al.</i> (2020)[62]	解決即時指針的自動讀取辨識 並在自然環境中得到較強的效率	Mask R-CNN
Gang <i>et al.</i> (2020)[10]	解決電纜支架的檢查方法在 傳統方法效率低等問題	Mask R-CNN
Wan <i>et al.</i> (2020)[52]	以垂統目標檢測鄉比 對多類別水果進行檢測	Faster R-CNN
Valdez <i>et al.</i> (2020)[51]	對水果缺陷進行目標檢測等問題	YOLO-V3
Jia <i>et al.</i> (2020)[18]	對重疊的水果進行目標檢測	Mask R-CNN

2.1.3 圖像增強處理

對於在擷取圖像時，可能有環境的影響，現場光源干擾、震動等，由於在這些情況下，導致圖像品質不佳，或是圖像中有太多雜訊，這時得透過一些方法對影像做合適的處理，達到去除雜訊，並將所需特徵突顯出，使圖像可以用於後續的圖像分析，這個過程稱之為圖像前處理。Singh *et al.*[46]提出新的圖像分割方法，此方法包括兩個部分做結合(1)基於中智學(Neutrosophic-Entropy)的聚類演算法(NEBCA)(2)HSV 色彩空間轉換，結果表示 HSV 色彩空間對分割區域裡面的特徵，可以更好的在視覺上被觀察。Jin *et al.*[19]先通過非下採樣的小波轉換(NSST)將圖像分為低頻與高頻，然後再將圖像轉換 HSV 色彩空間，實驗結果發現，與其他常規方法相比，可以更突顯圖像裡的訊息。Ganesh *et al.*[9]提出基於實例分割 Mask R-CNN 來檢測水果，並且進行分類，其結果顯示，在自然光底下，使用 RGB+HSV 圖像當作輸入，其精度可以看到有明顯提升。Tian *et al.*[49]使用多種圖片增強方式來增加模型強度，其中有使用模糊圖片來增強模型，此研究為了驗證圖片增強方式，每一次都刪除一種增強方式並且獲得其 F1 值與 IOU 值，發現圖像模糊化若被刪除，其模型強度下降最多，可以看出圖片模糊化是有利於模型精

度。Bahrami *et al.*[2]學者發現圖像模糊化會造成邊緣區域的對比度降低，可消除其他雜訊，使缺陷區域更好預測。Jung *et al.*[20]將圖像進行平均色處理可以改善血管的可視化效果，而且可以提高邊界的清晰度，可以看出將圖片進行平均色處理視有助於影像辨識。Gregori *et al.*[13]學者提出彩色圖像的模糊平均檢測方法，實驗結果表明，此方法適合不同類型之圖像雜訊。Shirgahi *et al.*[44]學者提出將籬木花在 RGB，HSB 和 CMY 等不同顏色模式設置平均值，並且訂定適當的模糊規則，最終將圖像辨識為籬木花的準確率為 98.6%，其結果表明將圖像色彩平均化視有助於圖像分類。

表 2-3 圖像增強處理文獻整理

作者(年分)[參考文獻]	目標	方法
Shirgahi <i>et al.</i> (2011)[44]	增加圖像分類效果	平均色處理、BLUR
Bahrami <i>et al.</i> (2017)[37]	模糊圖像使雜訊降低	BLUR
Jin <i>et al.</i> (2018)[19]	轉換到 HSV 空間使圖像得到多特徵	HSV
Gregori <i>et al.</i> (2018)[13]	減少背景複雜之雜訊	平均色處理、BLUR
Ganesh <i>et al.</i> (2019)[9]	提升分類精度	HSV
Tian <i>et al.</i> (2019)[49]	模糊圖像增強模型	BLUR
Jung <i>et al.</i> (2020)[20]	改善血液可視化效果	平均色處理
Singh <i>et al.</i> (2020)[46]	使圖像在視覺化能更好被觀察	HSV

2.2 卷積類神經網路(Convolutional Neural Network，CNN)

自從卷積神經網路(Convolution Neural Network，CNN)在 2014 年 ImageNet 大規模視覺辨識挑戰賽(ILSVRC)上贏得比賽，卷積類神經網路(Convolutional Neural Network，CNN)開始變成深度學習網路領域的較常發展對象[5][14]，它是一種前饋式神經網路，由 3 個 Con2D 層 *Convolution layer* 和 2 個池化層 *Pooling layer* 與全連接層堆疊而成。在大型影像處理方面相較於其他機器學習模型獲得的結果相比，有較好的效果 Spanhol *et al.*[47]。

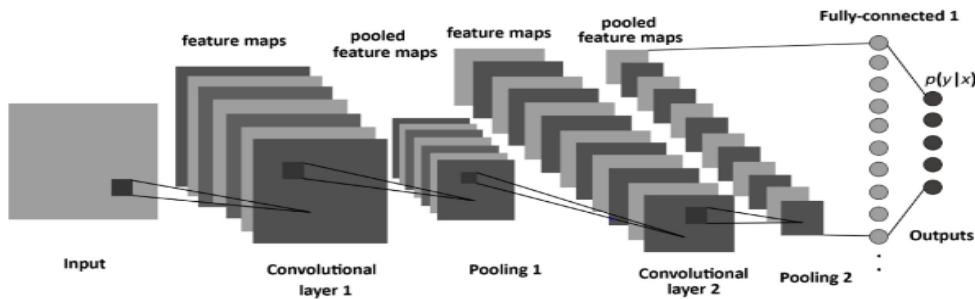


圖 2-1 卷積類神經網路架構圖[65]

由於卷積神經網路目前是深度學習之影像辨識裡面，最為基礎的深度學習模型，目前被廣泛使用。Dias *et al.*[7]使用 CNN 和支持向量機(SVM)方法在復雜的背景下自動對蘋果所開的花提取特徵，該方法產生了相對準確的蘋果開花面積。Hou *et al.*[16]提出一種基於卷積神經網路之算法結合選擇性搜索算法(Selective Search)來辨識在電子磅秤上堆疊的水果，近而獲得較顯著的準確率。Jahanbakhshi *et al.*[17]使用改進的 CNN 模型來檢測檸檬的缺陷並且對圖像進行分級，與其他方法進行比較，結果表明卷積神經網路對於其分級結果來說非常準確。Wu *et al.*[55]使用基於雷射光反向散射成像與卷積神經網路來對於蘋果缺陷進行自動檢測分級，其獲得了 92.5% 的高準確率。Khan *et al.*[22]學者使用深 CNN 模型來對不良品之水果圖像進行訓練，其結果發現此方法優於當時其現有模型。Osako *et al.*[33]學者微調了 VGG16 網路架構來建立模型，其結果表明卷積神經網路適合為水果圖像分類。Momeny *et al.*[31]學者使用 CNN 來對櫻桃圖像進行分類與檢測，其透過最大池化與平均池化的組合來改善原始 CNN，其結果表明 CNN 適合替代傳統檢驗水果方式。綜合上述可以得知，現今卷積神經網路被運用在視覺辨識領域上都有得到很好的績效，因此本研究將卷積神經網路作為本研究主要基本模型。

表 2-4 卷積類神經網路文獻整理

作者(年分)[參考文獻]	目標	方法
Spanhol <i>et al.</i> (2016)[47]	解決機器學習性能不佳等問題	CNN
Hou <i>et al.</i> (2016)[16]	解決水果圖像重疊等問題	CNN

續表 2-4 卷積類神經網路文獻整理

作者(年分)[參考文獻]	目標	方法
Han <i>et al.</i> (2018)[14]	解決訓練集有限等問題並提出 CNN 結合遷移學習與網路增強等方式	CNN
Han <i>et al.</i> (2018)[14]	解決訓練集有限等問題並提出 CNN 結合遷移學習與網路增強等方式	CNN
Dias <i>et al.</i> (2018)[7]	解決在複雜的背景下自動對蘋果所開的花提取特徵	CNN
Khan <i>et al.</i> (2018)[22]	對不良品之水果圖像進行訓練	CNN
Jahanbakhshi <i>et al.</i> (2020)[17]	檢測檸檬的缺陷並且對圖像進行分級	CNN
Wu <i>et al.</i> (2020)[55]	對於蘋果缺陷進行自動檢測分級	CNN
Osako <i>et al.</i> (2020)[33]	尋找適合水果圖像分類之模型	CNN
Momeny <i>et al.</i> (2020)[31]	對櫻桃圖像進行分類與檢測是否有缺陷	CNN

2.3 集成式學習(Ensemble Learning)

在 2015 年 Giba[66]在 Otto 舉辦的比賽當中，使用了 38 個模型，並且透過 Feature Engineering 以及堆疊技巧，建立三層的預測架構，並獲得了第一名。集成學習的熱潮就開始了。所謂的集成式學習(Ensemble Learning)，簡單來說就是「三個臭皮匠，勝過一個諸葛亮」的概念，集結大眾概念，可以獲得更好的解釋力。以神經網路來說，就是綜合多種模型的表現，提升最後的分類或預測結果。Tuncer *et al.*[50]學者針對肝細胞癌之議題做研究，學者提出相對應模型，並測試了 23 種常見的分類器與集成分類器。Cha *et al.*[5]學者發現針對乳突疾病之圖像使用集成分類器，其使原始模型產生顯著改善。一般常見的架構有三種(1)Bagging(2)Boosting(3)Stacking，這三種架構在結合模型時使用的概念並不一樣，以下為介紹。

(1)Bagging

將樣本重複抽取並且抽取後放回，並且產生多個子資料集後，依序建議多個模型，最後再將模型結果彙整。若是預測問題，就是將模型結果平均；若是分類問題，就是使用投票方式，判斷哪個類別出現最多次。Yang *et al.*[54]將基於隨機

森林的集成學習分類器使用在渦輪機葉片損傷的檢測，提高葉片缺陷檢測的準確性。Zhou *et al.*[61]對於預防疾病傳播與提高治療效果提出了基於集成深度學習模型，用來檢測最新型的 COVID-19，並且使用遷移學習用於初始化模型參數，預訓練了三個深度卷積神經網路模型，最後使用了集成學習之分類器與其他三個分類器進行比較，結果表明集成學習模型優於其他三種分類器。尤於在近幾年來眾多學者開始使用集成學習，其精度提升之優勢可以很明顯看到，因此本研究將使用此方法，來提升模型精度。

(2)Boosting

此方法事先進行簡單模型之預測，結果會有預測錯誤之資料，此演算法概念主要是將錯誤的資料權重放大，再進行第二次建模，再將得到之錯誤權重再放大，以此類推進行下去。但尤於此模型只能以相較於其他分類器模型較弱之分類器，若是分類器強度都很高，其結果會互相干擾，影響最後判斷。Rad *et al.*[36]對於人體胚胎進行研究，提出了一種基於 Boosting 的集成學習來分割人體胚胎，其結果表方法證實此方法辨識的平均精度落在 85.2% 左右，優於當時最新模型 4.9% 的精度。Lee *et al.*[25]對於模糊編碼之圖像問題，提出了基於圖像集成重複補償的模糊編碼方法，其結果表明此方法比現有方法有更高的分類精度，並且也證明了該方法在圖像邊緣模糊之問題的解決優勢。Yu *et al.*[58]針對蒸汽管道高溫區域進行快速診斷，提出基於集成學習理論與其他概念做結合，實驗結果表明，此演算法可以有效的診斷蒸氣管道高溫區域。

(3)Stacking

將訓練好之模型對訓練集進行預測，模型所得到之預測值將成為新模型之輸入值，得到最終預測值。Książek *et al.*[23]針對於肝癌細胞之檢測，提出了基於 Stacking 之集成學習，使用七個分類器，並結合成一種新的分類器以獲得更高之檢測精度，其結果表明演算法精度落在 90% 左右，可在癌症篩選實驗室所使用，提供臨床醫生進行準確判斷。Kardani *et al.*[21]針對於邊坡土壤之議題來做研究，此議題對於地質工程研究有很大之貢獻，改善了邊坡穩定性預測，學者先提出了基於 Stacking 之集成學習方式，與 11 種分類器單獨來做比較，其結果發現混和之堆疊分類器相較於其他單獨分類器來說高了 7% 左右。Boukellouz *et al.*[4]學者為了

減少醫療的核磁共振估計的誤差，提出了三種分類器的堆疊，分類器有 ANN、Random Forest 與 KNN，其結果表明具有堆疊的集成學習模型優於其他模型。

尤於上述相關文獻可以看出，基於集成式學習之方法可以有效的提升模型精度，因此本研究將會使用集成式學習作為後續建模之方法。

表 2-5 集成式學習文獻整理

作者(年分)[參考文獻]	目標	方法
Yu <i>et al.</i> (2012)[58]	針對蒸汽管道高溫區域進行快速診斷	Random Forests
Rad <i>et al.</i> (2018)[36]	提出了一種基於 Boosting 的集成學習來分割人體胚胎	Ensemble Learning
Cha <i>et al.</i> (2019)[5]	使用集成學習改善原始模型	Ensemble Learning
Lee <i>et al.</i> (2019)[25]	提出了基於圖像集成重複補償的模糊編碼方法	Ensemble Learning
Boukellouz <i>et al.</i> (2019)[4]	為了減少醫療的核磁共振估計的誤差	ANN、Random Forest、KNN
Giba (2020)[51]	比較多種集成學習模型	Ensemble Learning
Tuncer <i>et al.</i> (2020)[50]	比較多種集成學習模型	SVM、KNN、Boosted Trees、Ensemble Learning
Yang <i>et al.</i> (2020)[54]	提高葉片缺陷檢測的準確性	Ensemble Learning
Zhou <i>et al.</i> (2020)[61]	預防疾病傳播與提高治療效果	Ensemble Learning
Książek <i>et al.</i> (2020)[23]	針對於肝癌細胞之檢測提高績效	KNN、SVM
Kardani <i>et al.</i> (2020)[21]	解決邊坡土壤之穩定性預測問題	Ensemble Learning

第三章 研究方法

本論文提出集成式學習(Ensemble Learning)結合深度學習模型，應用於芒果品質檢測並且分級之探討，由於要先觀察圖像資料集有甚麼問題，因此本研究第一步會先將圖像資料做資料清洗，並且進行資料觀察時發現，將未處理之圖像輸入後，其模型分級效果不理想，發現背景複雜與圖像中有多重目標，使模型無法準確判斷要分類的目標物，且需要耗費大量時間在模型的訓練上，因此第二步將使用 Mask R-CNN 去除背景雜訊以利後續建模，第三步使用去除背景之圖像進行模糊化處理，並輸入簡易模型進行分級，降低雜訊對於模型之影響，增加模型強度，第四步使用去除背景之圖像進行平均色處理，並輸入簡易模型進行分級，突顯果實成熟度的平均顏色讓模型學習，第五步將每個模型之分級結果透過集成學習重新訓練，增加模型精度，得到最終分級結果與績效。如圖 3-1 所示。

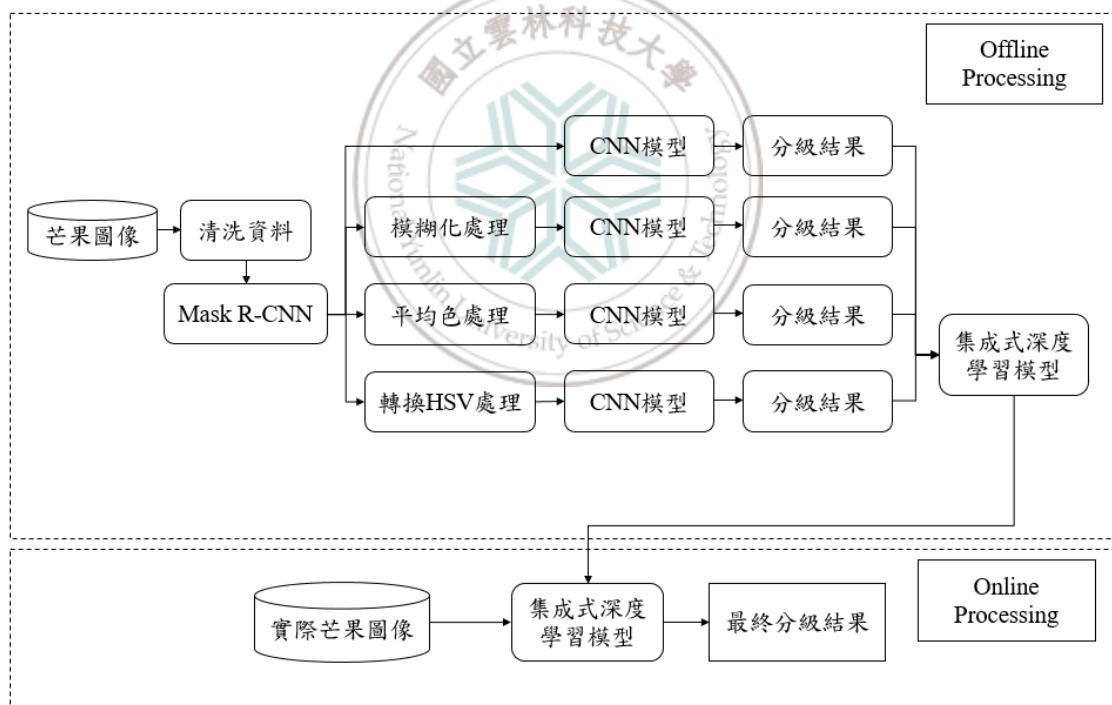


圖 3-1 研究流程圖

3.1 圖像資料集介紹與資料清洗

本研究所使用之芒果圖像資料集為 NTHU EE BIIC Lab 人類行為訊息暨互動計算研究室、台灣瓦克國際股份有限公司與屏東枋山鄉的建賢合作社共同合作所

蒐集之圖像，其同意提供本研究使用圖像之單位為教育部人工智慧競賽與標註資料蒐集計畫[68]。

分別有 A 類數據集 2035 張，B 類數據集 2361 張，C 類數據集 2004 張如圖 3-2、3-3、3-4。其資料集欄位名稱為圖像編號與等級，每筆資料共 2 個欄位，國外出口用標記 A；國內銷售用標記 B；食品產業加工用標記 C，如表 3-1 所示。

表 3-1 芒果圖像資料欄位

圖像編號	等級
32783.jpg	A
24935.jpg	B
30841.jpg	C



圖 3-2 A 類芒果圖像



圖 3-3 B 類芒果圖像

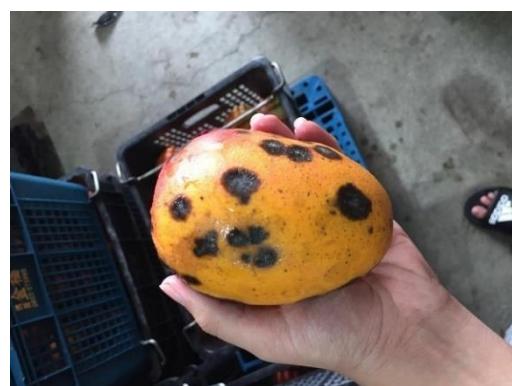


圖 3-4 C 類芒果圖像

本研究所使用之資料集，由於初始資料集並沒有將圖像分類好到各自類別，因輸入模型之需求來說，要將圖像分為 A 類、B 類以及 C 類，因此本研究透過業者所提供之圖像與分類資料，先建立各自圖像輸出之資料夾，並將每張圖片之答

案輸入模型，讓模型先將圖像資料分別放入各自類別，並且將原始圖像資料分為訓練集 4479 張與測試集 800 張，驗證集 1121 張。

3.2 Mask R-CNN 應用於芒果辨識的作法

3.2.1 檢測實物區域

Mask R-CNN 檢測實物區域架構圖如圖 3-5 所示，分為四個步驟，以芒果圖像舉例(1)先對一張芒果圖像分割成多個區域(Anchors)，此區域錨點(Anchors)為不同大小的方框，其方框會重疊。(2)對芒果圖像中實際要分割的芒果標註其位置座標。(3)對分割的區域錨點(Anchors)與實際芒果的位置座標進行面積重合度(Intersection over Union，IOU)計算，此步驟會計算區域錨點(Anchors)中哪些是芒果或背景雜訊，重疊度高為芒果，重疊度低為背景雜訊。(4)將區域錨點(Anchors)中屬於芒果的座標與實際芒果圖像的座標進行計算，得到相對應的長寬縮放比例。最後檢測區域(Region Proposal Network，RPN)的輸出結果會是前景機率較高的一定數量錨點作為可靠區域(Region Of Interest，ROI)，並輸入分類器網路中進行分類計算。

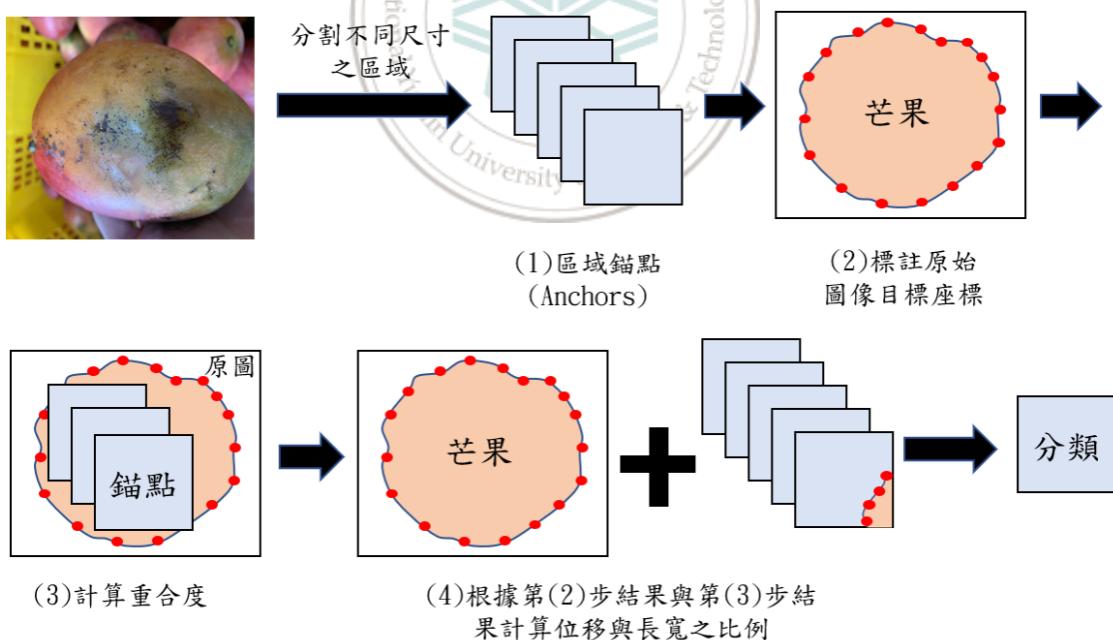


圖 3-5 Mask R-CNN 檢測實物區域架構圖

3.2.2 Mask R-CNN 模型之完整步驟

Mask R-CNN 模型架構圖如圖 3-6 所示，分為五個步驟，(1)分析主特徵，又被稱作骨幹網路。用來從輸入之芒果圖像中分析出不等長之尺度的重要特徵，這

些獲得的特徵資料被稱作 Feature map。(2)特徵融合，用特徵金字塔網路(Feature Pyramid Network，FPN)整合主特徵中不等長之尺度的特徵。在後續的 RPN 與分類器會使用到。(3)分析可靠區域，主要透過RPN來實現。該網路主要是計算出前景與背景的預測值，並算出以錨點為基礎的偏移，然後對前景機率較大的可靠區域用演算法去除重複，並從最後結果中取出指定個數的 ROI 用於後續網路的計算。(4)ROI 池化，使用 ROIAlign 的方式來進行。並且拿取特徵融合後的圖像，按照 ROI 中的候選區預位置從圖像中取出對應之目標。ROIAlign 層當中， $i \times (r, j)$ 目標座標位置， d 表示兩點之間的距離， $\Delta h \Delta w$ 表示 i 與 $i \times (r, j)$ 之間的差值，可以解決特徵與圖像融合時錯位之問題，其表示以下公式(1)表示，以 ROIAlign 得到之結果送入分類器進行分類與座標計算，並且經全連接層完成目標分割。

$$\frac{\partial L}{\partial x_i} = \sum_r \sum_j [d(i, i^*(r, j)) < 1] (1 - \Delta h)(1 - \Delta w) \frac{\partial L}{\partial y_{rj}} \quad (1)$$

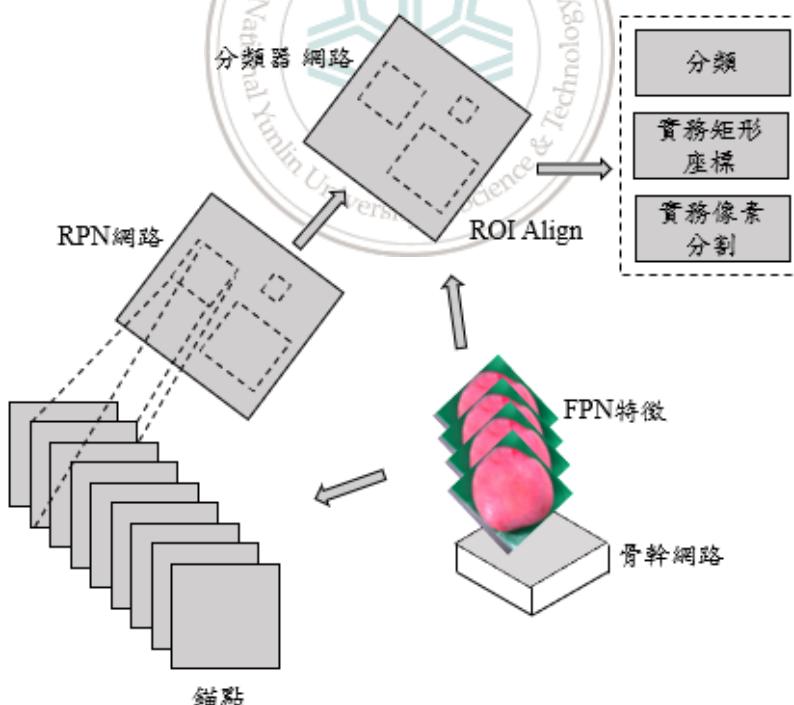


圖 3-6 Mask R-CNN 架構圖

本論文所使用的 Mask R-CNN 其方法為，輸入原始無去除背景之圖像，並將圖像切割成符合目標物大小之矩形，切割點為目標物上、下、左與右離各邊界最

近之點位置，並且利用其遮罩保留其目標物並將背景雜訊去除，輸出為背景去除之圖像，示意圖如下圖 3-7 所示。

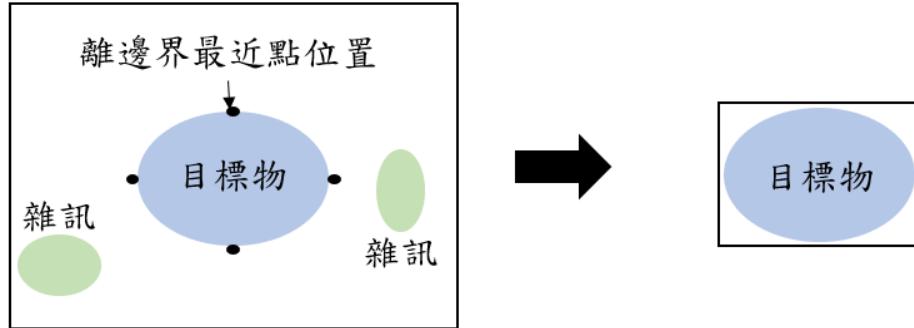


圖 3-7 本研究 Mask R-CNN 去除背景雜訊示意圖

3.3 圖像增強處理應用於芒果辨識的作法

在完成 Mask R-CNN 將圖像背景雜訊去除後，進行多種圖像前處理。舉例來說，(1)對芒果圖像進行模糊處理(2)對芒果圖像進行圖像模糊平均色處理(3)對芒果圖像進行轉換 HSV 顏色空間處理。

3.3.1 模糊化處理

經由原始模型的結果來看芒果的 A、B 與 C 類型的辨識之準確率發現準確率不高，可能因為圖像雜訊太多，由於模糊處理在圖像去除雜訊方面較為廣泛應用，因此本研究將圖像進行模糊處理，減少圖像雜訊。

較常使用之模糊處理大概分為兩種，一種是平均模糊(Averaging Blurring)，另一種為高斯模糊(Gaussian Blurring)，Gaussian Filter 的模糊化效果比起 Averaging 會比較明顯，但是效果卻更為自然。因此本研究使用高斯模糊來將圖像模糊化。

兩者差別在於高斯模糊需要多一個參數 σ (顏色空間的標準差)，此數值越大意味著越遠的像素會有較大的權值，使得模糊效果更明顯，本研究這邊設為 0 讓程式自動計算。Gaussian Filter 公式以式(2)表示：

$$G(x, y) = \frac{1}{2\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2)$$

Gaussian Filter 的中心點 (x, y) 須為 $(0, 0)$ ，下例為一個 3×3 的 (x, y) 值矩陣如圖 3-8，將此矩陣的 x, y 值套入 Gaussian Function 並正規化後就可以得到 3×3 的 Gaussian Filter。

(-1, 1)	(0, 1)	(1, 1)
(-1, 0)	(0, 0)	(1, 0)
(-1, -1)	(0, -1)	(1, -1)

圖 3-8 3×3 的權重矩陣 [21]

3.5.2 平均色處理

從圖像模糊化之模型結果來看芒果的 A、B、C 類型辨識之準確率發現 B 類辨識效果不好，並且經由資料觀察發現 B 類之圖像幾乎都是屬於變色芒果，模型容易被顏色混淆。本研究對芒果圖像進行 Mask R-CNN 目標檢測，並得到屬於該目標(橘色)的遮罩(淺綠色)如圖 3-9 所示，並計算遮罩內 R 、 G 、 B 各像素值加以平均，再以計算出的數值填補背景，讓圖像背景從空白轉為計算的平均色，使模型可以學習更多 B 類芒果的綜合特徵，而不是只學習到單一特徵，此方法可以突顯 B 類果實成熟度的平均顏色讓模型學習。

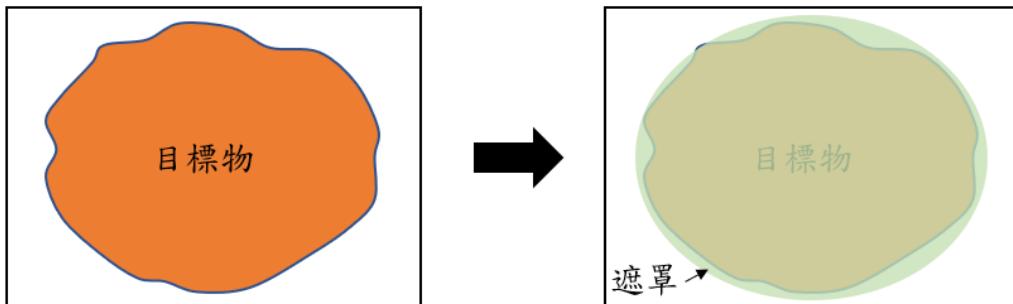


圖 3-9 Mask R-CNN 遮罩示意圖

3.3.3 轉換 HSV 處理

HSV(Hue, Saturation, Value)定義了另一種顏色空間，可視為一個圓錐體。它與RGB模型相似。HSV色彩空間有三個組成部分：色調 H 、飽和度 S 和亮度 V 。經由觀察資料發現，很多圖像對於當下拍攝情況可能較差，因此每張圖片明暗程度都不一或圖像太過運動模糊，因此本研究將透過HSV轉換不同顏色空間，來突顯芒果圖像的表面缺陷，其轉換公式如(3)(4)(5)， (R, G, B) 分別為顏色紅、綠、藍

的座標， Max 為 (R, G, B) 中最大值， Min 則為最小值，色調 H 的值通常介在於 0 到 360，飽和度 S 介在 0 到 1 之間，亮度 V 介在 0 到 1 之間。

$$H = \begin{cases} 0, & (S=0) \\ 60 \times \frac{G-B}{S \times V}, & (\max(R, G, B) = R \& G \geq B) \\ 60 \times \frac{2+(B-R)}{S \times V}, & (\max(R, G, B) = G) \\ 60 \times \frac{4+(R-B)}{S \times V}, & (\max(R, G, B) = B) \\ 60 \times \frac{6+(G-B)}{S \times V}, & (\max(R, G, B) = R \& G < B) \end{cases} \quad (3)$$

$$S = \frac{\max(R, G, B) - \min(R, G, B)}{\max(R, G, B)} \quad (4)$$

$$V = \max(R, G, B) \quad (5)$$

3.4 卷積神經網路應用於芒果辨識的作法

本研究將清洗過後之訓練集當作卷積類神經(Convolution Neural Network，CNN)之輸入，包含了輸入層、卷積層、池化層、以及全連接層。本研究所使用之 CNN 網路架構為 Xception，如下圖 3-10 所示，再將圖像輸入模型之前有進行資料增強處理，對圖像旋轉角度、隨機水平移動、隨機垂直移動、改變圖像形狀、隨機縮放、隨機水平翻轉、隨機垂直翻轉，增加資料集數量以增強圖像強度，Input_shape 為 $(150, 150, 3)$ ，其第一階段 Depthwise 卷積的訓練參數，通道數 d 為 3，對每個通道的特徵圖進行 3×3 的卷積核，是由 36 個卷積層、最大池化層與兩個全連接層所組成的網路，以下會介紹說明。

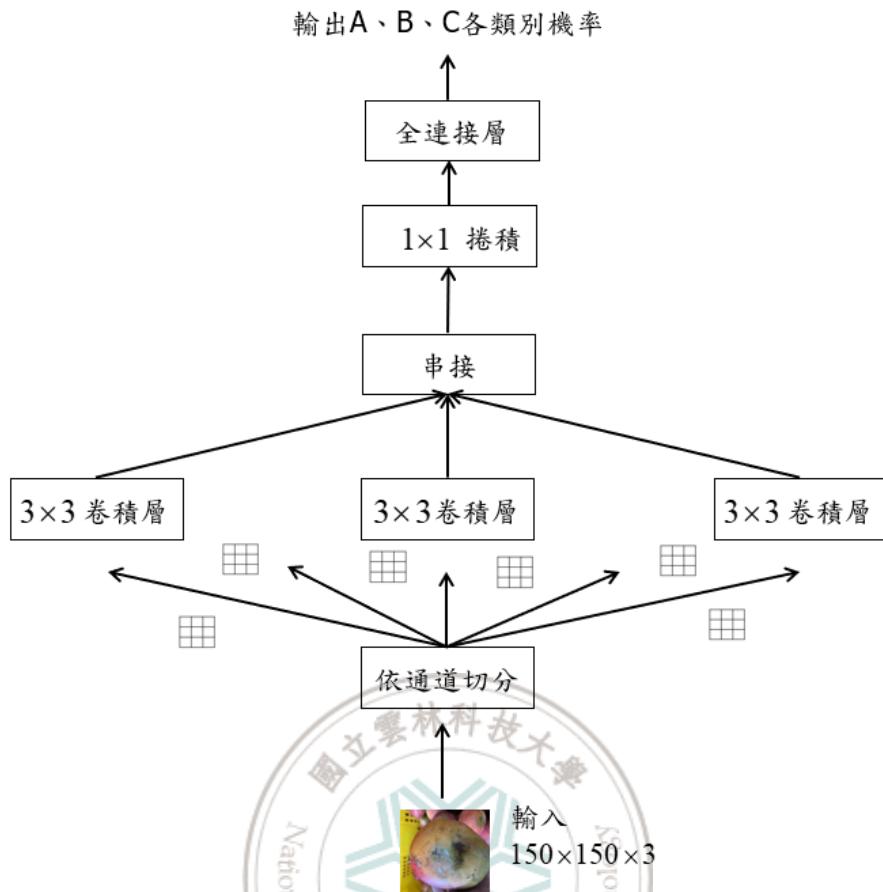


圖 3-10 本研究 Xception 架構圖

3.4.1 深度可分離卷積(Depthwise Separable Convolution)

一開始 Xception 為了將空間與通道的特徵分開學習，因此會對特徵圖進行深度可分離卷積，可分為 2 階段卷積:(1)Depthwise 卷積(2)Pointwise 卷積，以下為介紹：

(1) Depthwise 卷積

假設輸入芒果圖像長度和高度 m 、 n ，並且本研究所輸入之圖像為彩色圖像，其通道數 d 為 3(R、G、B)，其卷積會將圖像依通道切成 d 個 $m \times n \times 1$ 的單一通道，並對這些通道的特徵圖分別進行 3×3 卷積獨立學習個別的空間特徵後，再重新串接起來，如圖 3-11 所示：

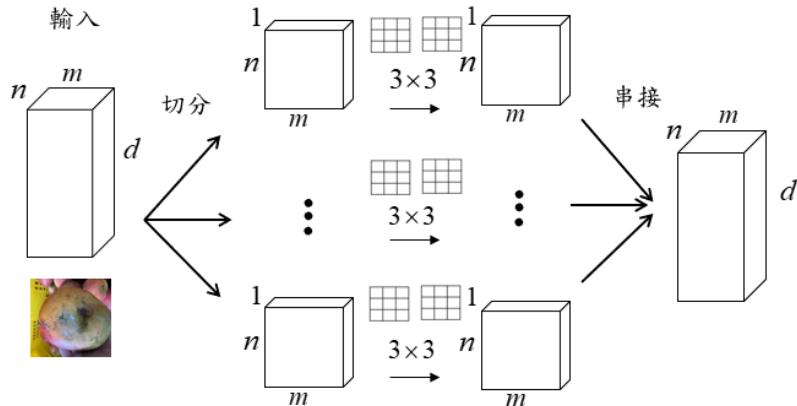


圖 3-11 Depthwise 卷積架構

(2) Pointwise 卷積

階段 2 的卷積會將第 1 階段的芒果特徵圖進行 1×1 卷積，獨立學習通道特徵，有多少 就會有多少 Feature map。如下圖 3-12 所示：

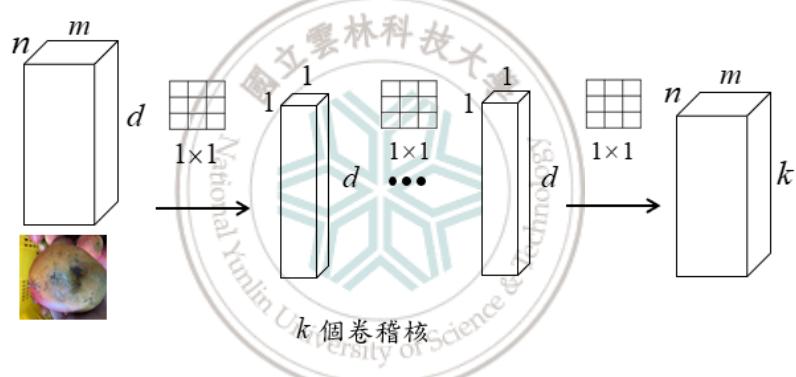


圖 3-12 Pointwise 卷積架構

池化層的工作為對卷積層輸出芒果的特徵圖進行重點挑選如圖 3-13 所示，舉例來說，當一張芒果圖像特徵(像素)太多時，可以把一些相對不重要的特徵丟棄、或者將多個特徵合併成一個，它的目的是為了要降低採樣頻率(Downsampling)，縮小特徵圖尺寸(降維)，並同時確保芒果圖像不會遺失過多的重要特徵。本研究有使用的是較常使用的最大池化 MaxPooling2D。

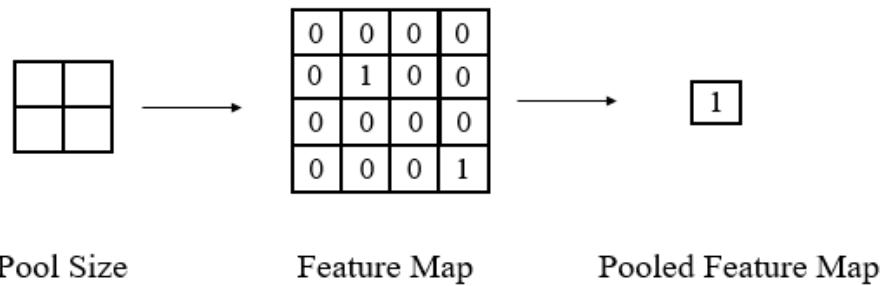


圖 3-13 池化層 MaxPooling2D

CNN 模型的展平層，其中層數是依據前一層 Feature Map 展平化後的數量而定的。如下圖 3-14 所示，舉例來說，展平層在做的事就是將特徵圖展平成 1D 向量，以便送入密集層。其實就是將芒果圖像萃取出來的重要特徵，送進後半部的全連接層進行分類工作；而因為前半部優良的特徵萃取能力，所以這時送進來的芒果圖像特徵已是最重要的特徵，因此全連接層就不需要過多的神經元就能好好的學習資料。最終會整合前面所有神經元透過激活函數 Relu 與 Softmax 與進行計算，此兩層的激活函數會整合前面神經元的輸出，並輸出此芒果圖像的分類結果，其公式為式(6)， O_j 為第 j 個神經元的輸出， $act(\bullet)$ 為激活函數， x_j 為第 j 個神經元的輸入， w_{ij} 表示前面一層所連結的權重值， b_j 為第 j 個神經元的偏差值。

$$O_j = act(x_j \times w_{ij}) + b_j \quad (6)$$

其 Relu 激活函數，公式為式(7)， O 為輸出， x 為輸入， i 為第 i 個神經元

$$O_i = \max(0, x_i) \quad (7)$$

其 Softmax 激活函數，公式為式(8)， O 為輸出， y 為模型預測值， a 為類別的總數， i 為第 i 個神經元，每個數值輸出介於 0 到 1 之間，每一個輸出總合為 1，其結果會得到 A 類、B 類與 C 類的機率，並透過機率判斷此張芒果圖像是屬於哪一個等級，並且計算最終準確率。

$$O_i = \frac{e^{y_i}}{\sum_{j=0}^a e^{y_j}} \quad (8)$$

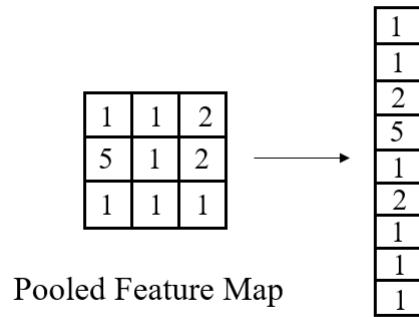


圖 3-14 展平層

3.5 集成式學習(Ensemble Learning)應用於芒果辨識的作法

藉由上述前處理模型所產生的輸出可以得到每一張芒果圖片屬於哪一個類別之分級結果，本節將會介紹如何透過集成式學習(Ensemble Learning)之概念，並挑選出最好之分類模型。本研究將會使用之集成深度學習模型之概念為 Stacking，以下為介紹。

3.5.1 Stacking

以簡單概念來說，以預訓練好四個機器學習模型，當芒果圖像需要做分類時，會各自得到四種前處理模型之分類結果，當作最終模型的輸入，並且得到最終分類結果，其 Stacking 演算法可以分為兩個階段(1)先訓練多個初始模型，其預測結果為 Meta-Data，作為最終模型(Meta-Learning Model)之輸入。(2)最終模型會取得 Meta-Data，整合出最終結果(Forecast Results)。概念圖如圖 3-15 所示。

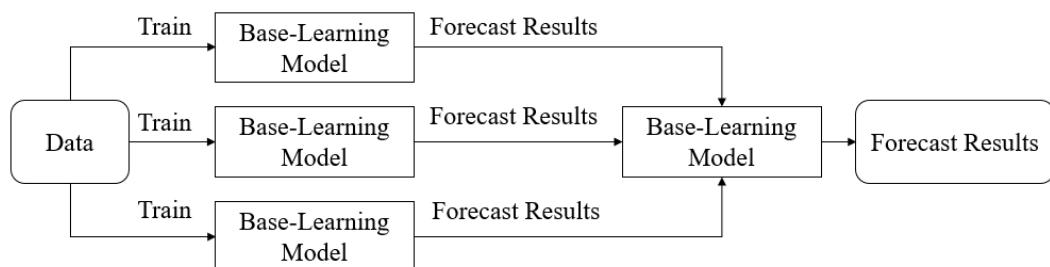


圖 3-15 基於 Stacking 之集成式學習概念圖[26]

本研究會將所有模型所得到之結果應用集成式學習(Ensemble Learning)重新訓練其所有模型所得到最終結果，並比較多種分類器所得到之績效，選擇在本研究

中辨識效果最好之分類器來當作最終模型，來得到每張芒果圖像之等級分類，其本研究之架構圖如圖 3-16 所示。

本研究將類神經網路模型預測出來的驗證與測試，拿驗證的資料集做機器學習訓練，訓練完再用測試資料集做最終的集成學習。

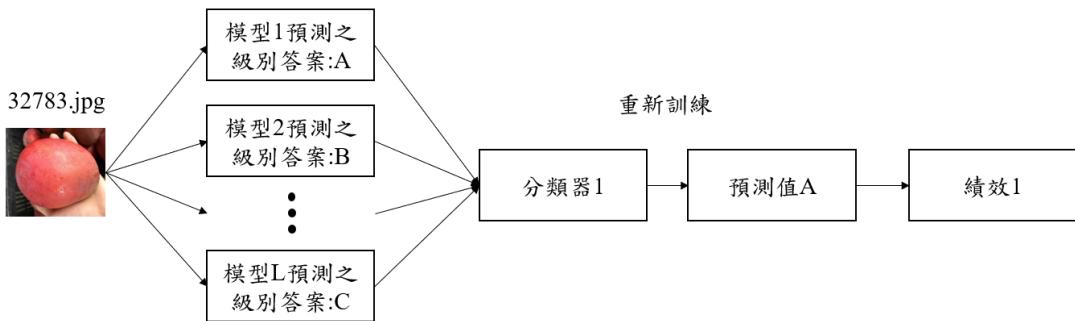


圖 3-16 本研究集成式深度學習架構

(1)Random Forest

[43][33]由 Bagging 演算法所衍生出的方法，先定義幾個隨機樣本，從各模型預測答案中隨機挑選幾個模型的答案，取完後放回，並從選擇的資料中訓練決策樹，並隨機挑選特徵，再使用特徵分割節點，最終是以多數決方式來決定此張芒果圖像分類結果為何，由於此方法以使用投票方式，判斷哪個類別出現最多次，與本研究所需之輸出符合。

(2)Support Vector Machine

[24][41]為一種線性或非線性分類器，對於小樣本、非線性、高維度都有非常好的效果，由於本研究透過集成學習其輸入無法在 2 維空間來做分割，因此透過映射函數，轉換為三維立體空間如圖 3-17，透過超平面(灰階平面)，就可以有效將樣本分類。其主要公式如式(9)所示

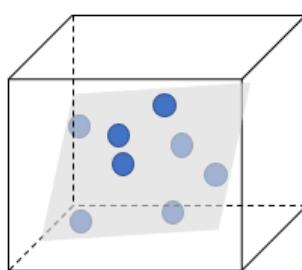


圖 3-17 SVM 三維空間示意圖

$$k(x_1, x_2) = 1 + x_1 x_2 + x^2 \quad (9)$$

(3) Decision Tree

可執行分類與回歸任務，主要公式如式(10)所示，舉例來說，從根節點開始，此節點會詢問模型預測之芒果圖像類別去決定往左分或往右分，以此類推再往下分節點，直到葉節點即結束，並比較分數，分數最高為答案，其公式主要是計算第*i*個節點裡的分數 G_i ， P_{i,k^2} 為在第*i*個節點裡，類別*k*的實例佔訓練實例的比例。其架構如圖 3-18 所示。

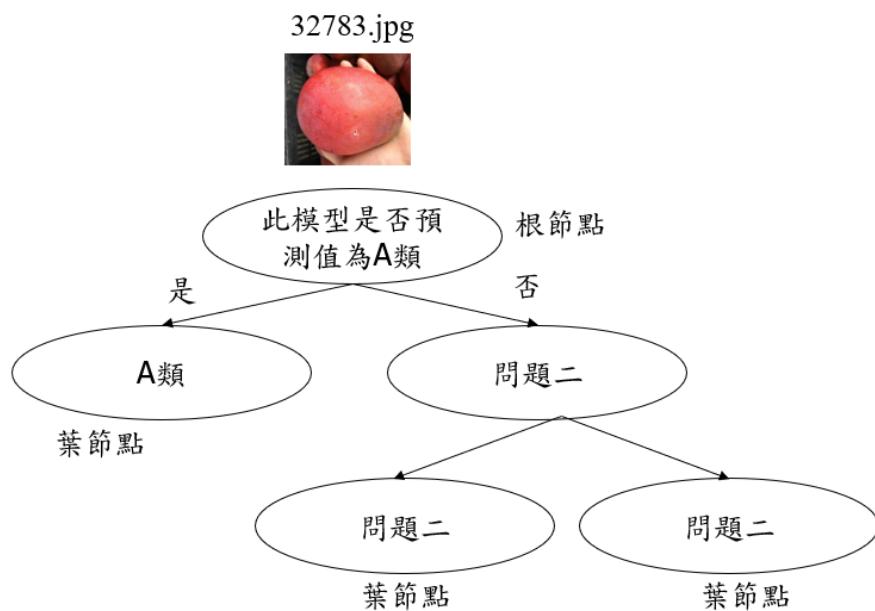


圖 3-18 決策樹架構圖

$$G_i = 1 - \sum_{k=1}^n P_{i,k^2} \quad (10)$$

第四章 實驗模擬

本研究所使用之芒果圖像資料集為 NTHU EE BIIC Lab 人類行為訊息暨互動計算研究室、台灣瓦克國際股份有限公司與屏東枋山鄉的建賢合作社共同合作所蒐集之圖像，且由教育部人工智慧競賽與標註資料蒐集計畫辦公室所同意提供。並使用本論文提出之方法來驗證其有效性，以下將分為六個部分，(1)參數設定與資料集介紹，(2)本研究之模型有效性，(3)集成學習之模型選擇，(4)實驗的績效評估(5)實驗質化合理性說明。所有實驗則是以 Python 完成，並操作在 Intel(R)Xeon(R)Gold 6230R CPU@2.10GHz 的 CPU、Nvidia TITAN RTX 24GB 的 GPU、搭配約 792GB 記憶體以及 Windows10 的作業軟體來完成。

4.1 參數設定與資料集介紹

本研究芒果圖像以手直接拿取芒果拍攝，拍攝由台灣瓦克國際股份有限公司與屏東枋山鄉的建賢合作社共同合作，將圖像蒐集後統一交給 NTHU EE BIIC Lab 人類行為訊息暨互動計算研究室來做處理，其同意提供本研究使用圖像之單位為教育部人工智慧競賽與標註資料蒐集計畫[68]，本研究所使用之資料集總數為 6400 張圖像，分別有 A 類數據集 2035 張，B 類數據集 2361 張，C 類數據集 2004 張，並且將資料拆成比例為 0.7 的訓練集與 0.3 的測試集與驗證集，訓練集總數為 4479 張、測試集總數為 800 張與驗證集總數為 1121 張。其著重在前處理，有對圖像做 Mask R-CNN 之處理，本實驗將圖像長、寬設置為 224×224 ，其訓練步伐為 100，並且對芒果圖像做去背、背景平均色處理、HSV 轉換色彩空間處理與圖像模糊化處理其參數設置為默認參數[64]。為了增加其辨識強度，因此在 CNN 模型中有對圖像做資料增強，本研究 CNN 模型之輸入圖像長、寬設置為 224×224 ，卷積層共有 36 個，中間有穿插 MaxPooling 與 Relu，最後將神經元 Flatten，設 Dropout 為 0.25，全連接層的 Relu 神經元個數為 1024 個，輸出為 Softmax，為此圖像在每一類別之機率，選擇最高當最終輸出。集成學習輸入為四種前處理之答案，其答案以 0、1 與 2 表示，分別代表 A 類、B 類與 C 類，其最終輸出為其預測之類別。

4.2 前處理之模型有效性

本研究會以目標檢測模型為例與圖像增強處理模型為例來證實本研究之模型有效性。

4.2.1 以目標檢測模型為例

一開始先將無處理之圖像輸入 CNN 模型並進行分級，發現其績效不如預期，分類錯誤之圖像發現，較多圖像背景複雜，且會有其他芒果出現，如圖 4-1 所示。此觀察方法是透過 Grad-CAM 方法，以圖 4-1(b)舉例，如圖 4-2 所示。

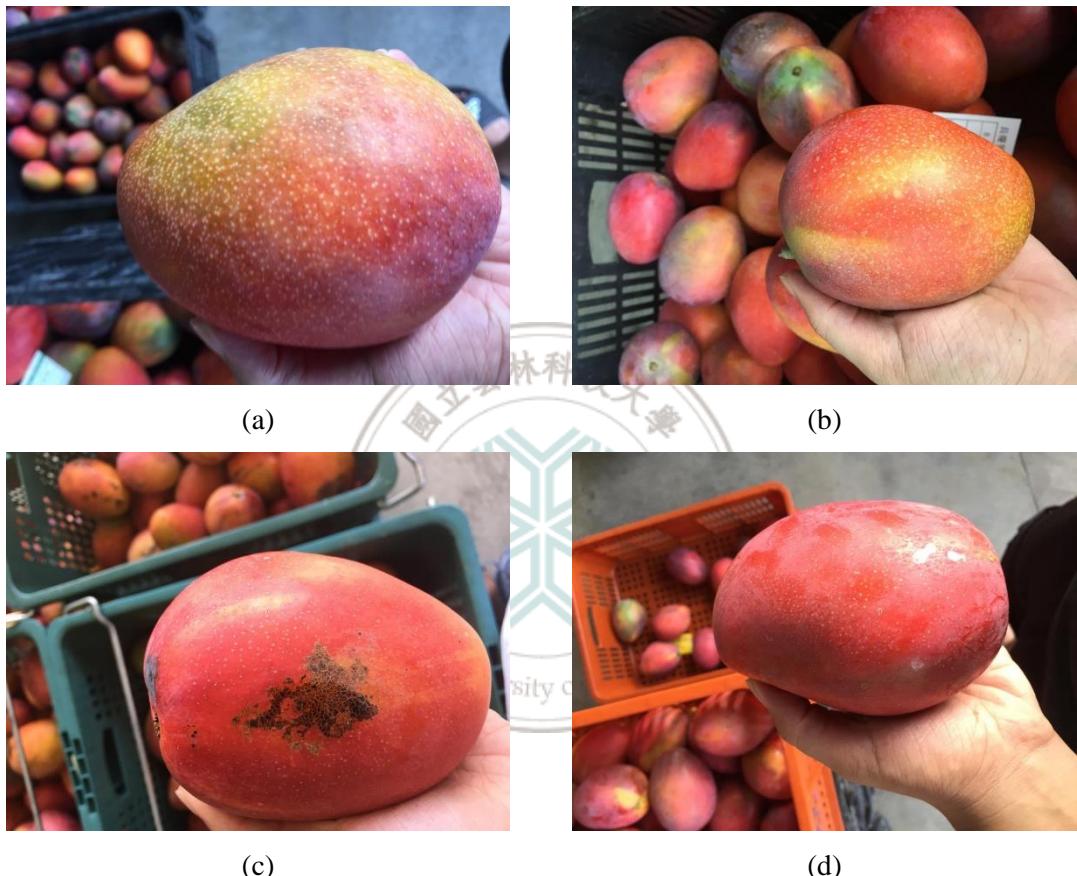


圖 4-1(a)(b)(c)(d)為背景複雜之芒果圖像

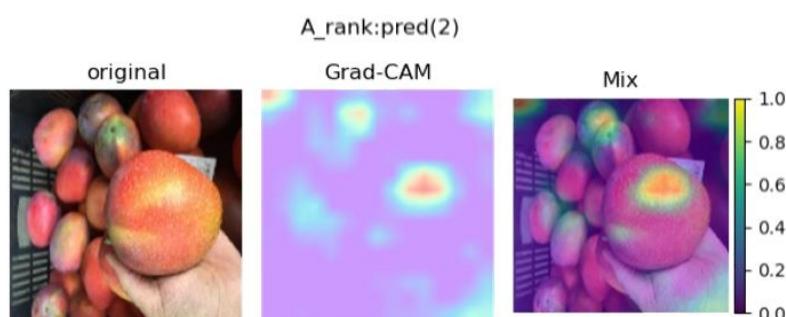


圖 4-2 芒果圖像 Grad-CAM 示意圖

因此本研究使用 Mask R-CNN 之目標檢測方式，將圖像背景雜訊消除，如圖 4-5 所示，並輸入本研究之 CNN 模型做訓練得到之績效，為了驗證本研究所使用 Mask R-CNN 目標檢測模型有效，本研究將比較三種模型之辨識精準度，將本研究之芒果數據集分別進行以下三種方法處理，為(1)未處理之原始圖像(2)圖像經由 Mask R-CNN 目標檢測模型(3)圖像經由目前目標檢測中性能較優的 YOLO_V3 模型，並將此三種方法得到之圖像分別輸入本研究 CNN 模型，得到最終績效，其視覺化如圖 4-3、4-4 所示。績效結果如表 4-1 所示，其結果發現經由 Mask R-CNN 進行去背景處理，相較於其他兩種方法來的好，可證實本研究使用 Mask R-CNN 之模型之有效性。



圖 4-3 原圖經由 Mask R-CNN 處理



圖 4-4 原圖經由 YOLO_V3 處理

表 4-1 目標檢測方法之績效比較

Model	Accuracy	Precision	Recall	F1
原始 CNN	0.74325	0.772	0.751	0.754
Mask R-CNN+CNN	0.80924	0.812	0.800	0.806
YOLO_V3+CNN	0.78625	0.801	0.794	0.801

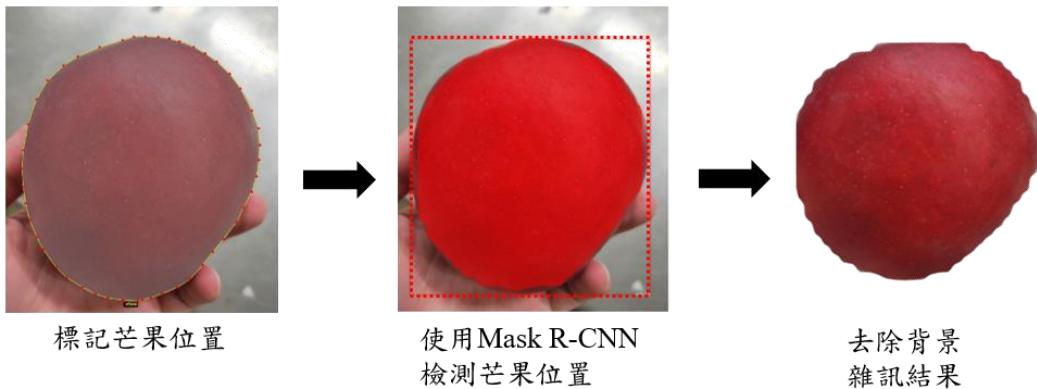


圖 4-5 Mask R-CNN 之處理流程示意圖

4.2.2 以圖像增強處理模型為例

為了使模型更好學習特徵來增加辨識精度，本研究使用了三種前處理方式(1)將芒果圖像進行 HSV 色彩空間轉換如圖 4-6 所示，突顯芒果圖像之不明顯特徵，(2)使用圖像模糊化如圖 4-7 所示，其目的是為了降低使用 Mask R-CNN 處理產出之圖片邊緣度對比度，使雜訊特徵不明顯(3)將芒果圖像進行背景平均色處理如圖 4-8 所示，由於經由 Mask R-CNN 處理並將資料輸入 CNN 模型得到其績效(績效)，發現 B 類辨識效果不佳，因此進行資料觀察發現 B 類芒果圖像多數屬於變色芒果，因此使用此方法降低變色芒果之雜訊干擾，並將此三種方法得到之圖像分別輸入本研究 CNN 模型，得到最終績效。結果如表 4-2 所示，其 Original 為 Mask R-CNN 加 CNN 模型，為了驗證圖像增強處理模型之有效性，因此以三種前處理之績效與沒有圖像增強之資料集的分類結果來做比較，其結果可發現經由圖像增強處理之圖像輸入模型，得到之績效比沒有圖像增強之資料集的分類結果來的高，可證實本研究使用圖像增強之模型之有效性。

表 4-2 圖像增強模型績效

Model	Accuracy	Precision	Recall	F1
Original	0.80924	0.812	0.800	0.806
Blur+CNN	0.82525	0.828	0.827	0.827
Average+CNN	0.83025	0.838	0.830	0.834
HSV+CNN	0.81345	0.812	0.814	0.813

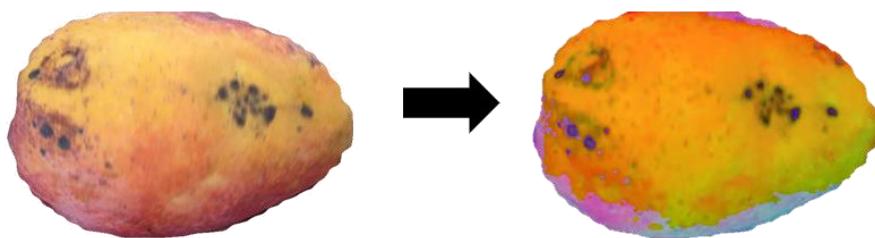


圖 4-6 轉換至 HSV 色彩空間處理示意圖

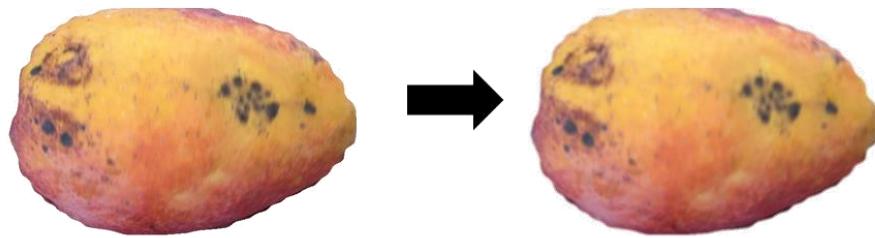


圖 4-7 圖像模糊化處理示意圖



圖 4-8 圖像平均色處理示意圖

4.3 集成學習之模型有效性

為了使辨識精度提升，其本研究將 4 種前處理模型得到之答案，進行集成學習，透過隨機森林(Random Forest)、支撑向量機(Support Vector Machine)與決策樹(Decision Tree)等機器學習分類器來做比較。其驗證方式為使用真實值與驗證集所預測之答案輸入我們的分類器去做訓練，並將測試集所預測之答案，再輸入以訓練好之模型，再做最終預測，得到最終集成學習所預測之答案。結果如表 4-3 所示，由於本研究所比較三種機器學習分類器發現其效果相差不大，但可發現其績效相較其單一前處理模型提升了至少 5%，可證實本研究所使用之集成深度學習模型之有效性。

表 4-3 集成學習使用之分類器績效

Model	Accuracy	Precision	Recall	F1
SVM	0.885	0.88	0.89	0.88
Random Forest	0.8825	0.88	0.89	0.88
Decision Tree	0.8825	0.88	0.89	0.88

4.4 本研究實驗之混淆矩陣

其利用原始圖像輸入至 Mask R-CNN 與 YOLO_V3 做目標檢測，並放進卷積神經網路訓練與分類，得到 Mask R-CNN 模型其績效較佳，並做混淆矩陣如表 4-4 所示，其中 Total population 為總驗證資料數量；True Condition 為實際的類別；Predicted Outcome 為模型輸出類別，0 為 A 類、1 為 B 類與 2 為 C 類，其混淆矩陣可以看出看 B 類分類效果較差。

表 4-4 目標檢測模型其混淆矩陣

		Mask R-CNN 模型		
		Predicted Outcome		
Total population(T=800)		0	1	2
True Condition	0	219	24	0
	1	75	212	6
	2	11	48	205

		YOLO_V3		
		Predicted Outcome		
Total population(T=800)		0	1	2
True Condition	0	227	16	0
	1	76	197	20
	2	11	48	205

由於透過混淆矩陣發現其 B 類辨識效果不佳，發現是因為 B 類芒果較多為變色芒果，其變色會影響模型分類，因此本研究使用圖像增強前處理方式來解決其辨識不佳等問題，如表 4-5 所示，為各圖像增強前處理之混淆矩陣，觀察後其發現辨識效果還是較差。

表 4-5 圖像增強模型其混淆矩陣

		HSV		
		Predicted Outcome		
Total population(T=800)		0	1	2
	0	207	35	1
True Condition	1	55	213	25
	2	3	33	228
		Blur		
Total population(T=800)		Predicted Outcome		
	0	0	1	2
	0	218	25	0
True Condition	1	55	220	18
	2	5	39	220
		Average		
Total population(T=800)		Predicted Outcome		
	0	0	1	2
	0	218	25	0
True Condition	1	45	236	12
	2	7	49	208

由於辨識效果還是較差，因此本研究集成四種前處理模型，並輸入機器學習分類器，其發現分級結果有明顯改善，其發現 SVM 模型 Accuracy 達到 88.5%，其混淆矩陣如表 4-6 所示。

表 4-6 集成深度學習模型其混淆矩陣

		SVM 模型		
		Predicted Outcome		
Total population(T=800)		0	1	2
	0	276	29	4
True Condition	1	28	229	19
	2	0	12	203
		Random Forest 模型		
Total population(T=800)		Predicted Outcome		
	0	0	1	2
	0	276	29	4
True Condition	1	30	227	19
	2	0	12	203
		Decision Tree 模型		
Total population(T=800)		Predicted Outcome		
	0	0	1	2
	0	276	29	4
True Condition	1	31	227	18
	2	0	12	203

4.5 實驗質化合理性說明

本研究透過 Grad-CAM 來觀察來說明此實驗的質化合理性。以下將詳細說明各前處理模型分類結果可視化。

本研究透過 Grad-CAM 將四種前處理模型進行可視化(1)Mask R-CNN 所切割之圖像輸入 CNN 模型(2)填補平均色於背景之圖像輸入 CNN 模型(3)經由 HSV 所轉換之圖像輸入 CNN 模型(4)模糊化之圖像輸入 CNN 模型。並找出此模型著重在看圖像的哪些區域，並且針對為何會將圖像分為 A 類、B 類與 C 類問題來解釋。由於圖像過多因此只舉例 9 種 case 其部份數據集來做 Grad-CAM 示意圖解釋，其標籤 Original 其模型之圖像；Grad-CAM 為找出模型看了此圖像哪些區域；Mix 為將兩張圖像和在一起。以下會詳細解說。

如圖 4-9 所示此模型結果其分類正確，此張圖像為 A 類，經由填補平均色於背景之圖像輸入 CNN 模型所得出之結果分為 A 類，其可以透過可視化結果可看出模型有注意到亮光部分，但其影響非常少，因此將此圖像分為 A 類，其可判斷此模型合理性。

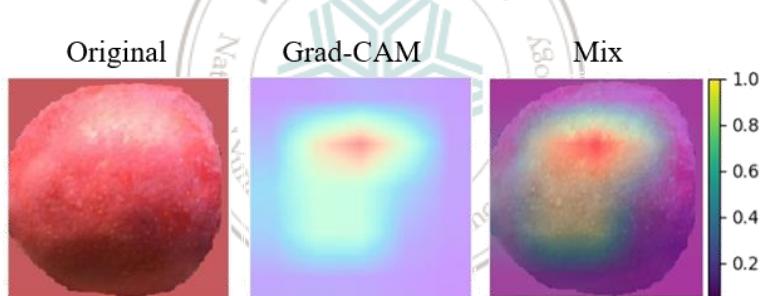


圖 4-9 Grad-CAM 以 00924.jpg 為例使用平均色模型，A 類被分為 A 類

如圖 4-10(b)此模型結果其分類錯誤，此張圖像真實值為 A 類，但經由 Mask R-CNN 所切割之圖像輸入 CNN 模型所得出之結果卻分為 B 類，其可以透過可視化結果可看出其中間紅色區域可能為模型判斷錯誤的依據，但其實這樣的現象不多，因此對模型整體來說並沒有過多影響。

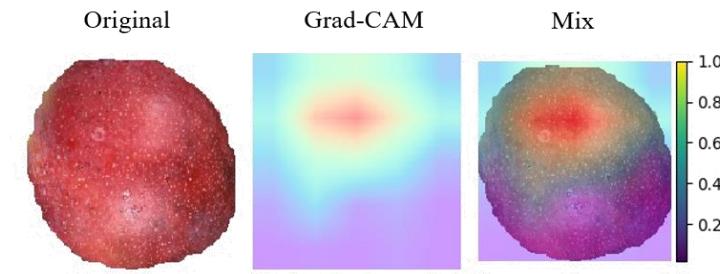


圖 4-10 Grad-CAM 以 03064.jpg 為例使用 Mask R-CNN 模型，A 類被分為 B 類

如圖 4-11 所示由於此模型結果其分類錯誤，此張圖像真實值為 A 類，但經由 HSV 所轉換之圖像輸入 CNN 模型所得出之結果卻分為 C 類，可發現輸出之芒果圖像表面增加了很多缺陷讓模型學習到，因此分類錯誤，且其數量只占少部分，並不影響整體。

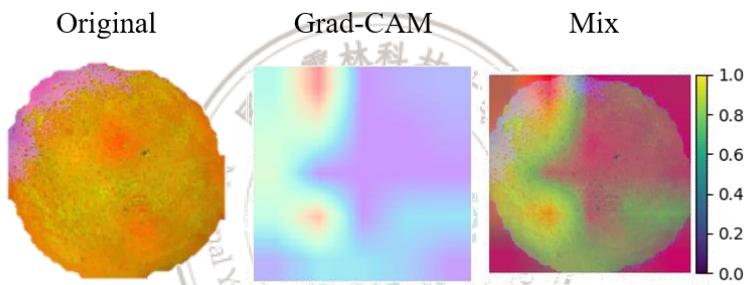


圖 4-11 Grad-CAM 以 05154.jpg 為例使用 HSV，A 類被分為 C 類

如圖 4-12 所示由於此模型結果其分類錯誤，此張圖像真實值為 B 類，但經由 Mask R-CNN 所切割之圖像輸入 CNN 模型所得出之結果卻分為 A 類，透過可視化結果可發現模型注意了明亮的地方，較暗的地方沒注意到，因此不足夠讓模型將此圖像分類到 B 類，其要在訓練時調整明亮度，才能解決此問題。

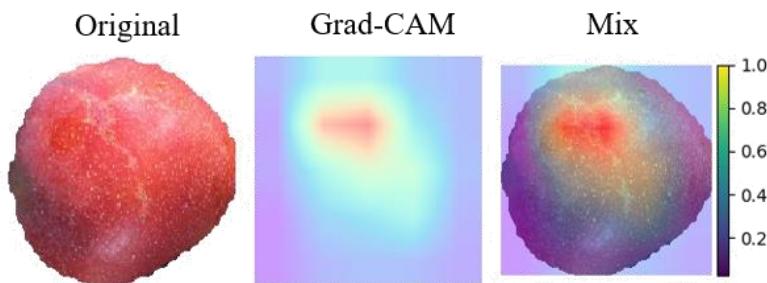


圖 4-12 Grad-CAM 以 06330.jpg 為例使用 Mask R-CNN 模型，B 類被分為 A 類

如圖 4-13 所示此模型結果其分類正確，此張圖像為 B 類，經由模糊化之圖像輸入 CNN 模型所得出之結果分為 B 類，其可以透過可視化結果可看出其模型有注意到其中間變色區域，且模型也確實有學習到，其可判斷此模型合理性。

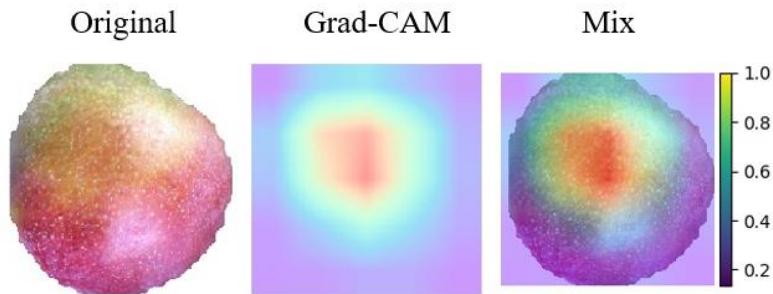


圖 4-13 Grad-CAM 以 07755.jpg 為例使用模糊化模型，B 類被分為 B 類

如圖 4-14 所示由於此模型結果其分類錯誤，此張圖像真實值為 B 類，但經由填補平均色於背景之圖像輸入 CNN 模型所得出之結果卻分為 C 類，其可以透過可視化結果可看出其平均色模型所輸出之圖像，讓模型多學習到背景之其他缺陷，由於此現象較少，所以對於整體系統來說並沒有影響。

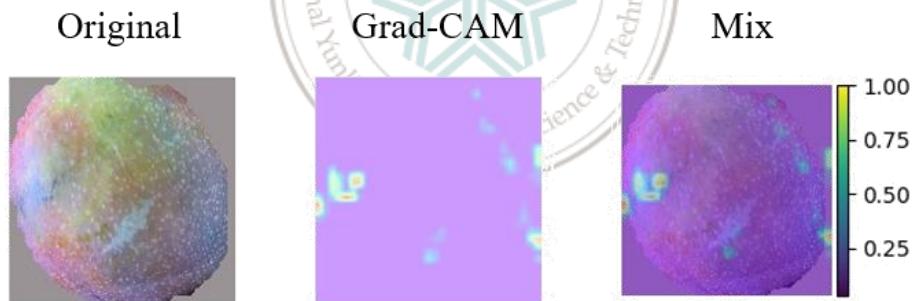


圖 4-14 Grad-CAM 以 03102.jpg 為例使用平均色模型，B 類被分為 C 類

如圖 4-15 所示由於此模型結果其分類錯誤，此張圖像真實值為 C 類，但經由 HSV 所轉換之圖像輸入 CNN 模型所得出之結果卻分為 A 類，其可以透過可視化結果可看出其經由模型輸出後，並沒有看到缺陷位置，在觀察原圖後發現其問題是光線問題，造成光線太亮降低了缺陷可見度，因此讓模型學習不到，此現象可能要調整光亮程度，使圖像更好讓模型判斷缺陷位置。

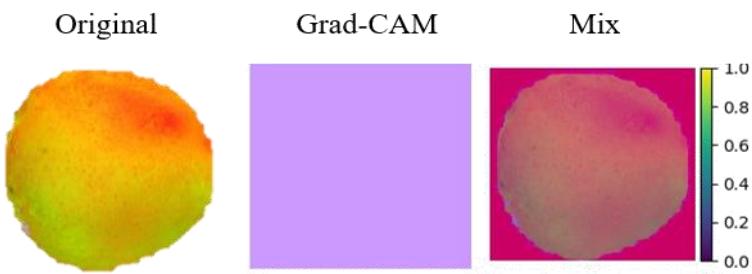


圖 4-15 Grad-CAM 以 02272.jpg 為例使用 HSV，C 類被分為 A 類

如圖 4-16 所示由於此模型結果其分類錯誤，此張圖像真實值為 C 類，但經由模糊化之圖像輸入 CNN 模型所得出之結果卻分為 B 類，其可以透過可視化結果發現，此張圖像透過模糊化減少了缺陷位置的特徵，因此模型判斷錯誤，由於此現象較少，所以對於整體來說較沒有影響。

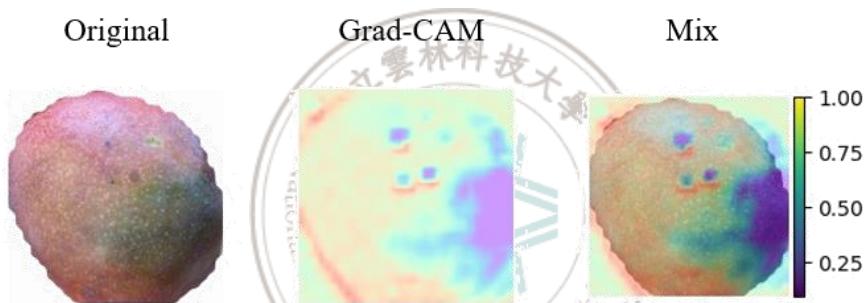


圖 4-16 Grad-CAM 以 01735.jpg 為例使用模糊模型，C 類被分為 B 類

如圖 4-17 所示此模型結果其分類正確，此張圖像為 C 類，經由 Mask R-CNN 所切割之圖像輸入 CNN 模型所得出之結果分為 C 類，其可以透過可視化結果可看出其經由模型輸出後，得到之圖像對於其缺陷位置，其模型有明顯學習到其缺陷，因此分類正確，其可判斷此模型合理性。

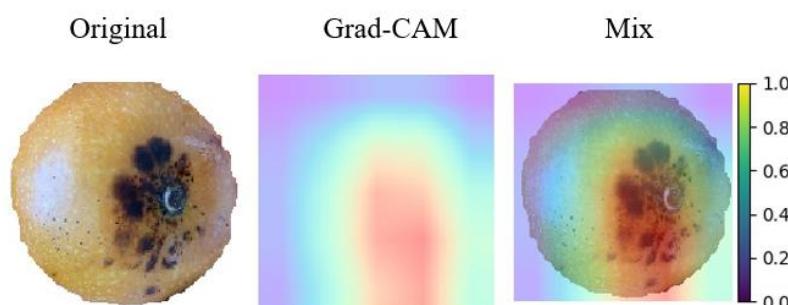


圖 4-17 Grad-CAM 以 02844.jpg 為例使用 Mask R-CNN 模型，C 類被分為 C 類

第五章 結論與未來研究

愛文芒果在旺季的時候每天可以出產到 40 噸的芒果到日本以及韓，這些數字可能還不包括出口到中國大陸，可看出愛文芒果在外銷上之重要性。目前由於缺乏新的技術及自動化的設備，導致其選果效率及品質都不盡理想，廠商也是希望可以透過 AI 辨識以及自動化的設備，來提升產地的效率及品質，進而推動台灣芒果的產銷履歷。過去學者在應用於芒果品質檢測問題時，針對不好辨識之芒果圖像進行分類辨識會遇到兩大問題(1)使用候選框來做目標檢測，方框內之目標的背景也同時會被候選框抓取，若是背景複雜或目標有重疊，因而被背景雜訊所影響，造成精度不佳等問題。(2)將未經過處理的圖像直接輸入至模型中，使模型在訓練過程中消耗大量的運算成本，造成運算時間過長與精度不佳等不符實務上的應用。本研究將提出幾個概念解決上述問題：對於問題(1)提出解決概念(a)透過 Mask R-CNN 提取目標輪廓。對於問題(2)提出兩點解決概念(b)使用多種前處理方式，提高模型強度與精度。(c)將多個模型透過集成式學習(Ensemble Learning)重新訓練結果。結果來說，本研究所提出集成式深度學習模型可以有效對芒果圖像進行分類，在其驗證下，本研究所提出之模型其績效相較於單一前處理模型來的高，其證明本研究方法之有效性。

尤於目前台灣仍然較少人在做採收後處理技術結合 AI 人工智慧，在未來研究中，期望可以希望能透過前處理技術與集成式深度學習模型建立一套自動篩選芒果的系統，降低成本，把 AI 的技術導入台灣整個水果產業鏈當中，可以完成芒果產銷履歷智能化的建構，在農民、人民跟出口商，三方都能得利，建立一個高效率的產銷循環，以促進台灣高競爭力水果的發展。

參考文獻

- [1] Azarmdel, H., Jahanbakhshi, A., Mohtasebi, S. S., & Muñoz, A. R.(2020). Evaluation of image processing technique as an expert system in mulberry fruit grading based on ripeness level using artificial neural networks(ANNs)and support vector machine(SVM). *Postharvest Biology and Technology*, 166, 111201.
- [2] Bahrami, K., & Kot, A. C.(2014). A fast approach for no-reference image sharpness assessment based on maximum local variation. *IEEE Signal Processing Letters*, 21(6), 751-755.
- [3] Bargoti, S., & Underwood, J.(2017, May). Deep fruit detection in orchards. In *2017 IEEE International Conference on Robotics and Automation(ICRA)*(pp. 3626-3633). IEEE.
- [4] Boukellouz, W., & Moussaoui, A.(2019). Magnetic resonance-driven pseudo CT image using patch-based multi-modal feature extraction and ensemble learning with stacked generalisation. *Journal of King Saud University-Computer and Information Sciences*.
- [5] Cha, D., Pae, C., Seong, S. B., Choi, J. Y., & Park, H. J.(2019). Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database. *EBioMedicine*, 45, 606-614.
- [6] Dalal, N.(2006). *Finding people in images and videos*(Doctoral dissertation).
- [7] Dias, P. A., Tabb, A., & Medeiros, H.(2018). Apple flower detection using deep convolutional networks. *Computers in Industry*, 99, 17-28.
- [8] Fukuda, S., Yasunaga, E., Nagle, M., Yuge, K., Sardsud, V., Spreer, W., & Müller, J.(2014). Modelling the relationship between peel colour and the quality of fresh mango fruit using Random Forests. *Journal of Food Engineering*, 131, 7-17.
- [9] Ganesh, P., Volle, K., Burks, T. F., & Mehta, S. S.(2019). Deep orange: Mask R-CNN based orange detection and segmentation. *IFAC-PapersOnLine*, 52(30), 70-75.
- [10] Gang, Z. H. A. O., Jingyu, H. U., Wenlei, X. I. A. O., & Jie, Z. O. U.(2020). A mask R-CNN based method for inspecting cable brackets in aircraft. *Chinese Journal of Aeronautics*.
- [11] Girshick, R.(2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*(pp. 1440-1448).

- [12] Girshick, R., Donahue, J., Darrell, T., & Malik, J.(2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*(pp. 580-587).
- [13] Gregori, V., Morillas, S., Roig, B., & Sapena, A.(2018). Fuzzy averaging filter for impulse noise reduction in colour images with a correction step. *Journal of Visual Communication and Image Representation*, 55, 518-528.
- [14] Han, D., Liu, Q., & Fan, W.(2018). A new image classification method using CNN transfer learning and web data augmentation. *Expert Systems with Applications*, 95, 43-56.
- [15] He, K., Gkioxari, G., Dollár, P., & Girshick, R.(2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*(pp. 2961-2969)
- [16] Hou, L., Wu, Q., Sun, Q., Yang, H., & Li, P.(2016, August). Fruit recognition based on convolution neural network. In *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery(ICNC-FSKD)*(pp. 18-22). IEEE.
- [17] Jahanbakhshi, A., Momeny, M., Mahmoudi, M., & Zhang, Y. D.(2020). Classification of sour lemons based on apparent defects using stochastic pooling mechanism in deep convolutional neural networks. *Scientia Horticulturae*, 263, 109133.
- [18] Jia, W., Tian, Y., Luo, R., Zhang, Z., Lian, J., & Zheng, Y.(2020). Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot. *Computers and Electronics in Agriculture*, 172, 105380.
- [19] Jin, X., Chen, G., Hou, J., Jiang, Q., Zhou, D., & Yao, S.(2018). Multimodal sensor medical image fusion based on nonsubsampled shearlet transform and S-PCNNs in HSV space. *Signal Processing*, 153, 379-395.
- [20] Jung, J. J., Yu, D. J. G., Zeng, A., Chen, M. H., Shi, Y., Nassisi, M., ... & Hoang, Q. V.(2020). Correlation of Quantitative Measurements with Diabetic Disease Severity Utilizing Multiple En-Face OCTA Image Averaging. *Ophthalmology Retina*.
- [21] Kardani, N., Zhou, A., Nazem, M., & Shen, S. L.(2020). Improved prediction of slope stability using a hybrid stacking ensemble method based on finite element analysis and field data. *Journal of Rock Mechanics and Geotechnical Engineering*.

- [22] Khan, M. A., Akram, T., Sharif, M., Awais, M., Javed, K., Ali, H., & Saba, T.(2018). CCDF: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep CNN features. *Computers and electronics in agriculture*, 155, 220-236.
- [23] Książek, W., Hammad, M., Pławiak, P., Acharya, U. R., & Tadeusiewicz, R.(2020). Development of novel ensemble model using stacking learning and evolutionary computation techniques for automated hepatocellular carcinoma detection. *Biocybernetics and Biomedical Engineering*, 40(4), 1512-1524.
- [24] Kumar, S. D., Esakkirajan, S., Bama, S., & Keerthiveena, B.(2020). A microcontroller based machine vision approach for tomato grading and sorting using SVM classifier. *Microprocessors and Microsystems*, 76, 103090.
- [25] Lee, J., Kong, T., & Lee, K.(2019). Ensemble patch sparse coding: A feature learning method for classification of images with ambiguous edges. *Expert Systems with Applications*, 124, 1-12.
- [26] Li, C., Zhang, W. H., Li, R., Wang, J. Y., & Lin, J. M.(2020). Research on Star/Galaxy Classification Based on Stacking Ensemble Learning. *AcASn*, 61(2), 21.
- [27] Li, Z., Niu, B., Peng, F., Li, G., Yang, Z., & Wu, J.(2018). Classification of peanut images based on multi-features and SVM. *IFAC-PapersOnLine*, 51(17), 726-731.
- [28] Lin, G., & Zou, X.(2018). Citrus Segmentation for Automatic Harvester Combined with AdaBoost Classifier and Leung-Malik Filter Bank. *IFAC-PapersOnLine*, 51(17), 379-383.
- [29] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C.(2016, October). Ssd: Single shot multibox detector. In *European conference on computer vision*(pp. 21-37). Springer, Cham.[ssd]
- [30] Moallem, P., Serajoddin, A., & Pourghassem, H.(2017). Computer vision-based apple grading for golden delicious apples based on surface features. *Information processing in agriculture*, 4(1), 33-40.
- [31] Momeny, M., Jahanbakhshi, A., Jafarnezhad, K., & Zhang, Y. D.(2020). Accurate classification of cherry fruit using deep CNN based on hybrid pooling approach. *Postharvest Biology and Technology*, 166, 111204.

- [32] Nemade, S. B., & Sonavane, S. P.(2020). Co-occurrence patterns based fruit quality detection for hierarchical fruit image annotation. *Journal of King Saud University-Computer and Information Sciences*.
- [33] Osako, Y., Yamane, H., Lin, S. Y., Chen, P. A., & Tao, R.(2020). Cultivar discrimination of litchi fruit images using deep learning. *Scientia Horticulturae*, 269, 109360.
- [34] Pereira, L. F. S., Barbon Jr, S., Valous, N. A., & Barbin, D. F.(2018). Predicting the ripening of papaya fruit with digital imaging and random forests. *Computers and Electronics in Agriculture*, 145, 76-82.
- [35] Qiao, Y., Truman, M., & Sukkarieh, S.(2019). Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming. *Computers and Electronics in Agriculture*, 165, 104958.
- [36] Rad, R. M., Saeedi, P., Au, J., & Havelock, J.(2018). Human Blastocyst's Zona Pellucida segmentation via boosting ensemble of complementary learning. *Informatics in Medicine Unlocked*, 13, 112-121.
- [37] Redmon, J., & Farhadi, A.(2017). YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*(pp. 7263-7271).
- [38] Redmon, J., & Farhadi, A.(2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
- [39] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A.(2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*(pp. 779-788).
- [40] Ren, S., He, K., Girshick, R., & Sun, J.(2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*(pp. 91-99).
- [41] Semary, N. A., Tharwat, A., Elhariri, E., & Hassanien, A. E.(2015). Fruit-based tomato grading system using features fusion and support vector machine. In *Intelligent Systems' 2014*(pp. 401-410). Springer, Cham.
- [42] Sengupta, S., & Lee, W. S.(2014). Identification and determination of the number of immature green citrus fruit in a canopy under different ambient light conditions. *Biosystems Engineering*, 117, 51-61.]

- [43] Shanmugam, P., Raja, J., & Pitchai, R.(2021). An automatic recognition of glaucoma in fundus images using deep learning and random forest classifier. *Applied Soft Computing*, 107512.
- [44] Shirgahi, H., & Danesh, N.(2011). Presented a fuzzy system to determine barberry product quality based on average color spectra by image processing. *Journal of King Saud University-Science*, 23(1), 105-109.
- [45] Simonyan, K., & Zisserman, A.(2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [46] Singh, P.(2020). A neutrosophic-entropy based clustering algorithm(NEBCA)with HSV color system: A special application in segmentation of Parkinson's disease(PD)MR images. *Computer Methods and Programs in Biomedicine*, 189, 105317.
- [47] Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L.(2016, July). Breast cancer histopathological image classification using convolutional neural networks. In *2016 international joint conference on neural networks(IJCNN)*(pp. 2560-2567). IEEE.
- [48] Tan, K., Lee, W. S., Gan, H., & Wang, S.(2018). Recognising blueberry fruit of different maturity using histogram oriented gradients and colour features in outdoor scenes. *Biosystems engineering*, 176, 59-72.
- [49] Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., & Liang, Z.(2019). Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Computers and electronics in agriculture*, 157, 417-426.
- [50] Tuncer, T., & Ertam, F.(2020). Neighborhood component analysis and reliefF based survival recognition methods for Hepatocellular carcinoma. *Physica A: Statistical Mechanics and its Applications*, 540, 123143.
- [51] Valdez, P.(2020). Apple Defect Detection Using Deep Learning Based Object Detection For Better Post Harvest Handling. *arXiv preprint arXiv:2005.06089*.
- [52] Wan, S., & Goudos, S.(2020). Faster R-CNN for multi-class fruit detection using a robotic vision system. *Computer Networks*, 168, 107036.
- [53] Wu, A., Zhu, J., & Ren, T.(2020). Detection of apple defect using laser-induced light backscattering imaging and convolutional neural network. *Computers & Electrical Engineering*, 81, 106454.

- [54] Yang, X., Zhang, Y., Lv, W., & Wang, D.(2020). Yang, X., Zhang, Y., Lv, W., & Wang, D. Image recognition of wind turbine blade damage based on a deep learning model with transfer learning and an ensemble learning classifier. *Renewable Energy*, 163, 386-397.
- [55] Yu, X., Lu, Y., & Gao, Q.(2020). Pipeline image diagnosis algorithm based on neural immune ensemble learning. *International Journal of Pressure Vessels and Piping*, 104249.
- [56] Yu, Y., Velastin, S. A., & Yin, F.(2019). Automatic grading of apples based on multi-features and weighted K-means clustering algorithm. *Information Processing in Agriculture*.
- [57] Yu, Y., Zhang, K., Yang, L., & Zhang, D.(2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Computers and Electronics in Agriculture*, 163, 104846.
- [58] Zarifneshat, S., Rohani, A., Ghassemzadeh, H. R., Sadeghi, M., Ahmadi, E., & Zarifneshat, M.(2012). Predictions of apple bruise volume using artificial neural network. *Computers and electronics in agriculture*, 82, 75-86.
- [59] Zhang, Y., Lee, W. S., Li, M., Zheng, L., & Ritenour, M. A.(2018). Non-destructive recognition and classification of citrus fruit blemishes based on ant colony optimized spectral information. *Postharvest Biology and Technology*, 143, 119-128.
- [60] Zhao, Y., Gong, L., Zhou, B., Huang, Y., & Liu, C.(2016). Detecting tomatoes in greenhouse scenes by combining AdaBoost classifier and colour analysis. *Biosystems Engineering*, 148, 127-137.
- [61] Zhou, T., Lu, H. L., Yang, Z., Qiu, S., Huo, B. Q., & Dong, Y.(2020). The ensemble deep learning model for novel COVID-19 on CT images. *Applied Soft Computing*, 106885.
- [62] Zuo, L., He, P., Zhang, C., & Zhang, Z.(2020). A robust approach to reading recognition of pointer meters based on improved mask-RCNN. *Neurocomputing*, 388, 90-101.
- [63] BIIC Lab(<https://aidea-web.tw/topic/72f6ea6a-9300-445a-bedc-9e9f27d91b1c>)
- [64] CV2.Blur(https://docs.opencv.org/master/d4/d13/tutorial_py_filtering.html)

- [65] CH.Tseng.(<https://chtseng.wordpress.com/2017/09/12/%E5%88%9D%E6%8E%A2%E5%8D%B7%E7%A9%8D%E7%A5%9E%E7%B6%93%E7%B6%B2%E8%B7%AF/>)
- [66] Otto Group Product Classification Challenge(<https://www.kaggle.com/c/otto-group-product-classification-challenge/discussion/14335>)
- [67] 行政院農業委員會
(<https://agrstat.coa.gov.tw/sdweb/public/trade/tradereport.aspx>)
- [68] 教育部人工智慧競賽與標註資料蒐集計畫(MOE AI competition and labeled data acquisition project)

