



應用電腦視覺方法於國軍副供站水果品質檢測之研究

DOI:10.3966/230674382021111104003

林念德、溫志皓

提要

- 一、隨著生活水準提升及健康意識抬頭，水果已成為日常必備的食品，其外部品質不僅是非常重要的感官屬性，更影響消費者偏好與產品市場的價值。以往執行水果外觀品檢，多仰賴人力而效率不彰，加上從農人口逐年減少，因此，如何精準、迅速的實施品質檢測，以取代傳統的人工作業，為近來政府推動「智慧農業」的重要課題。
- 二、新興的電腦視覺系統和影像辨識已在食品行業中獲得廣泛使用，隨著資通訊技術與設備日益精良，由「深度學習」技術所發展的「卷積神經網路」架構已被證明成為水果品質檢測的強大工具。若能將此技術運用於國軍副供站的水果品質檢測，可確保所供應的水果品質無虞，並彌補因組織調整所精簡的人力。
- 三、研究所得結果：藉由「YOLOv3」演算法所延伸出非破壞性、快速且正確率接近100%的電腦視覺方法，可輔助費時耗力且缺乏客觀性的人工目視檢驗，同時提升國軍副食供應站的水果品質分類及檢驗作業效率。

關鍵詞：品質檢測、電腦視覺、卷積神經網路、YOLO

壹、前言

一、研究背景與動機

臺灣因地處亞熱帶，水果種類多元，「香蕉」及「番石榴」屬於全年生產之重要經濟果樹，^{1、2、3}經行政院農業委員會（農委會）統計，108年度臺灣主要果菜批發市場，年度交易數量總額為61萬6,315公噸，其中以「香蕉」及「番石榴」分別占果品總交易量的6.6%及9.7%為最多。以國軍為例，經國軍副食供應中心（副供中心）統計，108年度國軍各部隊向副食供應站（副供站）採購水果類副食品，占總銷貨數量前兩名亦為「香蕉」及「番石榴」，由於無種植季節區別，可全年供應，年銷貨數量均超過600公噸以上。

在產銷過程中，完善的水果品質分級方法是整體提升品質的主要關鍵，目前傳統產業常用的重量分級機雖可挑選符合規格的水果，卻經常出現相當大的外部品質差異，仍需經過人工再次判斷。⁴因此，如何將水果篩選工作自動化，並以副供站為例，建立一套品質檢測分級模式，作為輔助水果檢驗之依

據，提升整體供貨水平，為本研究重要動機之一。

副供站設立之目的旨在負責提供部隊副食品，如何確保提領產品品質無虞，是最關鍵的基本問題。現行副食品申購作業中的「站部品檢」、「抽驗」或「全項檢查」均耗時費力，其中採買人員更因交換輪替頻繁、實務經驗不足，經常產生品質標準不一或檢查疏漏等問題。此外，近年國防部因應兵力結構調整，檢討保留後勤核心能量，精簡部分後勤人力，也間接影響副供組織調整，原有工作人員逐年遞減。品檢員每日除需負責檢驗約1,700公斤之副食品，更要掌握各類副食品申請調配、抽驗品檢等作業，其工作量及業務壓力亦隨之增加。在後勤人力精簡壓力下，副供站需要準確又有效率的品檢方法，是否可能應用電腦視覺方法建立品質檢測分類模型，藉由與人工目視檢驗比較之實驗結果，探討「電腦視覺」、「專業作業人員」與「一般作業人員」等三者間對於辨識水果效能的差異程度，以輔助副供站水果品質檢驗方法，為本研究重要動機之二。隨著資訊設備進步，

- 1 古關喜之，〈臺灣香蕉外銷自由化之後的銷日香蕉生產、外銷和日本市場〉《地理研究》（臺北市），第66卷，西元2017年，頁111-127。
- 2 朱堉君、蘇博信、周浩平、陳明吟、曾敏南、陳昱初，〈番石榴健康管理技術專刊〉（屏東縣：行政院農業委員會高雄區農業改良場，西元2013年）。
- 3 張明郎，〈番石榴產業發展現況〉《臺中區農業改良場特刊》（彰化縣），第108期，西元2011年，頁1-9。
- 4 Zhang, B., Huang, W., Li, J., Zhao, C., Fan, S., Wu, J., & Liu, C., "Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review," Food Research International, Vol.62 (2014), pp.326-343.

大量數據資料的儲存、運算與分析更為快速且容易，結合「電腦視覺(Computer Vision)」和「機器學習(Machine Learning)」等新興技術發展提供低成本、高精度且非破壞技術的檢測方法成為未來提高農業效率的必要條件。⁵目前廣泛使用的物件偵測技術係採「卷積神經網路」(Convolutional Neural Network, 以下簡稱CNN)，能有效解決電腦視覺在影像辨識領域中，結構越複雜越耗時計算的問題。其中，「YOLO」(You Only Look Once, YOLO)演算法更直接地檢測影像中的物體，⁶並節省運算速度而降低時間成本。所以，如何應用CNN架構，以快速的模型演算法，朝成本低廉、精度穩定的方向建立影像辨識演算法，是本研究關注的動機之三。

二、研究目的

基於上述背景與動機，歸納本研究目的如下：

- (一) 應用電腦視覺方法建立品質檢測之模型，並以副供站供應之「香蕉」及「番石榴」為研究物件，建構標準一致的外部品質檢測分類模式，作為品質分

類依據。

- (二) 透過實驗結果分析，檢視「電腦視覺」、「專業作業人員」與「一般作業人員」三者間，對於辨識水果效能的差異程度。
- (三) 探討以CNN為模型基礎架構，並與人工目視檢驗進行比較，期建立快速、具高精確率的影像辨識演算法，以輔助傳統水果品質檢驗方法。

貳、文獻探討

一、智慧農業

「智慧農業」一詞可追溯至1990年代的「精準農業(Precision Agriculture)」，主要指運用衛星遙測技術(Remote Sensing)來解決農田間存在之差異，遂依其差別準確地灑水、施肥或噴藥，同時降低努力成本，進而達成農業永續發展。⁷換句話說，智慧農業是運用農民既有的耕作知識，加上利用資訊通訊技術(Information and Communication Technologies, ICT)收集、處理和分析數據，

5 Shin, J., Chang, Y. K., Heung, B., Nguyen-Quang, T., Price, G. W., & Al-Mallahi, "A. Effect of directional augmentation using supervised machine learning technologies: A case study of strawberry powdery mildew detection," Biosystems engineering, Vol.194 (2020), pp.49-60.

6 Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. "You only look once: Unified, real-time object detection," Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, (2016).

7 余祁暉，〈國際智慧農業發展策略〉《臺灣經濟研究月刊》(臺北市)，第40卷第3期，西元2017年，頁21-30。

同時配合擁有前瞻科技的設施設備，以智慧化的模式改善作物生產力，降低農業的不可預測性，提高農場營運效率及利潤，並減少環境足跡的一種持續性活動。

我國「智慧農業」計畫中，主軸區分為『智慧生產』與『數位服務』兩面向。⁸其中『智慧生產』方面，是針對農業人口老化，勞動力流失等問題，推行智慧農業機具、輔具的應用。有學者指出，⁹透過智慧農業下的科技裝置、數據分析，可協助農民運用更省力省時工具，提高農作生產效能。主要包含『省力型輔助工具』、『省工搬運裝置』、『智能分級模組』。其中『智能分級模組』為農產品品質檢測設備開發，利用特徵及波段擷取進行非破壞性檢測與病徵分析研究。而其中品質檢測設備，不論在設施或機具方面，均可以有效改善耕作過程及管理農場。

二、水果品質及分級檢測

由於水果品種及產量大幅增加與提升，加上健康意識抬頭，水果已成為每天的必備食物來源之一，對水果品質相對也要求提高。「品質 (quality)」一詞，是指對象的特質、屬性或基本參數。¹⁰水果會因為各種外部品質的特徵、屬性，進而影響消費者的購買意願。所以，水果的外觀品質若能達到購買者心中的期望，則能促成消費行為。一般而言，水果的外部品質是經由顏色、紋理、大小、重量、形狀及其肉眼可見之瑕疵來檢查。^{11、12}對於水果而言，外觀是非常重要的感官特性，它不僅影響其市場價值，也影響消費者挑選水果之喜好及購買行為，還會在一定程度上影響其內部品質。因此有學者指出外部品質是農產品最重要且最直接的感官品質特性。¹³

以往傳統水果分級方式，是以人工肉

- 8 楊舒涵、楊智凱、葉有順，〈智慧農業創造優質從農環境，塑造農業新典範〉《農政與農情》（臺北市），第326期，西元2019年，頁6-11。
- 9 楊智凱、葉有順、王乙涵、張祥傑、陳偉豪、方理、…張建禕，〈智慧農業機具與輔具之應用〉《科儀新知》（新竹縣），第220期，西元2019年，頁4-19。
- 10 Mohd Ali, M., Hashim, N., Bejo, S. K., & Shamsudin, R., “Rapid and nondestructive techniques for internal and external quality evaluation of watermelons: A review,” *Scientia Horticulturae*, Vol. 225 (2017), pp.689-699.
- 11 Costa, C., Antonucci, F., Pallottino, F., Aguzzi, J., Sun, D.-W., & Menesatti, P., “Shape Analysis of Agricultural Products: A Review of Recent Research Advances and Potential Application to Computer Vision,” *Food and Bioprocess Technology*, Vol. 4, No.5 (2011), pp.673-692.
- 12 Zhang, B., Huang, W., Li, J., Zhao, C., Fan, S., Wu, J., & Liu, C., “Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review,” *Food Research International*, Vol.62 (2014), pp.326-343.
- 13 同註12。

眼方式進行外部品質檢測，要花費大量的人力、時間及成本，也較容易受到人的精神因素影響判斷而產生錯誤，造成品質不易受到認可，需要有經驗之人員才能判別，同時缺乏統一客觀的標準，^{14、15}勞力成本也相對提高，無法在時間及效率上達到良好的效果。^{16、17}近來，農業及食品加工業為解決傳統品質評估方法的缺點，廣泛運用「非破壞性檢測」的物理及光電磁原理來檢測樣品。此技術可區分『電腦視覺』以及『光譜技術』，後者包括聲波、核磁共振等。囿於『光譜技術』所需花費的設備成本較高，且其取得的數據龐大、複雜且須在特定實驗室才能完成，故在許多食品檢測應用的接受度較低。¹⁸因此，選擇有效實用的『電腦視覺』，開發適切、高效能和準確的影像處理技術，不僅可加快速度，滿足應用於即時檢測之要求，更可將繁複的影像處理過程輸入至電腦中，

代替大量的人工運算時間，以優化品質檢測的自動化流程。

三、人工智慧與電腦視覺

人工智慧是近年來非常熱門的技術，目的在於使用統計方法或是機器學習的演算法，處理大量資料，藉以進行分類、估計、預測的功能。其中，人工神經網路的發展，從2015年的ImageNet舉辦的比賽中，有了大幅的進步與效果展現。自此，人工神經網路對於圖像分類的精準度，已經優於人類。當時所使用的即是CNN的深度學習網路技術。本研究所使用的YOLO方法，是一種基於CNN架構的「電腦視覺技術」。

「電腦視覺」是利用電腦來模仿人類的視覺系統；更深一層的電腦視覺，則是針對一個客觀的物體，根據影像構建出明確清晰且有意義的描述。¹⁹也可將「電腦視覺」稱之為「人工視覺」，其目的在於透過電腦處理影

- 14 李汪盛，〈水果品質檢驗新技術近紅外線分析法〉《桃園區農業專訊》（桃園市），第29期，西元1999年，頁3-3。doi:10.29566/XLZY.199909.0002
- 15 林子傑，〈農產品非破壞性檢測技術〉《臺中區農業改良場特刊》（彰化縣），第105期，西元2010年，頁39-42。doi:10.29563/ZHWHGX.201012.0021
- 16 同註15，頁39-42。
- 17 鄭榮瑞、鍾瑞永、楊清富，〈近紅外線非破壞性水果品質檢測器〉《臺南區農業專訊》（臺南市），第58卷，西元2006年，頁9-12。
- 18 Chen, Q., Zhang, C., Zhao, J., & Ouyang, Q., "Recent advances in emerging imaging techniques for non-destructive detection of food quality and safety," TrAC Trends in Analytical Chemistry, Vol.52 (2013), pp.261-274.
- 19 Ballard, D. H., & Brown, C. M., Computer vision (United States: Prentice Hall, 1982).

像，並提取現實世界中物體的各項特徵，²⁰運用電腦結合攝影機，代替人眼對物體進行觀察、測量以及辨識與檢測的技術，透過學習和運算，建立具有智慧的「電腦視覺」。

『物體偵測 (Object Detection)』是「電腦視覺」中的重要應用，目前發展非常成熟的「YOLO」系列演算法已被廣泛運用於各個影像辨識領域中，²¹例如醫學影像、智慧運輸、安全監控等領域，近年來更成為影像辨識主流之一。物體偵測的演算法可以分為兩類，一類是提取區域檢測 (Region Proposal) 的二階段式 (Two Stage) 演算法；另一類是僅使用 CNN 架構直接預測目標的類別與位置的單階段式 (One Stage) 演算法。「YOLO」系列演算法將檢測問題轉換為迴歸問題，不需經過選取目標程序，而是直接利用迴歸方式計算出多個邊界框的坐標及各項物體類別的機率數字，可使訓練速度提升，辨識速率加快。

四、文獻小結

在全球農業面臨勞力不足等問題，生產風險隨之增加。因此，各國積極地推動智慧農業，而臺灣為提高水果產業產品品質，除需重視市場標準，更要精簡勞力成本，導入資通訊技術，利用非破壞性檢測設備，快速檢核外部品質，藉以提高作業效率並帶動整體

經濟增長。同時，資訊硬體設備不斷精良，電腦視覺及影像辨識技術發展出的演算法也日益進步，深度學習趨向成熟且穩定。特別是應用於水果品質分級，CNN 提供更高的檢測精確度。其中，又以「YOLO」屬流行的物體檢測演算法，其具有『速度快』、『易訓練』且『結構簡單』等特性，在影像辨識的速度、準確率取得領先地位。

綜合前述觀點，本研究將藉電腦視覺方法所擴展的 CNN，探討應用於水果品質檢測之相關技術，並採用「YOLO」演算法做為基礎架構，期建立省時、便利且具精確度之辨識模型，以提升整體水果供應品質及檢驗分類模式。

參、研究方法

一、研究架構

本研究根據國軍副供中心所印頒的「水果類副食品品質規格表 (以下簡稱品規表)」，向副供站取得各等級的研究物件，並由副供站之品檢員先行完成人工目視的品質檢驗，挑選符合規範之「特級品」、「優級品」及「不良品」。接著，建置照明光源充足之拍攝實驗場景，使用數位相機對每一個研

20 Vélez, J., Moreno, A., S?nchez, Angel, S?nchez, J., Visi?n por computador (Spain: Ed. Universidad Rey Juan Carlos, 2003).

21 Redmon, J., Divvala, S., Girshick, R., & Farhadi, A., "You only look once: Unified, real-time object detection," Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, (2016).

究物件進行影像擷取。再來運用「LabelImg」工具，標記影像中物件之位置，且予以定義及建立分類名稱，蒐集影像資料後分別建立訓練、測試及驗證資料庫。透過「YOLO」演算法取得特徵值以產生出分類模型，並使用測試資料計算其模型精確率，最後與人工目視品檢結果進行比較。

二、實驗環境與設備

本研究實驗環境中，由於照明條件影響拍攝影像品質甚鉅，是獲得評估食品影像的重要前提，而且高品質的圖像可有效減少後續影像處理步驟的時間，進而降低影像處理的成本。²²因此，本研究架設一長度、寬度及高度均為40公分的桌上攝影棚，並搭配LED照明設備，以提供取像環境均勻且充足之光源，可更有效拍攝物件。

另外，為使拍攝實驗結果之物件特徵擷取容易，實驗攝影棚使用白色背景布，藉以獲得良好影像品質。拍攝完畢後，影像資料分別輸入電腦中儲存，再利用影像標記工具「LabelImg」框選本研究所要標記的研究物件，製作訓練深度學習的資料庫，並進行不同等級分類之標籤命名。

三、研究物件

(一) 選擇研究物件

根據研究背景與動機之論述，挑選「香蕉」以及「番石榴」作為本研究之研究物件，並依照國軍副供中心核定的『國軍副食品品質規格表（品規表）』內容為各等級的分類標準，而其相關標準係分別參考「行政院農業委員會」與「臺北農產運銷股份有限公司」等專業單位所編印之『蔬果品質分級標準暨包裝規格手冊』²³以及『果菜分級包裝手冊』²⁴的規格明確訂定，詳如表一所示。每一項物件在進行影像擷取前，均由專業品檢人員（副供站站長及品檢員）完成初步人工分類，確保建立之模型符合規格表律訂的分類標準。

本研究共取得51根香蕉、53顆番石榴，共合計104個研究物件，其中分別包括「香蕉」：特級品17根、優級品23根及不良品11根；「番石榴」：特級品16顆、優級品20顆及不良品17顆。影像部分陸續拍攝研究物件，共六個等級各約1,000張，故總計6,130張影像資料。

(二) 訓練、測試與驗證資料

在電腦視覺中，對於所有預先收集、建立的完整數據集，將隨機分為三組：「訓練、

22 Du, C.-J., & Sun, D.-W., "Recent developments in the applications of image processing techniques for food quality evaluation," Trends in Food Science & Technology, Vol.15, No.5 (2004), pp.230-249.

23 蔬果品質分級標準暨包裝規格手冊（行政院農業委員會農糧署），西元2012年。

24 果菜分級包裝手冊（臺北農產運銷股份有限公司），西元2016年。

表一 水果類副食品品質規格表

編號	品項	級別	品質規格	圖示
1	香蕉	特級品	同一品種，果肉熟度適中，串型完整，無壓傷，色澤優良、無病蟲害及其他傷害。每把平均果指重量200公克以上（每年11月至翌年4月150公克以上）。	
2		優級品	同一品種，果肉熟度適中，果型完整，無壓傷，色澤良好，無嚴重病蟲害及其他傷害。每把平均果指重量140公克以上（每年11月至翌年4月123公克以上）。	
3		不良品	因病害、蟲害或栽培、採收、收穫後處理及流通過程等造成之腐爛、傷疤、水傷、裂痕、擦傷、壓傷等，影響外觀及品質者。	
4	番石榴	特級品	同一品種，成熟適度，形狀完整，清潔，無腐爛碰傷，無病蟲害及其他傷害，每顆350公克以上，大小規格分開包裝。	
5		優級品	同一品種，成熟尚適度，形狀及鮮度良好，無腐爛，無碰傷，無嚴重病蟲害及其他傷害，每顆200公克以上，大小規格分開包裝。	
6		不良品	果肉過熟或有果裂現象，果皮有腐爛、壓傷、水傷，嚴重之病蟲害斑等。	

資料來源：國軍副食品品質規格表

測試以及驗證資料」。²⁵相關教科書提及在分類問題的模型建構過程中，訓練與測試資料的分割比例是主觀的。為考量本研究模型在同一研究物件、不同角度之辨識精確度以及穩定程度，將數據集區分為訓練資料及測試資料，其比例約略為90%及10%；而此數據也

正是「YOLO」演算法所採用的分割比例。²⁶

本研究拍攝研究物件總共6,130張，接著進行影像資料篩選過程，將「物件特徵不明顯」以及「重複角度」之資料剔除並重新整理。最終，經過篩選處理後之影像，計有1,768張資料集，並參考上述比例作為實驗設

25 Koirala, A., Walsh, K., Wang, Z., & McCarthy, C., "Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of 'MangoYOLO' , " Precision Agriculture, Vol.20, NO.6 (2019), pp.1107-1135.

26 Redmon, J., & Farhadi, A., "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, (2018).

計，分割為「訓練集」：1,591張，用於建立模型，以及「測試集」：177張，作為測試並優化模型，使模型完成訓練後，可驗證其精確度後再針對影像作出合理的修正。

四、YOLO的逐次改變

常見的「深度學習」模式大致有兩類，分別為運用『區域檢測』的二階段式演算法以及直接利用『迴歸分析』的單階段式演算法。其中「YOLO」系列演算法屬於後者，採用深度學習技術作為檢測分類基礎，輔以CNN架構模式的單階段式演算法。將目標檢測問題視為迴歸分析問題，使用單一網路結構設計，用以檢測位於影像中任一位置的物體，而發展出之物體偵測演算法。「YOLO」系列演算法目前發展至第三代版本，由於每代版本均基於前代之基礎不斷改良進步，因此將自『YOLOv1』、『YOLOv2』至『YOLOv3』說明物體檢測的概念及原理，略述如下：

(一) YOLOv1原理

第一代版本非常簡潔，其重要的步驟為：「將輸入之影像縮放(resize)至特定尺寸，執行CNN結構，由模型輸出信賴指數，並透過設定之門檻值與非極大值抑制法(Non-Maximum Suppression, NMS)得出偵測結果」。其原理概述如下：

1. 輸入原始影像，將影像縮放至416X416的尺寸大小。
2. 在訓練階段，影像劃分為SXS之網格(Grid Cell)，影像的物體中心會位於某網格，該網格即標記此物體，使網格

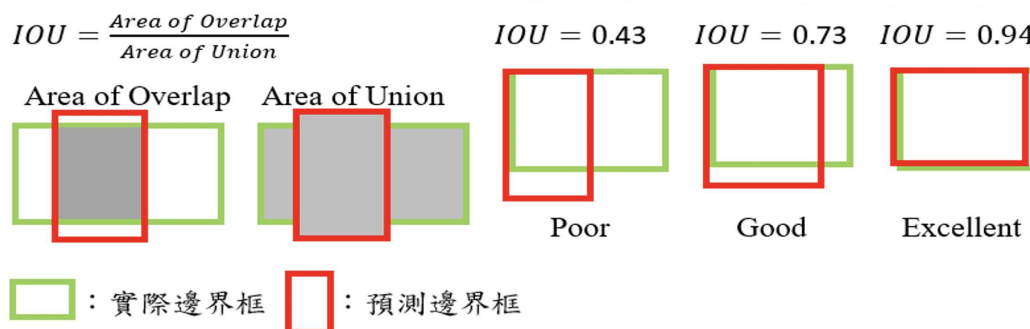
自動學習偵測物體之中心位置。每個網格分別預測「邊界框(Bounding Box)，B」和「是否含有(Classes)物體的機率，C」。其中邊界框包含五個預測值，分別為 x 、 y 、 w 、 h 、 $confidence$ 。 (x, y) 代表網格的邊界框之中心座標、 (w, h) 表示邊界框的寬度和高度，而 $confidence$ 代表邊界框之信賴分數(Confidence Score)，意即是否含有物體 $P_r(Object)$ ，乘以預測邊界框與實際邊界框(Ground Truth)的準確度 IOU_{pred}^{truth} 之數值，以公式表示如下：

$$confidence = P_r(Object) \times IOU_{pred}^{truth}$$

其中的IOU (Intersection Over Union, IOU) 係指「兩個物體的交集(Overlap)」除以「兩個物體的聯集」，IOU數值介於0到1之間，數值越高表示兩個框彼此位置及大小越相似，如圖一所示。

3. 網格還會預測「屬於某類物體類別的條件機率」 $P_r(Class_i|Object)$ 。此機率指的是，若該網格含有某物體且預測為 $Class_i$ 類別之機率。故將此物體的條件機率乘以前述各別邊界框的信賴分數，即為邊界框中針對特定類別物體的信賴分數，以公式表示如下：

$$\begin{aligned} & P_r(Class_i|Object) \times confidence \\ &= P_r(Class_i|Object) \times P_r(Object) \times IOU_{pred}^{truth} \\ &= P_r(Class_i) \times IOU_{pred}^{truth} \end{aligned}$$



圖一 IOU示意圖

資料來源:Koirala et al. (2019)。²⁷

4. 綜上，當影像偵測出可能為此物體的中心點（假設為紅點），其所對應長寬的紅色框即為預測邊界框之一，如圖二（A）所示。輸入的影像中，共有 $7 \times 7 = 49$ 個網格，每網格各有兩個邊界框，候選的邊界框共 $49 \times 2 = 98$ 個，如圖二（B1）所示。此98個邊界框必須大於預先定義之門檻值（Threshold），否則視為過度重疊而捨棄，再用NMS方式將重疊的邊界框刪除，重覆執行此步驟，直至每個類別留下預測值最高的邊界框，即選出物體之最佳位置，如圖二（B2）的紅框、紫框和黃框。同時，YOLO輸出 7×7 的網格算出各類別的機率，並取機率最大的某類物體代表為此網格的類別，如圖二（C）。最後結合「選出之物體位置」和「對應網格的物體類別」，便可辨識

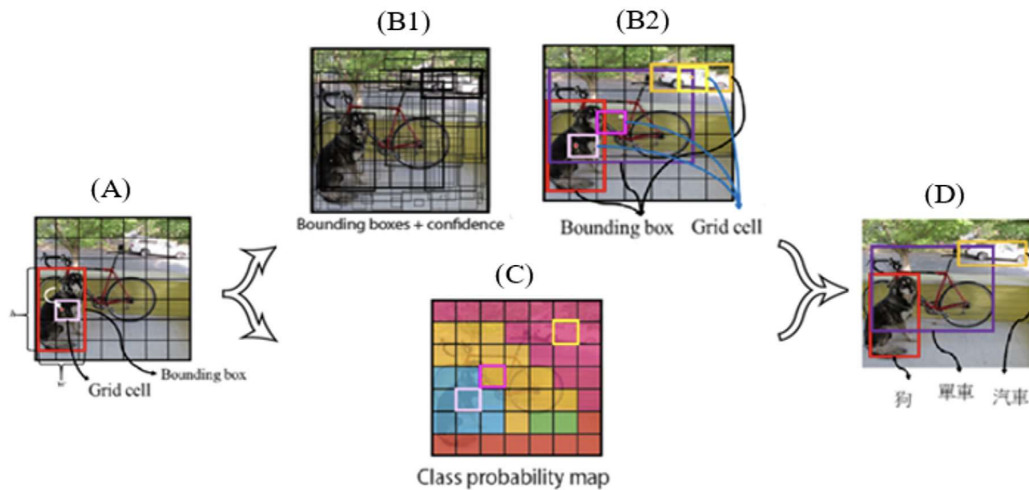
此物體之位置且屬於何種類別，如圖二（D）所示。

（二）改善後之YOLOv2

相對於第一代版本，「YOLO」第二代持續保持處理速度的基礎，除了將網路結構簡單化，還加入新穎概念以增強性能，使其更易學習，進而從更準確、速度更快、辨識更多物體的三個方面進行改善。

1. 在準確方面，網路結構增加「Batch Normalization, BN」。其作用是使每一層網路結構中間加入一個標準化，讓資料的分布穩定，訓練的收斂速度加快，防止過擬合。使用「高解析度分類器（High Resolution Classifier）」輸入影像，使分類器可適應更高解析度的圖像；另外，網路結構加入「Passthrough層」，經由合併前一層較小尺寸的特徵圖（Feature

27 Koirala, A., Walsh, K. B., Wang, Z., & McCarthy, C., "Deep learning – Method overview and review of use for fruit detection and yield estimation," Computers and electronics in agriculture, Vol.162 (2019), pp.219-234.



圖二 YOLOv1物件辨識示意圖

資料來源: Redmon et al. (2016)。²⁸

Map) 保留更精細的特徵，進行多元尺度的訓練，以使模型能適應多樣化的影像。

2. 由於YOLOv1在每個網格中只預測一項類別以及兩個邊界框（共98個邊界框），且直接預測信賴分數的值。而改進後的YOLOv2增加成 13×13 網格的特徵圖，並導入其他學者的方法，²⁹使用錨點框（Anchor Box）作為預先定義邊界框的候選區，再依據錨點框的位置偏移量進行預測，定位出各五個邊界框（共845個邊界框），檢測網格中是否存在物體，如圖三。

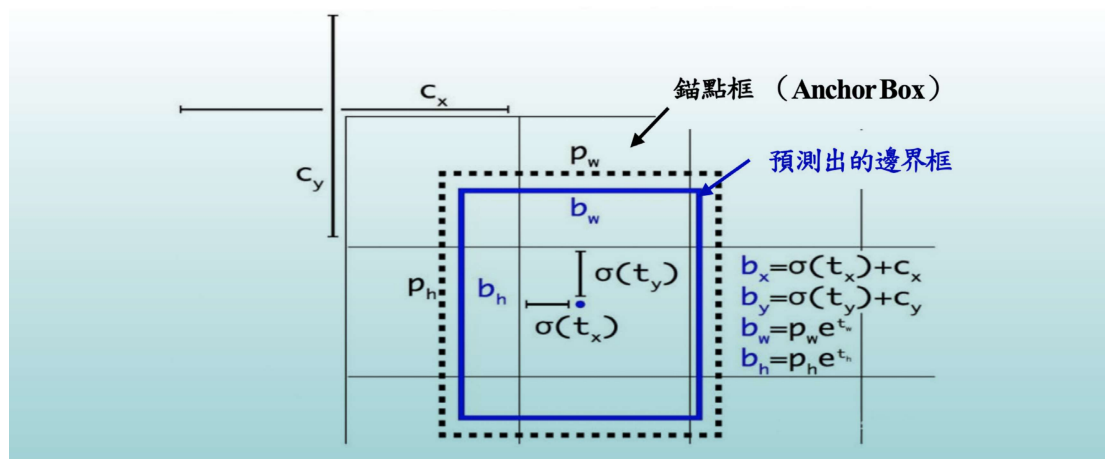
3. 在速度方面，YOLOv2使用全新的網路結構「Darknet-19」，具有十九層卷積層及五層最大池化層（Max Pooling），並加入BN技術來穩定訓練以及對模型進行標準化。
4. 最後在辨識物體方面，使用ImageNet比賽用的數據及結合數量較大的數據集進行模型訓練，使得第二代版本的YOLO能夠辨識超過9,000種的物體，即使檢測多種項目的測試樣本，也能成功分辨物體。

（三）提升後之YOLOv3

1. 結合多尺度的特徵預測：第三代YOLO

²⁸ 同註21。

²⁹ Ren, S., He, K., Girshick, R., & Sun, J., "Faster r-cnn: Towards real-time object detection with region proposal networks," Paper presented at the Advances in neural information processing systems, (2015).



圖三 錨點框、預測邊界框示意圖

資料來源：Redmon and Farhadi (2017)。³⁰

版本結合Lin et al.等學者提出之特徵金字塔網路 (Feature Pyramid Networks, FPN) 概念，³¹將輸入影像映射成三個尺寸 (13X13、26X26、52X52) 的SXS個特徵圖。此三個特徵圖分別獨立檢測，並使用兩次上採樣層 (upsampling)，使小目標檢測效果明顯提升。

2. 定位資訊預測：三個不同尺寸特徵圖各分配三個錨點框，作為預測邊界框的先驗框。每個網格預測三個邊界框以對應三個錨點框。每個網格輸出 $3 \times (4+1+C)$ 個參數值，而此參數同時也是每個尺寸

下最終輸出的特徵向量。

3. 分類資訊預測：利用處理重疊多標籤分類的羅吉斯回歸分類器來計算錨點框；從九個錨點框中找到信賴分數目標存在可能性得分最高的設為該網格的邊界框，再以邊界框與實際邊界框的IOU值作為判定標準，與每個實際邊界框對應。
4. 使用更新的「Darknet-53」作為特徵擷取的網路結構，以ResNet網路的快捷 (Shortcut) 方式，引入殘差 (Residual) 結構，避免出現梯度消失或梯度爆炸。Red-

30 Redmon, J., & Farhadi, A., "YOLO9000: better, faster, stronger," Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, (2017).

31 Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S., "Feature pyramid networks for object detection," Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, (2017).

mon and Farhadi等作者也提及，³²YOLO v3的網路架構以原始影像為輸入資料，其中「建立模型的類別數量乘以2,000次的總數」為演算法建議設定的迭代次數，便可直接提取目標區域的特徵參數值，再經過一次CNN卷積處理，影像會加入評估後的邊界框位置以及預測的物體分類機率作為輸出資料。因此對於物體偵測的訓練過程簡單且直觀，能有效提升辨識速度。

五、模型評估

(一) 評估方法

建立分類模型後，需要透過評估指標來衡量及驗證模型的性能與效益。而「混淆矩陣 (Confusion Matrix)」在監督式學習領域的分類問題中，已被視為重要的決策工具

之一。例如運用於即時識別果園內水果之種類。³³「混淆矩陣」可同時列舉分類問題的所有可能情況，將不同分類算法的混亂程度予以可視化，亦能定義許多分類性能評估指標的算法。本研究係以香蕉及番石榴等兩種水果各三個等級區分，共六類分類問題，因此我們關注於模型可以藉由多元混淆矩陣呈現正確分類出水果級別的效果。多元混淆矩陣可同時列舉分類問題的所有可能情況，將不同分類算法的混亂程度予以可視化，亦能定義許多分類性能評估指標的算法。

多元混淆矩陣的直行代表實際的類別，橫列代表預測的類別，矩陣內顯示了 A_i 至 A_n 多類別分類任務的混淆矩陣之基本形式，如表二所示。其中， N_{ij} 代表實際上屬於 A_i 類的樣本數，但被歸類為 A_j 類。

表二 混淆矩陣表

預測類別		實際類別			
		A_1	$\cdots A_i \cdots$	A_n	
		N_{11}	$\cdots N_{i1} \cdots$	N_{n1}	
	A_1				
	$\vdots$				
	A_j	N_{1j}	$\cdots N_{ij} \cdots$	N_{nj}	
	$\vdots$				
	A_n	N_{1n}	$\cdots N_{in} \cdots$	N_{nn}	

資料來源：Deng, Liu, Deng, and Mahadevan (2016)。³⁴

32 同註26。
33 Saedi, S. I., & Khosravi, H., “A deep neural network approach towards real-time on-branch fruit recognition for precision horticulture,” Expert Systems with Applications, Vol.159 (2020), p.113594.
34 Deng, X., Liu, Q., Deng, Y., & Mahadevan, S., “An improved method to construct basic probability assignment based on the confusion matrix for classification problem,” Information Sciences, Vol. 340-341 (2016), p.250-261.

其餘關於模型表線的正確率 (Accuracy)、召回率 (Recall) 及精確率 (Precision)、平均精確率 (Average Precision, AP)、整體平均精確率 (Mean Average Precision, mAP)、F1度量 (F-1 measure, F1) 及 Kappa 係數 (Kappa coefficient, Kappa) 等七類指標，個別意義分別說明如下列公式。

1. 正確率 (Accuracy)：無特定的關注類別，僅作為評估整體模型之正確預測能力。表示分類模型預測正確的總數占所有樣本的比例。

$$Accuracy = \frac{\sum_{i=1}^n N_{ii}}{\sum_{i=1}^n \sum_{j=1}^n N_{ij}}$$

2. 召回率 (Recall)：在分類模型中，表示正確分類的樣本數占實際為該類樣本之比例，又稱查全率、命中率或真陽性率。

$$Recall_i = \frac{N_{ii}}{\sum_{k=1}^n N_{ik}}$$

3. 精確率 (Precision)：指分類模型中，衡量預測特定類別的樣本數，而實際亦為該類別的樣本數占有多少比例，又稱查準率。

$$Precision_i = \frac{N_{ii}}{\sum_{k=1}^n N_{ki}}$$

4. 平均精確率 (Average Precision, AP)：常使用於物體偵測的評估指標中，³⁵如下列公式所示。簡單來說將PR曲線平滑化，並取橫軸Recall值 (r) 0至1的11個等分點的Precision值 (p)，計算其平均值，即為AP值。表現較好的分類模型，AP值越高，代表模型在單個類別的好壞。

$$AP = \frac{1}{11} \sum_{r \in \{0, 0.1, 0.2, \dots, 1\}} p(r)$$

5. 整體平均精確率 (Mean Average Precision, mAP)：是物體檢測的重要指標，係由每個類別的AP值加總後再取平均，同樣介於0至1之間，它可代表通盤考量每一個類別狀況下所作出的平均值，如下列公式所示，mAP值越大表示模型在所有類別更精確。

$$mAP = \frac{1}{classes} \sum_{i \in classes} AP_i$$

6. F1度量 (F-1 measure, F1)：F1度量是基於召回率 (Recall) 和精確率 (Precision) 的調和平均數，其值介於0至1之間。公式如下所示：

$$F1_i = 2 * \frac{Recall_i * Precision_i}{Recall_i + Precision_i}$$

35 Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A., "The Pascal Visual Object Classes (VOC) Challenge," International Journal of Computer Vision, Vol.88, No.2 (2010), pp.303-338.

7. Kappa係數(Kappa coefficient, Kappa)：
Kappa係數的計算基礎來自於分類模型的混淆矩陣，其值介於0至1之間。可以用於衡量分類精度。 P_o 表示每一個類別中，被正確分類的樣本數量和除以總樣本數，亦即總體分類的精度。 P_e 表示每一個類別中，真實樣本數量和被預測命中樣本數量的乘積。公式如下所示：

$$kappa = \frac{P_o - P_e}{1 - P_e}$$

最後，本研究針對模型性能選擇採用「混淆矩陣」以及「mAP值」作為綜合衡量之評估指標，以進行「電腦模型」與「人工目視」的測試分析。

(二) 驗證結果

電腦視覺在提出新的技術「深度學習」後，有了極大的突破，影像的準確性則在於儘可能預測出影像中的內容，讓電腦能夠在識別或分類超過1,000個項目的特定任務中超越人類表現。因此，本研究將採立意抽樣，招募受試者參與驗證階段，其中受試者之身分區分兩種類型，分別為專業從事相關副供作業人員，如副供站之站長、品檢員等5員；另一類則為一般專業人員，如擔任部隊伙委、採買人員等5員，合計共10名受試者。接著將60張研究物件之驗證資料影像，以電腦播放圖片方式，計算受試者完成辨識的所需時間以及分類水果各等級之準確率，再比較

電腦模型與人工目視兩者辨識之整體平均精確率及時間。

六、研究結果

(一) 研究資料與模型建立

本研究採用的「YOLOv3」演算法，屬於深度學習領域中使用卷積神經網路的模型架構，通常需要透過大型數據集才能實現並滿足模型的通用性(Kamilaris & Prenafeta-Boldú, 2018)。研究一開始使用影像資料中的5,517張「訓練集」進行學習，但卻無法獲得良好分類結果的訓練權重檔，遂針對影像資料進行分析，發現大部分影像資料特徵並不明顯，且部分影像因拍攝角度過小而造成影像資料重複性過高，導致模型未能有效擷取及學習物件特徵。因此，重新進行影像資料篩選過程，將「物件特徵不明顯」以及「重複角度」之資料剔除並整理。

最終，經過篩選處理後之影像，計有1,768張資料集，如表三所示，再分割為「訓練集」：1,591張，用於建立模型，以及「測試集」：177張，作為測試並優化模型，使模型完成訓練後，可驗證其精確度後再針對影像作出合理的修正。

接續利用1,591張影像資料，進行模型的初始訓練，將所有影像之特徵值輸入模型運算，參數經由不斷迭代，使模型的輸出結果能吻合物件辨識的訓練目標；再將分割後的177張用於測試模型的辨識效能，並優化後續結果。接著，從所有「待訓練的模型」中做比

表三 研究物件數量處理後統計表

編號	品項	區分	數量	原始影像數量	人工挑選後之影像數量	比例
1	香蕉	特級品	17	1,005	289	1
2	香蕉	優級品	23	1,022	391	1.3
3	香蕉	不良品	11	1,003	187	0.6
4	番石榴	特級品	16	1,005	272	0.9
5	番石榴	優級品	20	1,091	340	1.1
6	番石榴	不良品	17	1,004	289	1
總計			104	6,130	1,768	

資料來源：本研究整理。

對，找出能使「訓練集」與「測試集」產生最小損失值(loss)的模型權重檔，即為本研究之最佳模型，隨後覆蓋至初始模型設定值，重新進行最後訓練階段，即完成訓練分類模型之建立。

然而，在原本訓練階段「YOLOv3」模型建議設定的迭代次數為每個類別各2,000回合，即可得出待訓練的分類模型。但套用至本研究之設定時，卻因迭代次數太少、模型過於簡單、學習能力較弱，且在照片數據複雜度較高的情況下，出現訓練不足、無法辨識研究物件的「欠擬合(under fitting)」現象。因此，本研究將模型參數進行微調整，總共區分三個部分，分述如下：

YOLOv3模型所標記分類的數量自原本的80個物體，降為本研究所需辨識的六種水果品質級別，以符合本研究目標。

因為前述分類類別改變，模型中的過濾器(filters)數量隨之修正計算為 $3 \times (4 + 1 + C) = 33$ 個。

初始模型迭代次數每類別各為2,000次，本次辨識物件計六個類別，依照原始設定共迭代1萬2,000次回合，修改後為每個類別2萬次，相當於六個類別，總共迭代12萬次回合。

最後，重新調整參數設定並使用「訓練集」運算，設定迭代每1萬次會得出一個權重檔，經由測試資料集進行評估，隨著迭代增加，模型性能相對提高。經過第12萬次迭代回合，運算時間總計約700多小時，訓練結束的權重檔代入模型運算，成功辨識出「香蕉」以及「番石榴」等六個級別，並獲得接近100%的正確率才終止迭代，完成本研究之訓練模型建立。

(二) 模型評估結果

本研究使用「混淆矩陣」、「整體平均精確率(mAP值)」與「辨識所需時間」等指標，針對模型實施性能評估。訓練完成之分類模型，整體效能表現良好，辨識結果具有一定程度的鑑別能力。使用177張「測試集」進行分析，各項性能評估整理如表四，結果說明如

後：「電腦模型」的『正確率』為0.99、『召回率』為1.00、六個級別之『精確率』為0.99，在『整體平均精確率』於門檻值設定mAP@0.50時為1.00；整體辨識所需時間約55秒。

(三) 模型驗證結果

本研究另外依每個級別再次挑選並拍攝各10張影像資料作為「驗證集」，合計共60張影像，分別運用「電腦模型」與「人工目視」等兩種影像辨識方式進行比較分析。

「電腦模型」之評估數據整理如表五，說明如下：『正確率』為1.00、『召回率』為1.00、

六個級別之『精確率』為1.00，以及『整體平均精確率』在mAP@0.50為1.00，整體辨識所需時間約為20秒。

在「人工目視」方面，共招募10位受試者，其中包含副供站站長2員、品檢員3員及國軍部隊專任伙委、採買作業人員各5員，受試者資料如表六所示。並以筆記型電腦撥放60張影像資料簡報，再請受試者依照圖片的外觀品質特徵進行辨識，逐題勾選相對應的級別分類。另外，區分『專業作業人員』與『一般作業人員』的驗證結果整理如表七及表八所示。

表四 電腦模型評估

評估數據							
Class	Accuracy	Recall	Precision	mAP@0.50	Times	F1	Kappa
香蕉特級品	0.99	1.00	0.99	1.00	55秒	0.99	0.93
香蕉優級品		1.00	0.99			0.99	
香蕉不良品		1.00	0.99			0.99	
番石榴特級品		1.00	0.99			0.99	
番石榴優級品		1.00	0.99			0.99	
番石榴不良品		1.00	0.99			0.99	
註：測試集177張							

資料來源：本研究整理。

表五 電腦模型驗證結果

評估數據							
Class	Accuracy	Recall	Precision	mAP@0.50	Times	F1	Kappa
香蕉特級品	1.00	1.00	1.00	1.00	20秒	1	1
香蕉優級品							
香蕉不良品							
番石榴特級品							
番石榴優級品							
番石榴不良品							
註：驗證集60張							

資料來源：本研究整理。

『專業作業人員』之評估數據說明如下：『精確率』為0.94，以及『整體平均精確率』在對於辨識兩種研究物件各六個等級的『正確率』為0.94、『召回率』平均為0.95、六個級別之均約為119秒。

表六 受試者人員說明表

項次	類別	職務內容	從事副供作業年資
1	專業作業人員	副供站站長	2年
2			1年10個月
3		副供站品檢員	6年
4			6年
5			6年
6	一般作業人員	伙委	3個月
7			2個月
8		採買人員	1.5個月
9			1個月
10			1個月

資料來源：本研究整理。

表七 人工目視驗證結果（專業作業人員）

評估數據							
Class	Accuracy	Recall	Precision	mAP@0.50	Times	F1	Kappa
香蕉特級品	0.94	0.93	1.00	0.94	119.2秒	0.96	0.93
香蕉優級品		0.96	0.92			0.94	
香蕉不良品		1.00	0.96			0.98	
番石榴特級品		0.91	1.00			0.95	
番石榴優級品		0.90	0.88			0.89	
番石榴不良品		0.98	0.90			0.94	
註：驗證集60張							

資料來源：本研究整理。

表八 人工目視驗證結果（一般作業人員）

評估數據							
Class	Accuracy	Recall	Precision	mAP@0.50	Times	F1	Kappa
香蕉特級品	0.81	0.82	0.94	0.81	146秒	0.88	0.77
香蕉優級品		0.67	0.78			0.72	
香蕉不良品		0.97	0.68			0.80	
番石榴特級品		0.75	0.90			0.82	
番石榴優級品		0.78	0.64			0.70	
番石榴不良品		0.94	0.92			0.93	
註：驗證集60張							

資料來源：本研究整理。

『一般作業人員』之評估數據說明如下：對於辨識兩種研究物件各六個等級的『正確率』為0.81、『召回率』平均為0.82、『精確率』為0.81、六個級別之『精確率』為0.81及『整體平均精確率』在mAP@0.50為0.81，每人整體辨識所需時間平均約為146秒。

肆、結論與建議

一、研究結論

果品分級不但是國內產銷過程中掌握品質的關鍵要素，也影響臺灣水果在國際市場上之評價與表現。臺灣農事環境正面臨作業人口高齡化的衝擊，而目前國軍副供站在國防部後勤人力縮減的政策推動下，每日耗時費力的副食檢驗作業已顯現品檢人員普遍不足情況，故建立適合副供事業的水果品質檢測模型有其必要性。本研究應用電腦視覺的影像處理與物體偵測技術輔助人工目視檢驗，以CNN的「YOLOv3」為基礎架構，挑選副供站供應之「香蕉」及「番石榴」為研究物件，發展品質檢測分類的電腦模型，並獲得正確率接近100%的辨識結果。

本研究以人工目視方式進行檢驗分類的「專業作業人員」，在相關實務工作已累積足夠的品質檢驗經驗，品檢準確率及效率相對於「一般作業人員」較高。然而，經演算訓練後的「電腦模型」，其分類表現更為精準，且所需的辨識時間相較與「專業作業人員」約有100秒以上的時間差，突顯電腦視覺的影像

辨識能力大幅領先人工目視，有相當高的參考與應用價值存在；而在與「一般作業人員」的比較下，「電腦模型」更有高達125秒的時間差優勢，可證明在國軍部隊之「非專業作業人員」伙委、採買等人力運用方面，能以電腦視覺有效輔助人工目視。

綜前所述，就研究結果而言，顯見本研究模型的辨識效能已達一定水準，成功建立快速、具高精確率的影像辨識模型，並提升傳統水果品質檢驗方法。同時，本研究所操作之電腦設備並無搭載高速運算影像之顯示卡，可有效模擬副供站實際配置的電腦作業系統，輔助副供站進行準確有效率之檢驗分類技術，建構標準一致的外部品質檢測分類模式，作為品質分類之依據，提供副供站準確且有效率之檢驗分類技術。

二、研究貢獻

每日實際副供作業中，品檢人員針對水果進行外部品質判定時，通常會將水果拿起並左右旋轉進行目視檢查，以綜合判斷此一物件之分類級別。然而，此一檢驗分類作業環節，也是人工品檢階段中，最容易因為人的精神因素影響，出現缺乏統一客觀的分類誤差。而在本研究之「電腦模型」與「人工目視」的評估驗證結果中可發現，電腦模型不僅有相當優異的辨識速度表現，可證明能有效降低非專業人員辨識的壓力，節省辨識的整體時間。更重要的，能大幅降低人為「有意」、「無意」的誤差產生。

本研究所建立的電腦分類模型係採取

「YOLOv3」演算法，運用CNN與機器學習的優點，確實地感知並擷取物件的特徵及其位置，進而推測其動態的變化。因此利用分類模型結合「視訊鏡頭」或「網路攝影機（webcam）」等光學設備，即時監控水果影像，加上多區域、多角度的辨識能力，配合動態的水果揀選分級流程，獲取一致性的外部品質判定，並可有效避免影像資料因為照相機拍攝的「單一角度」或「照明光源」其他外在條件的干擾因素，出現與「特級品」或「不良品」等級的相似角度，提高辨識結果正確率，降低分類誤差產生。

三、未來展望

為滿足國軍相關單位每日菜單搭配及伙食烹調需求，副供站提供之副食品高達上千種品項，本研究考量副供人員執行品檢時，在後勤人力逐日壓縮下，可輔助人工目視，故以供銷數量最高的兩種水果「香蕉」及「番石榴」為本研究範圍。若國軍未來發展導入物聯網（The Internet of Things, IoT），持續蒐集大量、有效的副供站產品影像，可能將產生龐大的圖片資料，建議後續依此為基礎，完整建構專屬副食品項之資料庫，增加參數獲得的擴充程度，進而提高模型預測精確率。除此之外，本研究於文獻初探階段，「YOLO」演算法僅發展到第三代版本，但隨著科技飛快進步，現已擴展至「YOLOv4」，其神經網路結構以及運算的隱藏層數相對更深入、複雜。考量後續研究的議題，建議未來運用「YOLOv4」模型演算法，以利巨量數據學習

及影像資料分析。

最後，在研究平臺操作系統上，大量副食品項的影像數據接收、辨識判讀結果的參數，須頻繁地溝通與回饋至轄管副供站單位之雲端服務器，中間感應器和服務器的距離可能相差數萬公里，就算資訊能以直線高速傳播，都可能有數毫秒的延遲，再加上來回計算時間，造成大量網路流量，雲端可能超出負荷。因此建議運用霧運算布局，因其接近終端用戶（各地區副供站），可有效減少延遲時間，不僅減低雲端系統負擔，又能快速回應短期大量的數據運算與資料分析的環境需求，使判斷結果標準一致性並增快運算速度，以利整體副食品檢的穩定及效率之需求。

作者簡介

林念德少校，聯合後勤學校專業軍官班97年班，陸軍後勤訓練中心正規班104年班，國防大學管院運籌管理學系碩士110年班，現任職於陸軍後勤指揮部補給處補給官。

作者簡介

溫志皓備役中校，空軍軍官學校88年班，國防大學管院資源管理研究所碩士94年班，國立中央大學企業管理研究所博士102年班，現任職於世新大學傳播管理學系助理教授。