

Analysis and Prediction for Defective Irwin Mangos Based on Neuron Networks for Image Recognition

Wei-Ting Wu¹, Chow-Sing Lin^{2,*}

^{1,2} Department of Electrical Engineering, National University of Tainan, Tainan, 70005,
Taiwan

^{2*}E-mail : mikelin@mail.nutn.edu.tw

Abstract

This paper used the image recognition algorithm model and Detectron2 framework to detect five types of defective Irwin mangos. The principal object was to recognize five different diseases concerning Irwin mangos by multi-object instance segmentation. There are five diseases respectively, poor coloration, anthrax, latex attached, mechanical harming and ink spot disease. We spent over 500 hours on data cleaning and data pre-processing due to the dataset (training set + validation set totally 59650 images) offered by vendor is inferior in quality. We also eliminated 8995 bad images. Finally, the dataset remained 50655 images. Data collection and data hazard are the significant challenges will face when apply AI to the real-world. Our research mainly used Mask R-CNN which is the image recognition algorithm of deep neural networks and transfer learning based on COCO Dataset Pretrained Model makes the detection results more precise. Then, we used Grabcut algorithm which makes accuracy of instance segmentation up to 99.9% in data pre-processing stage. Further, we applied X101-FPN backbone for making neural network deeper which compared to R50-FPN was effectively improved 90% accuracy. Eventually, we achieve the 67.2 AP in ours experiment.

Keywords: Image Recognition, Mask R-CNN, Detectron2, Transfer Learning, Diseases Detection.

* Corresponding author: mikelin@mail.nutn.edu.tw
DOI : 10.53106/222344892022101202007

影像辨識神經網路應用於愛文芒果不良品分析與預測

吳威廷, 林朝興

國立臺南大學資訊工程系

摘要

本研究採用 Mask R-CNN 影像辨識演算法模型以及 Detectron2 框架，應用於愛文芒果影像進行五類不良品分類。主旨在於針對愛文芒果五類不同病種進行多物件實例分割影像辨識，五類病種分別為著色不良、炭疽病、乳汁吸附、機械傷害以及黑斑病，總共五類不良病種。因廠商提供之資料集(training set + validation set 共計 59650 張)資料雜訊嚴重，在資料清洗(Data Cleaning)以及資料前處理(Data Preprocessing)階段花費超過 500 小時以上，以及將不良資料汰除 8995 張，最終資料集共為 50655 張。資料集蒐集不易以及資料雜訊多的情形，為人工智慧應用於真實場域所會面臨的挑戰。為求辨識精準度而非速度，此研究主要應用技術為深度神經網路影像辨識演算法 Mask R-CNN 以及基於 COCO Dataset Pretrained Model 應用遷移式學習(Transfer Learning)。再運用 GrabCut 演算法，在資料前處理階段使實例分割(Instance Segmentation)達到 99.9% 的精準度。以及應用 X101-FPN 骨幹，以增加神經網路深度，相較 R50-PFN，使辨識精準有效提升 90%，最終達成 67.2 之 AP。

關鍵詞：影像辨識、Mask R-CNN、Detectron2、遷移式學習、病種辨識

1. 緒論

隨著近年來人工智慧(Artificial Intelligence)[1]研究蓬勃發展，相關應用以及技術實踐也呈現爆發式成長。由人工智慧所進行延伸的重要技術，機器學習(Machine Learning)[2]，更是使人工智慧此詞彙推向高峰的重要轉捩點。機器學習透過處理運算龐大數據量，並且讓機器學習演算法對數據進行歸納分析，使模型能夠自行對新數據進行解讀。其中技術應用像是，監督式學習(Supervised Learning)[3]、半監督式學習(Semi-supervised Learning)[4]以及非監督式學習(Unsupervised Learning)[5]均是目前炙手可熱的機器學習應用，也未人工智慧後續發展奠定了基礎。

而目前，基於機器學習技術發展，最為熱門且應用最為廣泛的技術絕對非深度學習(Deep Learning)[6]莫屬，深度學習因近年來硬體技術的成長與發展，促成了第三次人工智慧的突破與浪潮。深度學習與機器學習的不同在於，機器學習是交由人類來定義特徵，再由機器運用人類所定義的特徵結合大量數據，對模型目標進行歸納，最終做出預測以及決策。

而深度學習則是運用類神經網路(Neural Networks)[7]，交由機器自行定義特徵向量，利用相似於人腦神經的結構，配合神經網路深度。相較於機器學習能夠更加精進機器的訓練難度以及計算能力，使其能夠學習處理更加複雜且龐大的大數據，像是影像辨識分類、語音解析、文字語意分析等。

而在深度學習中，其中最建樹的一支應用便是影像辨識技術，而在人工智慧影像辨識領域中最廣泛應用且最熱門的技術便是卷積神經網路(Convolutional Neural Networks, CNN)[8]，而 CNN 模型的雛型最早是由卷積神經之父，Yann LeCun 所提出[9]，應用於阿拉伯數字的手寫字體物件辨識(Object Recognition)中，再加上硬體技術崛起後，自此便掀起了一波 CNN 研究與應用風潮。

CNN 使用卷積核可以做到有效的特徵提取，使影像與圖片中各個局部的特徵都能夠被 CNN 模型所擷取到，再加上因為 CNN 具有局部權重共享的特性，使整張圖片不會只被偏重辨識單一特徵，能夠較全面的被模型所辨識。

也因為上述 CNN 所具備的特性，CNN 相較於其他的深部學習模型，在影像辨識、語音辨識以及圖像辨識與分析上，具備一定的領先與優越性。CNN 也被認為是第一個真正成功且能夠應用於真實場域的深度學習技術，故此研究中也是由 CNN 技術做為物件辨識基底技術。

除前述的人工智慧之外，智慧加值的概念也開始導入各項產業之中，農業精準化技術正是其中一項。芒果是臺灣三大外銷高經濟生鮮果品之一，但到至今仍未建立一套有效且低成本的不良品偵測系統，故需要在採收處理後結合科技輔助應用，以此實現自動化以及精準化的智慧農業應用。此研究預期最終能夠結合臺灣本地資料庫建立自動芒果篩果系統，並與農委會合作結合區塊鏈技術整合搭建芒果智能履歷，提供果農與民眾芒果選購系統。

如圖 1 所示，愛文芒果不良品包含多樣種類，此研究主要目的為針對愛文芒果，其中五類不良品進行影像辨識分類，包括機械傷害、黑斑病、炭疽病、乳汁吸附、著色不佳等五種病徵。運用 Mask R-CNN[10]技術，快速且正確地分析與預測愛文芒果不良品之病徵。訓練出精準度高且有效的模型，將能夠真正的幫助傳統農產業導入人工智慧自動化篩果技術，促成傳統農業轉型，以節省現行的果篩人力成本。



圖 1: 愛文芒果果篩分類準則圖[18]

2. 相關研究

2.1 Mask R-CNN 網路架構

此論文研究為愛文芒果五類不良品分類，以此為基礎再選擇合適的影像辨識演算法模型。由於在實際場域中，芒果在輸送帶上並非高速移動物體，需求攝影鏡頭幀數僅需 30FPS，故我們首要的需求條件為高準確度，其次才是辨識速度。我們比較了多項影像辨識相關技術，參照圖 2 及表 1 可以發現，在基於 COCO Dataset[19]作為資料集所訓練出來的所有演算法模型之中，YOLO[11]的辨識速度最快，但 AP 僅為 21.6。而 Mask R-CNN 在 ResNeXt-101-FPN 之 backbone 訓練下，AP 可達 37.1。

綜合以上因素，我們挑選了辨識速度稍慢，但幀數可達 30FPS，而精準度提高了 71% 的 Mask R-CNN 作為技術基礎。

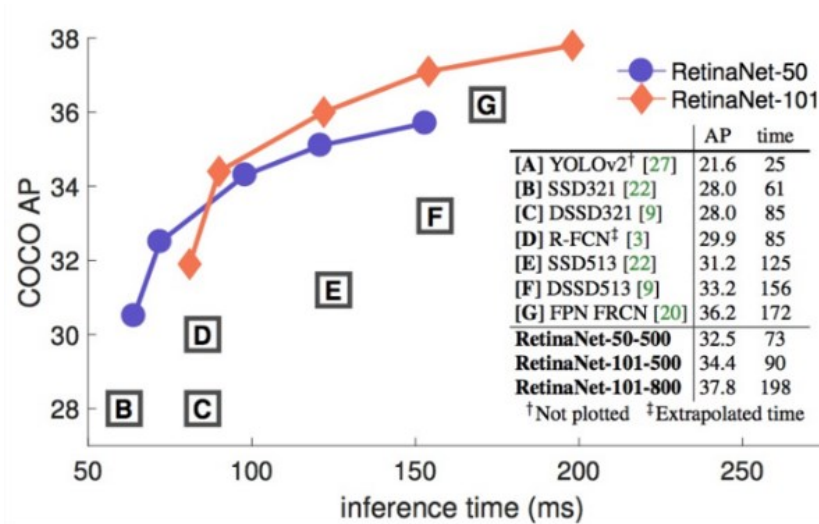


圖 2：各類影像辨識技術準確率[11]

表 1：Mask R-CNN 準確率[10]

	backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
MNC	ResNet-101-C4	24.6	44.3	24.8	4.7	25.9	43.6
FCIS + OHEM	ResNet-101-C5-dilated	29.2	49.5	-	7.1	31.3	50.0
FCIS+++ + OHEM	ResNet-101-C5-dilated	33.6	54.5	-	-	-	-
Mask R-CNN	ResNet-101-C4	33.1	54.9	34.8	12.1	35.6	51.1
Mask R-CNN	ResNet-101-FPN	35.7	58.0	37.8	15.5	38.1	52.4
Mask R-CNN	ResNeXt-101-FPN	37.1	60.0	39.4	16.9	39.9	53.5

除此之外，我們選用 Mask R-CNN 還有一大考量因素，就是運用 Mask R-CNN 可以做到實例分割(Instance Segmentation)的特性，如圖 3 所示，實例分割可將前景與後景進行分割，不僅能辨識出目標物件特徵也能達到分割效果。因此我們運用實例分割技巧進行資料前處理，精準辨識芒果之輪廓特徵，留下我們想要辨識的目標物件，生成無背景之芒果圖像，最大程度降低訓練資料雜訊。而實例分割與語意分割(Semantic Segmentation)僅有部分神經網路能實現，要做到實例分割絕大多數需基於 FCN(Fully Convolutional Neural Networks)[12]的架構下，而 Mask R-CNN 是目前基於 COCO 資料集下實例分割的最佳作法。

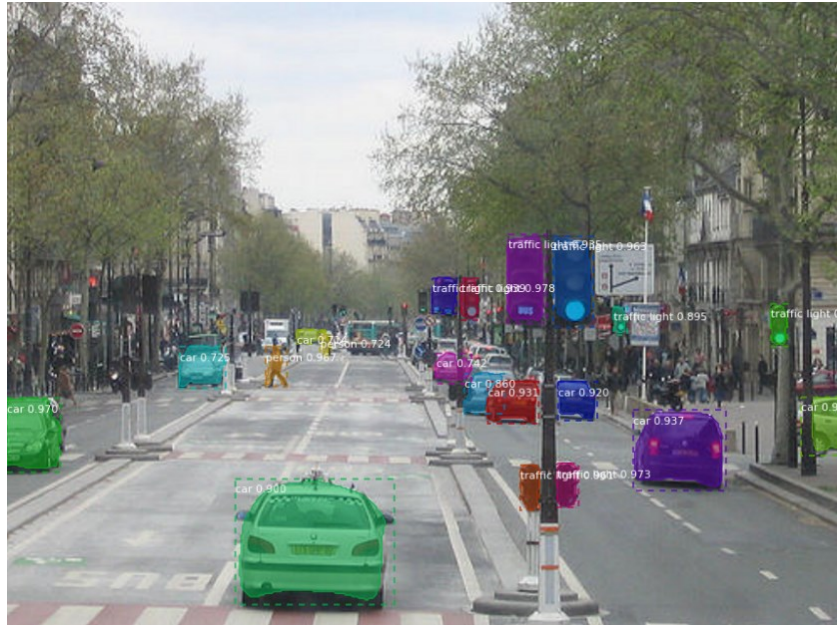


圖 3：Mask R-CNN segmentation[20]

Mask R-CNN 是由 Faster R-CNN[13]演變而成的，Mask R-CNN 的架構如圖 4 所示，Mask R-CNN 的基礎架構採用與 Faster R-CNN 相同的 two-stage procedure。透過第一階段，也稱作 Region Proposal Network(RPN)來候選出 object bounding boxes。在第二階段 Mask R-CNN 導入了 RoI Align(Region of Interest Align)，以此取代了 Faster R-CNN 中的 RoI Pooling，對 sampled RoI 定義的 multi-task loss 如(1)式：

$$L = L_{cls} + L_{box} + L_{mask} \quad (1)$$

RoI Align 運用插值法確保影像分割時輸入座標與輸出座標一致，改善 Faster R-CNN 輸入與輸出像素不一致的問題，讓影像分割的準確度得到大幅度的改善，並且超過了當時所有最先進的影像辨識演算法模型。因 Mask R-CNN 上述的特性，特別適合應用於本研究中，故使用 Mask R-CNN 於芒果輪廓分割，進行資料前處理的資料清洗與輪廓辨識。

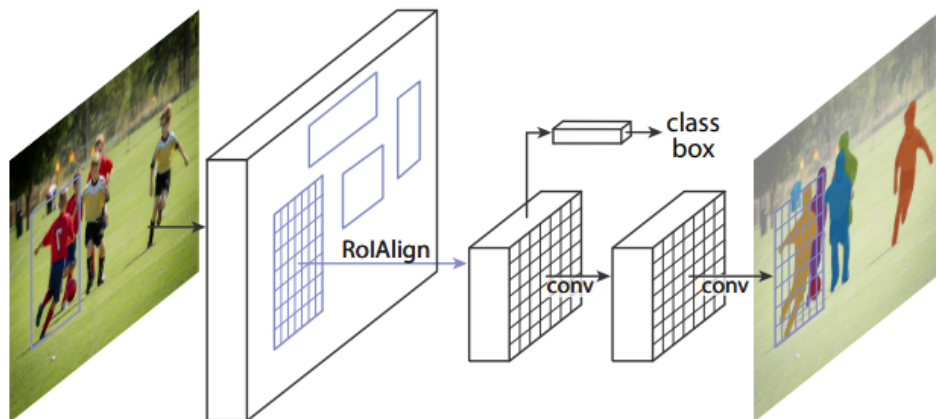


圖 4：Mask R-CNN 實例分割框架[10]

2.2 遷移式學習

由於我們進行訓練的技術基底 Mask R-CNN 與後續使用的 Detectron2 框架[21]皆是基於 COCO dataset[14]。在此基礎上，我們可以在 COCO dataset pretrained model 上應用

遷移式學習(Transfer Learning)[15]，以此技術增強我們訓練模型的精準度。遷移式學習的主要目的是，解決資料取得不易的問題，這一直是人工智慧模型所會面臨到的問題。而遷移式學習是將某個領域或是模型的訓練結果遷移至另一個模型之中，藉此舉來提高訓練效果，此應用角度就像是某個先備模型的知識被遷移應用至另一個目標模型上。

從圖 5 可看出，遷移式學習可應用於不同類型之機器學習方法上，而對於不同方法會對應不同的 learning setting。在遷移是學習中，會定義 domain 與 task。domain(D)是由兩個部分組成:特徵空間(feature space, χ)以及邊緣分佈(marginal probability distribution, $P(X)$)， $X = \{x_1, \dots, x_n\} \in \chi$ 。而給定一個 domain 為 $D = \{\chi, P(X)\}$ ，task 包含兩個部分，標記空間(label space, Y)與目標預測函數(objective predictive function, $f(\cdot)$)，task 表示為 $T = \{Y, f(\cdot)\}$ ，可以從 training data 中習得。在最常見的情況下，我們考慮一個 source domain D_S 與 target domain D_T ，source domain 的 data 為 $D_S = \{(x_{S_1}, y_{S_1}), \dots, (x_{S_{n_S}}, y_{S_{n_S}})\}$ ，滿足 $x_{S_i} \in \chi_S$ 為 data instance，且 $y_{S_i} \in Y_S$ 為相對應的 class label。因此可表示 target domain data 為 $D_T = \{(x_{T_1}, y_{T_1}), \dots, (x_{T_{n_T}}, y_{T_{n_T}})\}$ ，滿足輸入為 $x_{T_i} \in \chi_T$ ，而 $y_{T_i} \in Y_T$ 為相對應的輸出。基於以上定義，給定 source domain 為 D_S ，learning task 為 T_S ；而 target domain 為 D_T ，learning task 為 T_T ，當滿足 $D_S \neq D_T$ 與 $T_S \neq T_T$ ，遷移式學習可運用 D_S 與 T_S 中的知識來優化 D_T 中的 $f(\cdot)$ ，以此增進學習效果。

Learning Settings		Source and Target Domains	Source and Target Tasks
Traditional Machine Learning		the same	the same
Transfer Learning	<i>Inductive Transfer Learning /</i>	the same	different but related
	<i>Unsupervised Transfer Learning</i>	different but related	different but related
	<i>Transductive Transfer Learning</i>	different but related	the same

圖 5：Relationship between Traditional Machine Learning and Various Transfer Learning Settings[15]

遷移式學習也會有運用不佳的時候，稱作負遷移(Negative Transfer)，此時會導致訓練效果不佳。由圖 5 可知，選用正確地遷移方式相當重要。而我們使用的遷移式學習屬於推導遷移式學習(Inductive Transfer Learning)，適用於具有足夠數量標記訓練資料樣本，並能以樣本建立模型，以取得較佳的效果，正好適用於本研究之應用。

2.3 Detectron2 框架

在進行病種辨識的階段，我們不只是單純使用 Mask R-CNN，而是進階使用由 FAIR 實驗室開發的 Detectron2 框架。Detectron2 是基於上一代 Detectron[22]與 maskrcnn-benchmark[23]開發的，主要應用於電腦視覺領域研究，而其中支援的一項技術，正是 Mask R-CNN。因為 Mask R-CNN 的作者在此團隊中，故 R-CNN 系列一直是 Detectron2 的重點支援項目，能夠進行語意分割、實例分割，以及全景分割。

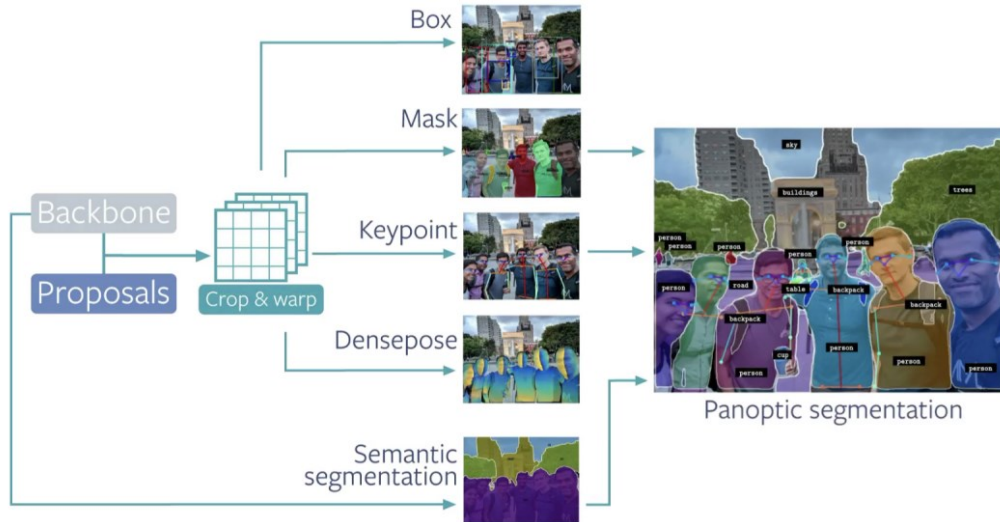


圖 6：Detectron2：Generalized R-CNN Models[21]

Detectron2 另一項優點是速度相較於原生演算法模型或是與其他框架進行對比，有明顯的優化。Detectron2 對於硬體資源也有更好的調用性，可支援多 GPU 運算，以及算力的分配調控等功能。故選用 Detectron2 作為第二階段病種訓練框架。而 Detectron2 不僅對於硬體資源，對於許多訓練參數都具備足夠高的自定義彈性。在 model 的部分，Detectron2 將其劃分為多項模組，從特徵提取器、生成器、骨幹，到最終集成模型都能自由調適，也可自定義模組，進行不同架構的擴充應用。

3. 愛文芒果病種辨識流程

3.1 辨識流程

研究流程如圖 7，主要分為兩部分：資料前處理(橘色部分)及病種辨識(綠色部分)。此外，兩個部分也分別對應到原生 Mask R-CNN 與 Detectron2，而基底技術皆為 Mask R-CNN。因原生 Mask R-CNN 只能使用(ResNet 50 & 101)骨幹且對硬體調不佳，故第二階段之病種辨識採用 Detectron2。資料前處理階段目的，為去除雜訊、輪廓辨識以及背景去除，而第二階段為最終愛文芒果不良品病徵辨識。

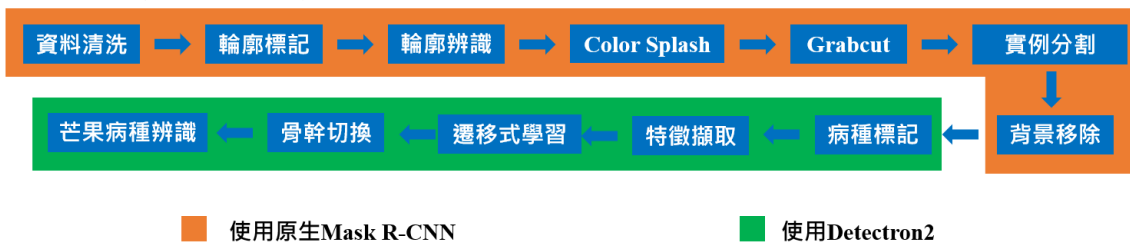


圖 7：研究流程圖

3.2 資料前處理

3.2.1 資料清洗

資料集以及驗證資料集(training set & development set)，總共為 59650 張愛文芒果各類病種資料圖片，其中又以"炭疽病"和"著色不佳"占了絕大部分的比例。在資料前處理中，第一個步驟為資料清洗，而資料清洗則是將原始資料中，如圖 8 所呈現之資料，會

嚴重影響辨識準確率的不良圖片給汰除(芒果上有貼紙、影像過於模糊、非單一目標物.....等)。由於我們要先做輪廓切割再進行病種辨識，因此會直接影響到辨識目標物或是妨礙到遮罩生成的圖片都必須汰除。如將其強行保留則會成為資料雜訊，嚴重可能會導致辨識結果變差。因為廠商提供之資料集品質良莠不齊，以不產生雜訊為前提，最大限度保留圖片資料以利後續訓練的情況下。原始的 59650 張圖片經過了人工資料清洗後，剩下 50655 張圖片，共計汰除了 8995 張照片。



圖 8：汰除資料樣式圖片

3.2.2 輪廓辨識

Mask R-CNN 為了生成 mask，在前置需要進行資料標記的工作，因此第二步驟為輪廓標記。對於位置標記，廠商有提供圖片資料標籤註記檔，但是這些標記檔將愛文芒果直接以矩形進行框選。對一般人工智慧影像辨識演算法來說，此資料標記即可進行訓練。但本研究為求準確度，並且期望能實際應用到場域之中，使用此類資料標記會使準確度有所耗損。因為愛文芒果為不規則形的辨識物體，若使用矩形進行標記框選，則會將愛文芒果以外的背景一併納入辨識物體的範疇之中。此舉無異於替訓練資料增加雜訊，屬於混淆影像辨識模型的作法，勢必會造成精準度下降的問題。因此，我們採用 VIA(VGG Image Annotator)作為資料標記工具，以 1000 張圖片作為訓練資料集，驗證資料集和測試資料集則各為 200 張圖片，採用不規則邊框行進行標記。實際輪廓標記情形如圖 9 所示，須以多點方式進行不規則形狀標記，將芒果輪廓外觀以不規則狀框選標記。在標記完成後可生成 JSON 文件作為資料格式，輸入至原生 Mask R-CNN 中，以 ResNet 50 & 101 做為 Backbone 進行模型訓練。完成訓練得到權重後，可進行初步推論得到芒果的輪廓辨識結果，因為輪廓辨識相當成熟，因此可由圖 10 看出芒果辨識的結果準確度極高，資料集中所有芒果輪廓辨識，精準度均在 99% 以上，並且均能正確的生成出覆蓋芒果的 mask。

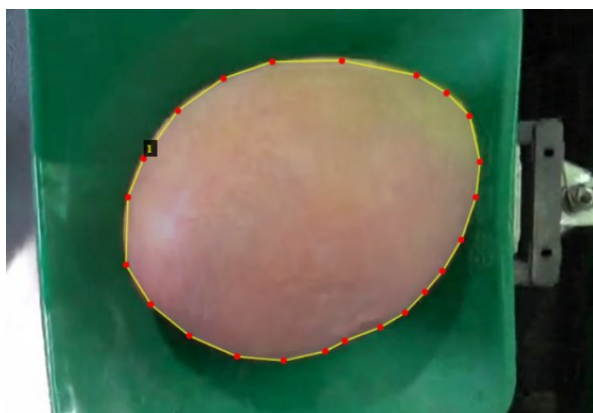


圖 9：不規則形狀輪廓標記



圖 10：愛文芒果輪廓辨識結果

輪廓辨識的主要目的是為了生成 mask 以利後續對芒果進行實例分割(Instance segmentation)，而做實例分割是為了確保在辨識病種時能夠將雜訊最小化。而生成 mask 後就可以進入到 color splash 階段，使用 color splash 進行 mask 的上色，能夠確認 mask 是否覆蓋完善。在使用 color splash 後會發現雖然 mask 基本上已經很完善的覆蓋在芒果上，但是仍有一小部份邊角尚未將芒果涵蓋。因此，本研究運用 Opencv 函式庫中的 Grabcut 演算法[16]對 mask 生成進行補強，使 mask 可以更加完善的將芒果輪廓完整包覆。由圖 11 可看出，由 color splash 到 Grabcut 將 mask 補全地更加完整。由此可知，Mask R-CNN 結合 Grabcut 演算法後使輪廓辨識更加完善。如此一來，現在我們就有準確度達 99%且貼合的芒果輪廓 mask。最後使用原生 Mask R-CNN 中的實例分割，將芒果做為前景物件與後景(背景)分離，將芒果完整的切割出來，完全去除圖片背景以利後續病種辨識的訓練。至於我們堅持將背景去除的原因是，因為病種辨識才是我們的主要目標而非輪廓辨識，故希望盡可能將雜訊降至最低，因此選擇將背景直接完全切除(檔案形式為 png)，而非單純使用灰階或二值化技巧進行處理。至此，資料前處理已處理完畢。

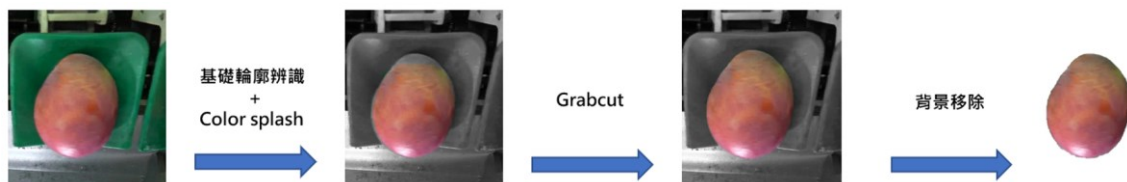


圖 11：輪廓辨識與去雜訊流程

3.3 病種辨識

病種標記同樣是先利用 VIA 進行不規則邊框標記，根據廠商提供的病種 label，如圖 12 所示，我們在一張張圖片進行不規則形狀之病種位置標記，同樣花費了相當巨量的時間在處理病種標記上。



圖 12：愛文芒果病種標記

3.3.1 正規化與特徵擷取

由於廠商提供的圖片大小不一，因此須將圖片進行圖片大小進行正規化。在圖 13 中，我們以 mask 之中心點向外擴展成統一大小，假若擴展範圍超出原圖片的範圍則做 padding 補上空缺。如此一來即可確保資料輸入進訓練模型時的圖片大小是一致的。

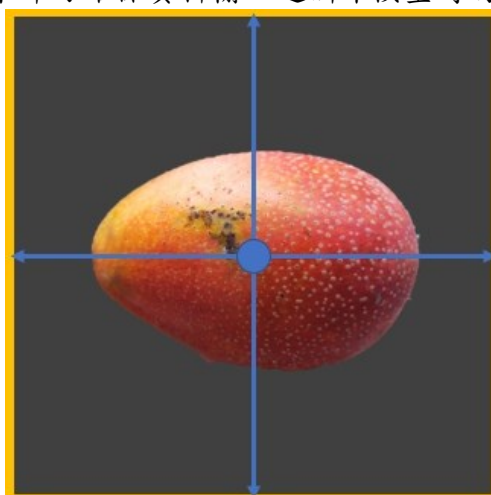


圖 13：圖片資料正規化示例圖

完成圖片大小之正規化後，為確保資料特徵多樣性，我們進行了特徵擷取。特徵擷是除了去除背景後的原色圖外，嘗試使用不同的特徵製成的資料去進行模型的訓練(灰階、二值化、HSV 調適等)。在本研究中我們採用了灰階以及二值化，由圖 14 可看出，二值化讓芒果病癥點過於淡化，效果不甚理想。而灰階則將病徵特徵特別強化。以訓練結果來看，其中又以灰階特徵的訓練效果最佳，故後續訓練輸入資料均轉化為灰階進行訓練。

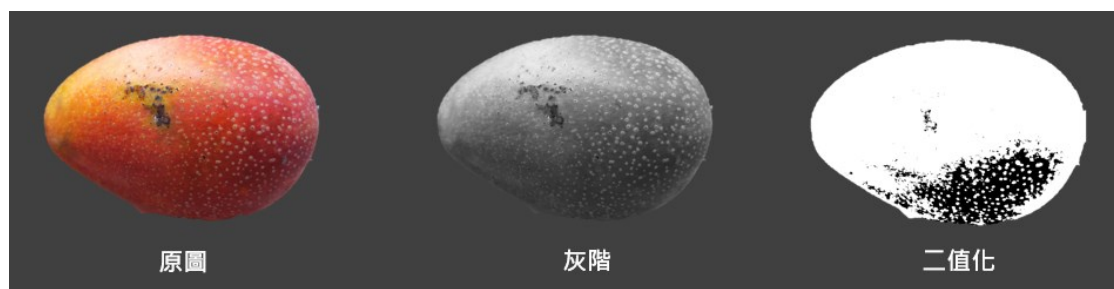


圖 14：愛文芒果特徵擷取

3.3.2 訓練階段

在資料準備完成後，要從原生 Mask R-CNN 轉向使用由 Facebook 的 FAIR 實驗室開發的 Detectron2。主要有三項原因，第一是因為 Detectron2 對硬體比原生 Mask R-CNN 有更高的可調用性，因本研究高階運算資源，需進行多 GPU 同步運算，故對硬體的調用性有所要求；第二，Detectron2 具有更多的骨幹選擇，如果想要的話，甚至可以自行建置骨幹架構。不同的骨幹對於不同的訓練需求以及不同的資料形式都會產生不同的訓練結果，更多的骨幹選擇，就能夠有更多不同的嘗試機會。如果使用原生 Mask R-CNN，僅能使用(ResNet 50 & 101)骨幹；第三，Detectron2 對於神經網路的參數能夠有更好的調控，除自架骨幹外其實也能夠針對整個模型本身進行參數操作，相當靈活，故而選用 Detectron2 做為病種辨識的框架。而資料備齊後，使用 Detectron2 的第一步需要將資料格式由 JSON 轉為 COCO data format，因為 Detectron2 只能夠接受 COCO data format 之資料格式。因此我們需特別針對"file name","image_id","height","width"四項參數，由一般 JSON 格式轉換為 COCO data format。

第二步是引用 Detectron2 dataset APIs DatasetCatalog&MetadataCatalog，先以 DatasetCatalog 對自建 dataset 做 register，用 MetadataCatalog 將 dataset 中的 metadata 讀出，再建立字典(dict)以供後續訓練模型使用。再針對自己所需的 config 參數做調節，我們以自建骨幹以及硬體調節進行嘗試，最高能一次使用八片 GPU 進行 training。

最終，我們必須針對先前 resize 的圖片大小以及 annotation 的特性來自建模型所需之 Dataloader，順利將前處理完成的資料輸入進模型之中，並且緊接著開始進行，病種特徵辨識模型訓練。

3.3.3 遷移式學習與骨幹

相較於從頭訓練單純辨識病種的模型，我們可以利用 Detectron2 提供之，基於 COCO Data-set Pretrained Model 的遷移式學習(Transfer Learning)進行模型訓練。雖然 COCO Dataset 原生並不包含芒果病種，但引入訓練過的權重文件確實會使辨識結果更加精準。遷移式學習之所以不包含訓練資料卻還能起到修正偏差的原因是，若基於某一訓練資料集的情況下，假設資料集訓練有效，可以將先前訓練的權重做為借鑒。等同在已經看過相當多不同類型資料與特徵的情況下，將我們自建的資料補入現有模型中，因此在一般情況下使用遷移式學習將有助於訓練模型。此外，由表 2 可知，Detectron2 提供了許多 backbone 選擇，不同骨幹對於訓練可因深度以及參數之差異，產生訓練速度以及精準度上的不同。而在本研究中，我們選用兩種最經典也最廣為使用之骨幹(R50-FPN & X101-FPN)以及自建骨幹進行模型訓練。我們將 R50-FPN 作為中型模型架構之骨幹，以及 X101-FPN 最為最消耗資源但 AP 最高的大型模型架構之骨幹選擇。而在本研究中，以 X101-FPN 深度最深，效果最佳，在 AP_1 達到 83.42 之精準度。

表 2：COCO Instance Segmentation Baselines with Mask R-CNN[21]

Name	lr sched	train time (s/tier)	inference time (s/im)	train mem (GB)	box AP	mask AP
R50-C4	1x	0.584	0.110	5.2	36.8	32.2
R50-DC5	1x	0.471	0.076	6.5	38.3	34.2
R50-FPN	1x	0.261	0.043	3.4	38.6	35.2
R50-C4	3x	0.575	0.111	5.2	39.8	34.4
R50-DC5	3x	0.470	0.076	6.5	40.0	35.9
R50-FPN	3x	0.261	0.043	3.4	41.0	37.2
R101-C4	3x	0.652	0.145	6.3	42.6	36.7
R101-DC5	3x	0.545	0.092	7.6	41.9	37.3
R101-FPN	3x	0.340	0.056	4.6	42.9	38.6
X101-FPN	3x	0.690	0.103	7.2	44.3	39.5

4. 實驗結果

4.1 資料集

由圖 15 可知，本研究資料集總共有五種病種樣式。如圖所示，分別是，乳汁吸附(D1)、機械傷害(D2)、炭疽病(D3)、著色不佳(D4)以及黑斑病(D5)共五種特徵病種。



圖 15：愛文芒果資料集病種類別圖

而資料集總數量與類別分布，如表 3 所示與圖 16 所展示。廠商提供之訓練資料集以及驗證資料集(training set & development set)，總計有 49650 張愛文芒果各類病種資料圖片，其中又以"炭疽病"和"著色不佳"占了絕大部分的比例。而測試資料集則因為廠商未提供而無法取得，故無測試資料集。

表 3：愛文芒果資料集病種分布表

Grand ML	Train	Dev	Test	Total
乳汁吸附(D1)	2579	391	Blinded	Blinded
機械傷害(D2)	507	90	Blinded	Blinded
炭疽病(D3)	23587	3504	Blinded	Blinded
著色不佳(D4)	15045	1997	Blinded	Blinded
黑斑病(D5)	2657	298	Blinded	Blinded

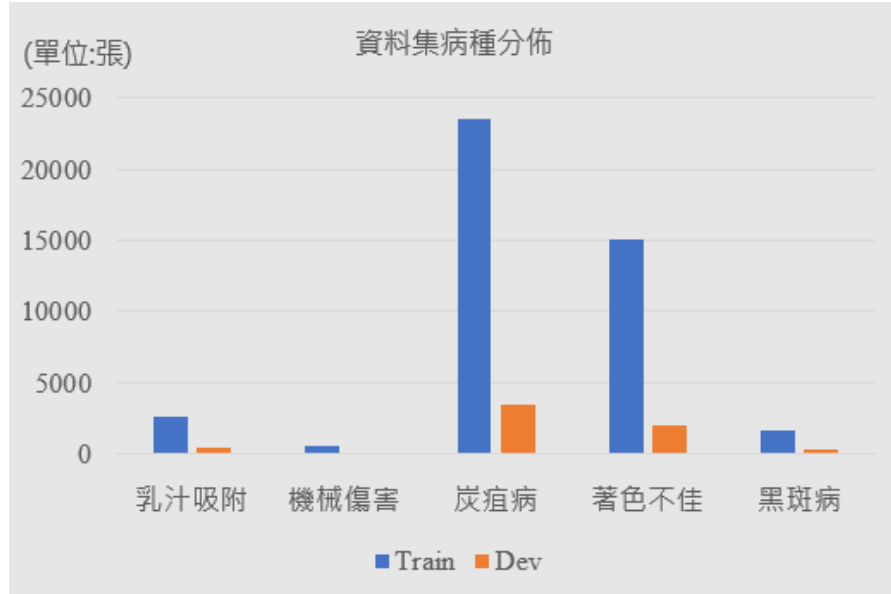


圖 16：愛文芒果資料集病種分佈圖

4.2 基準線與評估標準

根據廠商提供的資料，經由(2)式與(3)式之計算方式與對照表 4 之參數可綜合計算，此資料集的精度基準線。此研究屬於二分類問題中的多標籤分類問題(Multilabel classification)，因此在(2)式會涵蓋透過真實類別與分類器預測所劃分之TP, FP, TN, FN等四項類別參數。

$$\text{Macro-F1}(F_1, ma): F_1, ma = 2 * \frac{\text{recall}_{ma} * \text{precision}_{ma}}{\text{recall}_{ma} + \text{precision}_{ma}}$$

$$\text{where, } \text{precision}_{ma} = \frac{\sum_{i=1}^5 \text{precision}_i}{5}, \text{recall}_{ma} = \frac{\sum_{i=1}^5 \text{recall}_i}{5}$$

$$\text{precision}_i = \frac{TP_i}{TP_i + FP_i}, \text{recall}_i = \frac{TP_i}{TP_i + FN_i} \quad (2)$$

$i = 1, 2, 3, 4, 5$

$TP_i = \text{True Positive of class } i$, $FP_i = \text{False Positive of class } i$

$TN_i = \text{True Negative of class } i$, $FN_i = \text{False Negative of class } i$

而透過類別參數可以建構混淆矩陣(Confusion Matrix)，以表達分類之分析狀況。可透過四項參數分別計算出，精度量測指標(F-score)計算所需的精確率(precision)與召回率(recall)。精確率與召回率之間具有 trade-off，當著重FP時應更看重 precision;反之，著重FN時，應更看重 recall。在此研究中期望精準辨識芒果病種，因此更需注重精確率。本研究之愛文芒果病種辨識共計有五種病種，以此綜合計算總精確率與召回率，如(3)式所示：

$$\text{precision}_{ma} = \frac{\sum_{i=1}^5 \text{precision}_i}{5} = 62.67\%$$

$$\text{recall}_{ma} = \frac{\sum_{i=1}^5 \text{recall}_i}{5} = 62.21\% \quad (3)$$

$$F_1, ma = 2 * \frac{\text{recall}_{ma} * \text{precision}_{ma}}{\text{recall}_{ma} + \text{precision}_{ma}} = 63.92\%$$

分別為，62.67%與 62.21%。再將精確率與召回率用以計算，調和精確率與召回率的分類指標 $F_1 - score$ ，最終可得 63.92%。

表 4：病種 baseline 計算參數

Diseases Parameters	D1 乳汁吸附	D2 機械傷害	D3 炭疽病	D4 著色不佳	D5 黑斑病
Recall	56.11	55.61	77.19	76.79	60.32
Precision	54.1	50.38	77.19	76.75	54.93
F_1 -score	54.8	40.43	77.19	76.63	56

4.3 實驗環境

在影像辨識領域，眾所皆知的是，如果要將辨識結果或準確度往上提升有最直接的兩種方式。一是增加有效資料數量，二是增強運算資源，將 batch size 拉大，使神經網路一次能看到更多圖片。因此能使用高階運算資源，對於我們做影像辨識無疑是有很大幫助的。在本研究我們使用國研院國家高速網路中心 GPU 運算架構作為運算資源，最高使用到八張 Tesla v100 GPU 進行模型訓練，運用高達 960GB 之 VRAM 於神經網路模型之訓練。

最終，我們所選用的運算資源，國網型號為 c.4xsuper 的主機進行運算，CPU 為 8 張 Tesla v100，32 CPU cores，記憶體為 720GB。

4.4 辨識結果與消融實驗

本研究最終評估辨識結果之指標採用平均精準度(Average Precision, AP)[17]或稱作 mean average precision，為影像辨識模型中常見之評量指標。

參見(4)式可知：

$$AP = \int_0^1 p(r) dr \quad (4)$$

AP 是由前述之精確率以及召回率所形成曲線之積分，也就是面積。因精確率與召回率介於 0~1 之間，故 AP 同樣為 0~1 之間的數值，為影像辨識模型提供了平衡精確率與召回率的評估指標。由圖 17 可看出在最終結果中，芒果病種皆有準確辨識出來，並且方框中的範圍也有正確涵蓋芒果的病種。在所有病種中，以炭疽病、機械傷害以及著色不佳之辨識準確度較高，AP 平均值可超過 70。我們經過幾個階段的實驗測試與技術應用，使準確度漸入佳境，從訓練時間把控、資料蒸餾、增加網路深度、應用遷移式學習等。而最終試驗所有骨幹中效果最好的骨幹為 X101-FPN，由表 5 可知，從初始 AP 為 26.78 至最終實驗結果可達 67.42AP。我們取出辨識結果如圖 17 所示，以絕大部分的病種而言，已經可以達到超過 70%以上之病種辨識度。但像是在乳汁吸附與黑斑病等資料樣本較小以及分界較不明確的病種在辨識表現上相對來說沒有著色不佳等大量樣本數的病種表現來的好。考量到 Mask R-CNN 需要進行有效資料標記，準確度會與最初標記時的先備知識(domain knowledge)有關，資料集初始乾淨程度，對於最終辨識環節會對準確度造成一定的影響性。但總體來說，在常見病種上，病種辨識度仍然能夠達到七成。

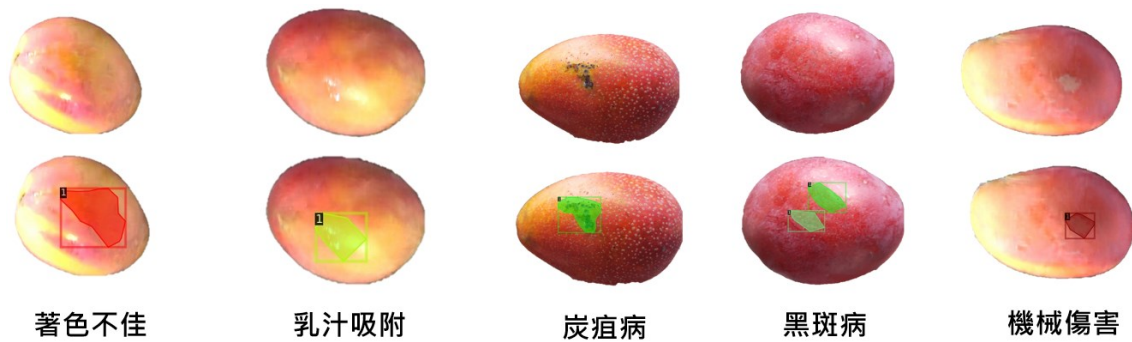


圖 17：病種辨識結果圖

表 5：消融實驗 AP 辨識結果

description	backbone	AP	AP ₅₀	AP ₇₅	APs	AP _m	AP _l
original baseline	R-50-FPN	26.78	55.64	51.88	4.53	32.72	61.48
+ train-time augm.	R-50-FPN	28.90	57.47	52.06	4.86	34.67	62.28
+ data distillation	R-50-FPN	30.74	59.41	56.97	5.51	36.04	64.91
+ deeper	X-101-FPN	58.49	70.03	69.90	9.74	52.57	80.71
+ data augmentation	X-101-FPN	61.35	73.15	71.54	9.84	54.45	81.34
+ transfer learning	X-101-FPN	63.64	74.41	73.98	10.21	55.74	81.62
+ test-time augm.	X-101-FPN	67.42	79.53	75.21	10.95	57.39	83.42

5. 結論與未來展望

本研究之主要貢獻為，替農產品果篩與病種辨識提供自動化解決方案。於現有解決方案中，提供不同於傳統影像辨識演算法之解決方案，運用 Mask R-CNN 與 transfer learning 等機器學習技術，為果篩與病種辨識建構一套自動化流程。在 X-101-FPN 骨幹架構中，最終本研究達到 67.42AP 之病種辨識率，較單純使用 R-50-FPN 有效提升近 91% 之 AP。而在資料樣本數較大之病種，辨識精準度也順利提高至七成，通過辨識基準線。唯待資料集先備知識完備之情況以及擴大資料樣本數，始可將辨識精準度有效往上提升。

我們期盼在未來能夠把辨識的準確率提升到更高，在更深度使用其他技術進行擴充，希望進一步提升辨識準確度。例如在擁有更多資料樣本數的情況下，可以應用 GAN 進行資料擴充。也期盼能夠將人工智慧影像辨識技術實際應用於農產品場域之中，並且完成數據監測使用介面，實際與產線結合達成即時監測與辨識，協助果農以更加智慧及自動化之方式進行果篩。取代現行耗時之人工果篩方式，以進一步協助傳統農產業，促進傳統農產業智慧化轉型。

6. 誌謝

本研究能順利執行，首要需感謝國立臺南大學林朝興教授擔任指導教授，感謝教授在研究過程中不斷提供建議及技術指導，循序漸進以修正研究過程中的問題，並以研發過程嚴謹為第一要務。再感謝國研院國網中心通過我們提交之提案並農產品 AI 影像辨識成立計畫案，提供計算以及儲存資源，協助本研究順利進行。最後感謝廠商提供此議題以及資料，能實際解決真實議題的機會實屬可貴，並讓我們於此次研究中深刻瞭解，除學術之外會在業界遭遇到的問題，讓研究過程相當扎實。

7. 參考文獻

- [1] Russell, Stuart, and Peter Norvig, " Artificial intelligence: a modern approach," 2002.
- [2] Jordan, Michael I., and Tom M. Mitchell, " Machine learning: Trends, perspectives, and prospects," Science, p. 255-260, 2015.
- [3] Cunningham, Pádraig, Matthieu Cord, and Sarah Jane Delany, "Supervised learning," Machine learning techniques for multimedia. Springer , Berlin, Heidelberg, p. 21-49, 2008.
- [4] Zhu, Xiaojin, and Andrew B. Goldberg, " Introduction to semi-supervised learning," Synthesis lectures on artificial intelligence and machine learning 3.1, p. 1-130, 2009.
- [5] Barlow, Horace B, " Unsupervised learning," Neural computation 1.3, p. 295-311, 1989.
- [6] Yan, Le Cun, B. Yoshua, and H. Geoffrey, " Deep learning," nature 521.7553, p. 436-444, 2015.
- [7] Bishop, Chris M, " Neural networks and their applications," Review of scientific instruments 65.6, p. 1803-1832, 1994.
- [8] O'Shea, Keiron, and Ryan Nash, " An introduction to convolutional neural networks," arXiv preprint, arXiv:1511.08458, 2015.
- [9] LeCun, Yann, et al, " Object recognition with gradient-based learning," Shape, contour and grouping in computer vision. Springer, Berlin, Heidelberg, p. 319-345, 1999.
- [10] He, Kaiming, et al, " Mask r-cnn," Proceedings of the IEEE international conference on computer vision, p. 2961-2969, 2017.
- [11] Redmon, Joseph, and Ali Farhadi, " Yolov3: An incremental improvement," arXiv preprint, arXiv:1804.02767, 2018.
- [12] Long, Jonathan, Evan Shelhamer, and Trevor Darrell, " Fully convolutional networks for semantic segmentation," Proceedings of the IEEE conference on computer vision and pattern recognition, p. 3431-3440, 2015.
- [13] Ren, Shaoqing, et al, " Faster r-cnn: Towards real-time object detection with region proposal networks," Advances in neural information processing systems 28, p. 91-99, 2015.
- [14] Lin, Tsung-Yi, et al, " Microsoft coco: Common objects in context," European conference on computer vision. Springer, Cham, p. 740-755, 2014.
- [15] Pan, Sinno Jialin, and Qiang Yang, " A survey on transfer learning," IEEE Transactions on knowledge and data engineering 22.10, p. 1345-1359, 2009.
- [16] Hua, Siyang, and Ping Shi, " GrabCut color image segmentation based on region of interest," IEEE 7th International Congress on Image and Signal Processing, p. 392-396, 2014.
- [17] Powers, David MW, " Evaluation: from precision, recall and F-measure to ROC," informedness, markedness and correlation, arXiv preprint , arXiv:2010.16061, 2020.
- [18] AI Cup Mango FAQ, Jan. 2022
[Online] Available : <https://reurl.cc/Qj7zeo>
- [19] COCO Dataset, Jan. 2022
[Online] Available : <https://cocodataset.org/#home>
- [20] Matterport Mask R-CNN, Jan. 2022
[Online] Available : https://github.com/matterport/Mask_RCNN

[21] Facebook AI Research Detectron2, Jan. 2022

[Online] Available : <https://github.com/facebookresearch/detectron2>

[22] Facebook AI Research Detectron, Jan. 2022

[Online] Available : <https://github.com/facebookresearch/Detectron>

[23] Maskrcnn-benchmark, Jan. 2022

[Online] Available : <https://github.com/facebookresearch/maskrcnn-benchmark>