

應用人工智慧技術於港灣構造物自動化辨識管理

洪維屏¹ 韓仁毓² 甘翊萱³ 蔡立宏⁴ 李俊穎⁵

¹交通部運輸研究所港灣技術研究中心副研究員

²國立臺灣大學土木工程學系教授

³國立臺灣大學土木工程學系研究助理

⁴交通部運輸研究所港灣技術研究中心主任

⁵交通部運輸研究所港灣技術研究中心科長

摘要

近年來隨著無人飛行載具發展成熟，其具有高度的移動性以及遠距自動導控功能，能夠取代人力快速且輕易到達過去不容易接近的區域，並獲取大量的可視化觀測資訊。另一方面，人工智慧技術在近期有大幅度的發展，透過適當的訓練程序與學習策略，可針對大量數據進行快速的分析與判斷。本研究以無人飛行載具所拍攝之影像資訊，針對港埠設施之反光板、碰墊及車擋，進行單物件及多重物件的模型訓練及辨識測試，並以臺中港為示範港區，驗證本研究提出之方法可應用於港區巡查任務，日後可利用這些資訊建立港區重要設施的監測管理機制，達到以人工智慧技術提升管理效能。

一、前言

臺中港為我國國際商港，以近洋航線為主的中部區域加值型物流港，是主要能源、重工、石化原料進口港及油品儲轉中心。港區使用單位及管理項目繁多，相關土地利用與規劃是港區營運的重要工作之一，然而其幅員廣大，且人力有限，要快速掌握各區域之使用狀況並妥善管理難度極高。另一方面，港區設施受波浪及劇烈天氣狀況（如颱風、地震等）衝擊影響之下，其穩定性與安全性備受考驗，如何以有限人力持續監測港區環境並及早採取預防或補強措施，為一重要議題。

近年來航空攝影測量領域隨著無人飛行載具(以下稱 UAV)影像測繪技術不斷推陳出新，其針對任務需求搭載不同感測器進行影像拍攝，廣為應用於地形測繪及基礎設施巡查等項目。另影像記錄媒介從類比影像演進到數位元影像後，許多在電腦視覺計算上的理論也逐漸被引進到數位攝影測量應用當中，影像匹配技術更是除了已有廣泛應用的電腦視覺、遙測、醫學影像等領域外，也開始應用於各種類型的 UAV 影像，如近距離攝影、航拍圖像及車輛辨識(Liu et al., 2011； Malik and Siddiqi, 2014)。除此之外，Tsai and Lin (2017)和 Zhuo et al. (2017)更指出可利用 UAV 影像已知位置或飛行高度的特性，得知更全面的環境資訊並即時更新原有數據，在環境變化監測方面應用有很大的幫助。

大量影像需要辨識的前提下，人工智慧影像學習演算有助於提高效率。CNN (convolution neural network)在 2012 年被發揚光大後，以其強大的影像分類能力及高準確性一直大受歡迎。近年來，更有許多以 CNN 演算法為原型的物件辨識與切割模型，以及與各式因應不同需求和訓練方式所產生的模型興起。R-cnn (Girshick et al., 2014) 是第一個成功將深度學習應用到物件辨識上的訓練模型，作法是將區域縮放到統一大小，再使用 CNN 提取特徵後進行分類，最後通過非極大值抑制輸出結果。同期出現的還有 Yolo(Redmon et al., 2016)利用回歸的特性，將整張圖做為樣本輸入，直接在圖像的多個位置上回歸出位置的目標邊框及所屬類別，是一種將物體定位和分類一起完成的階段演算法。因其可使用 GPU 進行運算，可處理較大的數據量且進行即時運算，逐漸成為主流模型。綜觀研究結果(Ammar et al., 2019；Li et al., 2020) Yolo 模型的表現不論是精度或是運行速度皆優於其他模型。

本文將以 UAV 影像內的港區構造物做為訓練樣本，利用 YOLOv4 模型進行物件辨識，並依據不同偵測目標分成三種模型，測試構造物自動化辨識管理之可行性，最後提供評估成果的指標，且根據研究結果提出合理的驗證標準。

二、研究方法

針對不同港埠設施進行自動化巡查及辨識之研究步驟如圖 1 所示，共分成三部分：(1)訂定偵測目標，(2)UAV 巡查影像蒐集，(3)YOLOv4 物件辨識。



圖 1 研究流程圖

2.1 訂定偵測目標

擇定碰墊、反光板、車擋及繫船柱為主要偵測目標，再將此標的物於影像中進行單物件偵測及多重物件偵測模型分組，總共分為三組。

2.2 UAV 巡查影像蒐集

臺中港區幅員廣大，為使自動化辨識效能達到最高，拍攝巡查影像過程中，首先針對目標區域進行航線規劃，目的為沿港區堤岸邊構造物進行連續拍攝，另 UAV 在巡查拍攝的相片可獲取當下 UAV 的經緯度、高度與接收衛星數來提供未來定位資料使用，即可透過影像得知各項構造物位置；除此之外，相片資訊也會記錄拍攝當下飛機姿態與相機姿態，例如本次拍攝區域為臺中商港中南突堤及中碼頭間，以航高 40 公尺、離岸距離 50 公尺、鏡頭角度(飛機姿態)45 度

進行，且 UAV 搭載法國 Parrot Anafi 4K 照相機，於每一位置同時拍攝 5 張影像，並設定拍照間隔為 10 公尺 1 張，飛行速度每秒 5 至 7 公尺。

2.3 YOLOv4 模型架構

YOLOv4 的模型架構中，依序可以分為主幹網路(BackBone)、連階層(Neck)及檢測頭(Head)三部分，如圖 2 所示。輸入影像先經過 BackBone 進行影像特徵提取，由於影像特徵通常具有共通性，因此 BackBone 常使用已經在大量數據上訓練好的網絡；另 BackBone 輸出的特徵會傳遞給 Head 進行預測，而在這兩者之間的神經網路層即為 Neck，Neck 的用意在於彙整 BackBone 輸出的特徵影像，有助於 Head 預測；Head 為最後的預測層，進行物件的定界框(Bounding Box)與類別(Class)的預測。

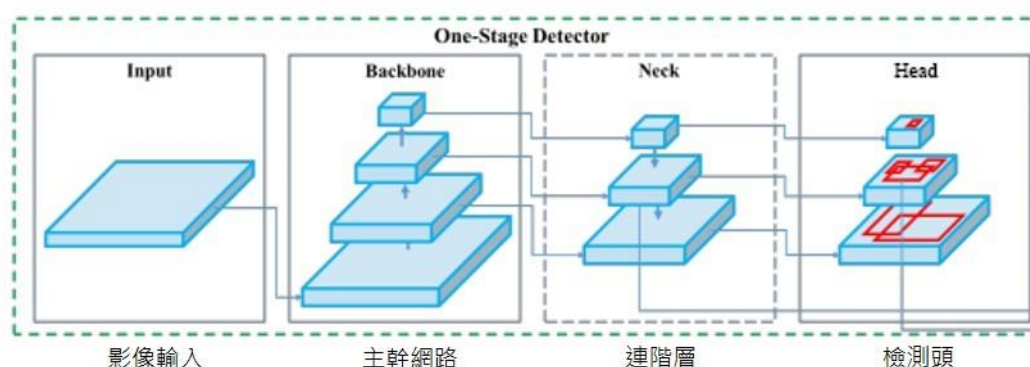


圖 2 YOLOv4 架構 (改自 Alexeu Bochkovskiy et al., 2020)

(1) Backbone：

因檢測模型的計算量較大，為加速並達到併行計算之優化，Wang et al.(2020)做了許多測試，發現 Darknet53 與 Cross Stage Partial Network (CSPNet)最適合用來做檢測；其中 CSPDenseNet 是基於 DenseNet 改良的網絡結構，DenseNet 為改善因模型中計算產生的梯度訊息損失問題，因此提出跨階段局部網路，意即每一層的輸出都會傳遞給其後的每一層網路，而非僅只傳遞給下一層，DenseNet 使最後獲得豐富訊息有效降低損失，但也導致運算量大幅上升。CSPDenseNet 則改良了其結構，每層僅一半的資訊透過 DenseNet 保留豐富的訊息，另一半則直接傳遞，達到降低計算量的目的，但也還能有 DenseNet 的效益，兩種演算方式如圖 3 所示。

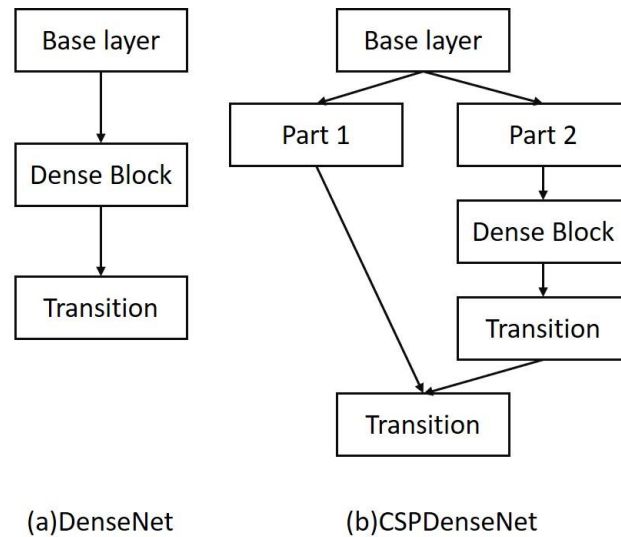


圖 3 DenseNet 與 CSPDenseNet 演算法(Wang et al.,2020)

(2) Neck :

此部分為了增加可偵測範圍與融合不同尺度特徵，結合 SPP(Spatial pyramid pooling layer)及 PANet(Path Aggregation Network)兩種模型。SPP 解決過去 CNN 模型在輸入與輸出大小有嚴格限制(如圖 4)，並且在該限制下會導致資訊變形或損失的問題。SPP 能使最後輸出的資料固定，因此可以使用多種不同的池化層大小來做不同尺寸的特徵萃取，並保證輸出的維度。PANet 強化低層訊息以解決資訊被忽略的問題，縮短低層及特徵的傳遞路徑，用於將不同尺度的特徵影像彙整並傳遞給 Head 進行預測。

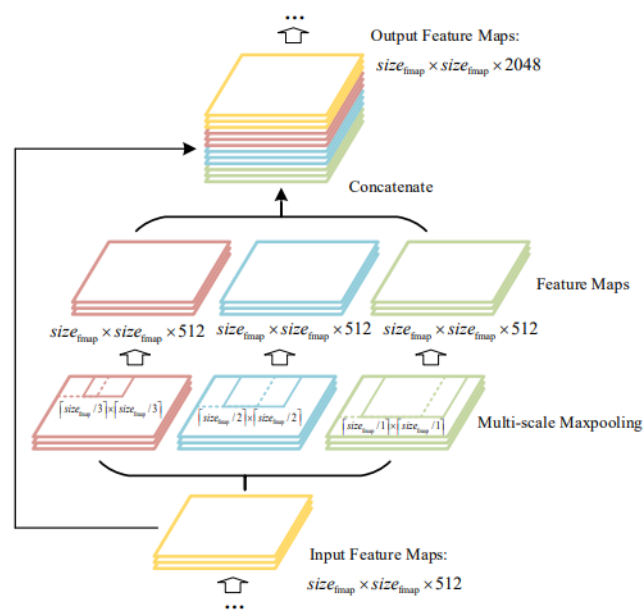


圖 4 SPP 示意圖

(3) Head：

Neck 整合好特徵後會傳遞給 Head 進行 bounding box 的預測。Yolov4 沿用 Yolov3 的 Head，總共分支出三個 Head，分別負責大物件、中物件及小物件的偵測，在各別的 Head 中會將圖片切分成多個固定數量的方框(13×13 , 26×26 , 52×52)，每個方框中有多個基準方框(Anchor)，Head 則預測目標物與 Anchor 之間的偏移量，每個方框可以預測三個 Bounding Box，最後則由信賴指數來剔除可能性不高的方框。

三、實驗結果與分析

根據無人機拍攝的影像並配合訓練樣本，進行了三種不同的模型訓練測試，包括單物件偵測、多重物件偵測(一)與多重物件偵測(二)，詳細資訊如表 1。整體模型進行流程為先獲取原始圖像，並利用相關標記軟體進行此次需要判斷之物件框選，接著將資料輸入做模型訓練，訓練完成後再輸入測試照片來做物件辨識，最後獲取辨識後結果。

表 1 不同模型訓練統整表

項目	單物件偵測	多重物件偵測(一)	多重物件偵測(二)
偵測目標	碰墊	碰墊 反光板(黃) 車擋	碰墊 反光板(黃+藍+紅) 車擋 繫船柱
訓練樣本數	100 張	100 張	100 張

3.1 實驗成果

3.1.1 單物件偵測

在影像訓練中，選 100 張場景構造物種類分布差異較大的照片，並利用圖 5 的 LabelImg(替深度學習模型做數據的軟體)進行碰墊選取，在 6000 次整體訓練下最終約使用 5~6 小時，可得到該模型的損失曲線(如圖 6)，最終損失值為 0.2104%，保有高於 99%物件特性，可確定模型之訓練完整性。模型最後會將輸入之照片來偵測碰墊位置，並標示偵測信心指數(如圖 7 標示值為 1.00)，其中設定信心指數未滿 0.6 則不標示，由此結果可發現因使用的訓練資料大多是同一方向，因此當相機的拍攝角度過斜則會導致碰墊的偵測出現問題。

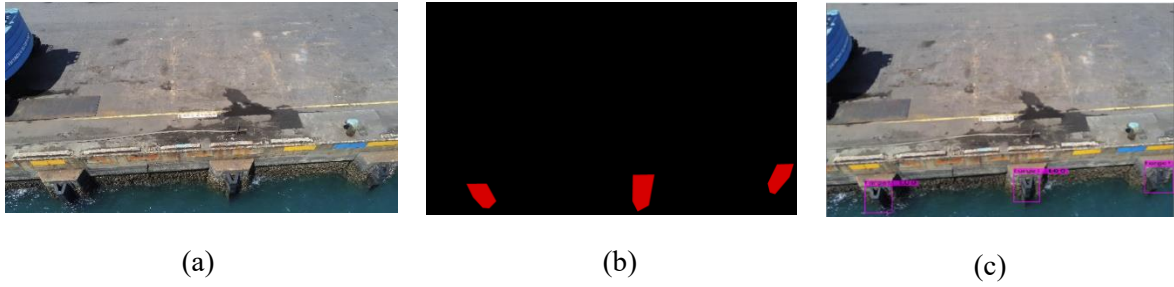


圖 5 單物件自動辨識 (a)欲訓練原始圖像 (b)人工框選預辨識物件示意圖
(c)模型辨識後框選物件結果圖

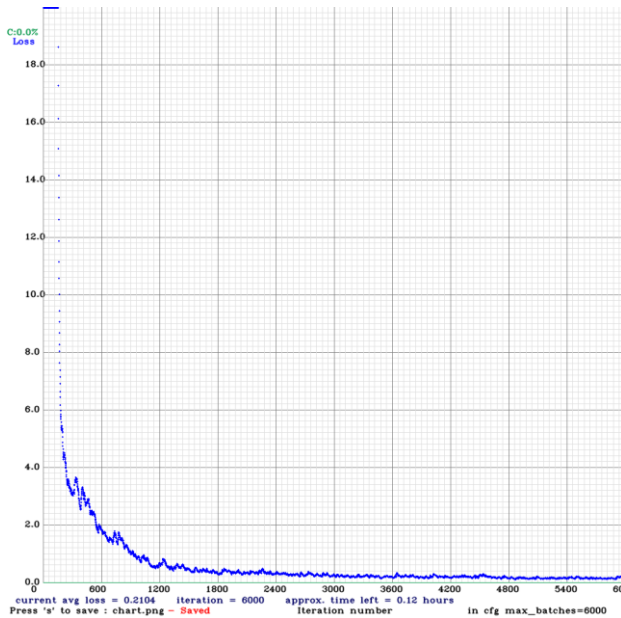


圖 6 單物件訓練後之損失曲線圖



圖 7 單物件偵測成果圖(標示出碰墊與偵測信心指數)

3.1.2 多重物件偵測(一)

應用與單物件測試相同之影像進行多物件辨識(如圖 8)，其可同時偵測碰墊、反光板及車擋，其中黃色反光板可有效被偵測標註，而車擋以藍色標註，整體信心指數高於 0.9。訓練次數為 6000 次，約花費 5~6 小時，得到該模型的損失曲線之損失值為 0.8342% (如圖 9)，顯示多重物件偵測率達 99%，多重物件偵測成果如圖 10 所示。

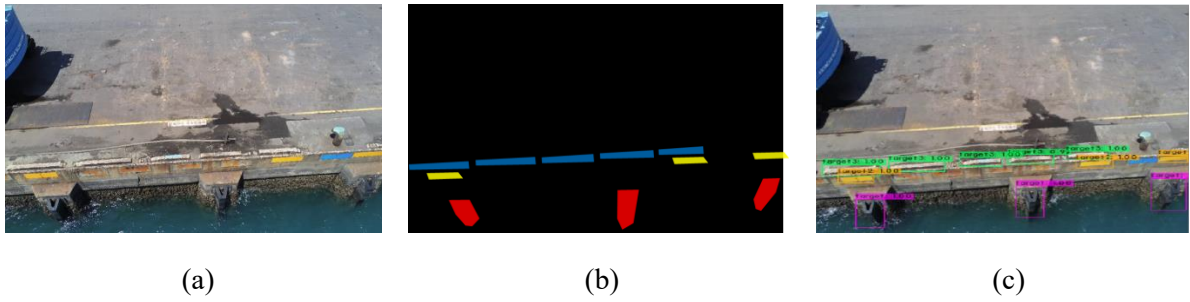


圖 8 多重物件自動辨識(一) (a)欲訓練原始圖像 (b)人工框選預辨識物件示意圖
(c)模型辨識後框選物件結果圖

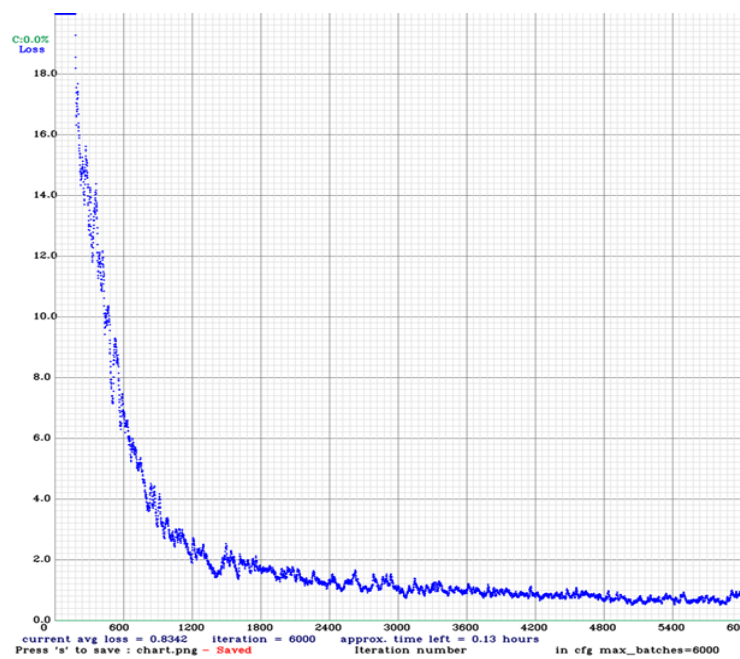


圖 9 多重物件(一)訓練後之損失曲線圖

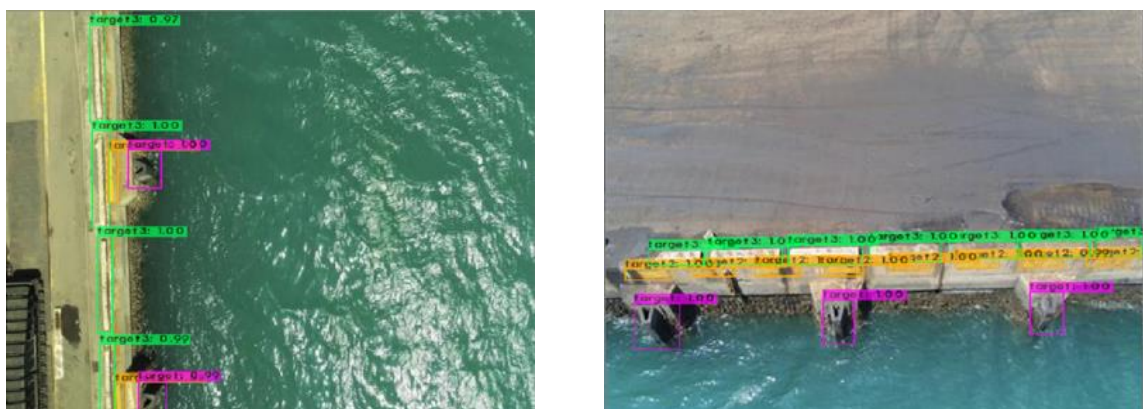


圖 10 多重物件偵測(一)成果圖

3.1.3 多重物件偵測(二)

在多重物件偵測得到初步成效後，證實多重辨識的可行性，於是進行第三類模型測試(如圖 11)，此時模型新增藍色反光板及繫船柱，並分析是否會影響模型的偵測效果。在 6000 次訓練下約耗時 5~6 小時，最後可得模型之損失曲線值為 0.8753%(如圖 12)，偵測成果如圖 13 所示。

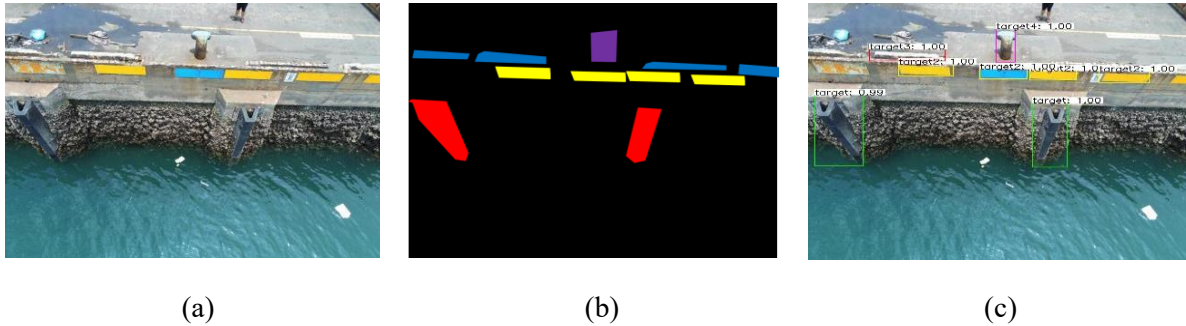


圖 11 多重物件自動辨識(二)(a)欲訓練原始圖像 (b)人工框選預辨識物件示意圖

(c)模型辨識後框選物件結果圖

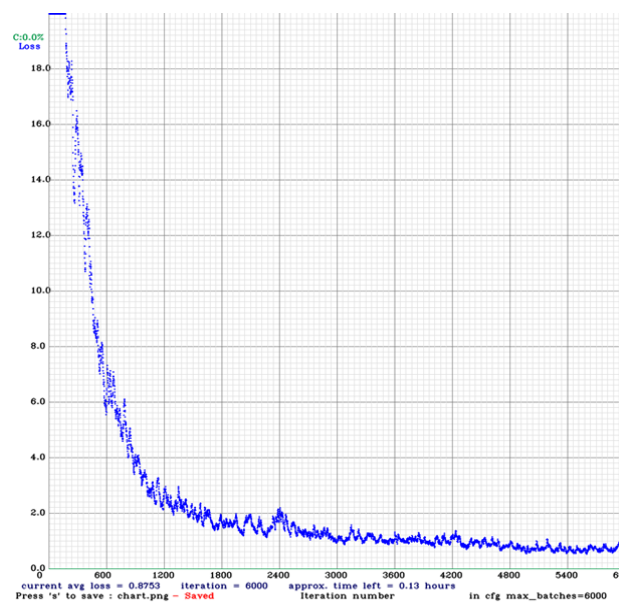


圖 12 多重物件(二)訓練後之損失曲線圖



圖 13 多重物件偵測(二)成果圖

3.1.4 物件辨識信賴度評估分析

為確認辨識成果之可信度，Yolov4 的評估指標主要採用 IOU(Intersection over Union) ，其於辨識後會給予每一結果信賴指數(confidence score)(如圖 14)。

$$\text{confidence score} = \text{Pr}(\text{object}) * IOU_{pred}^{truth}$$

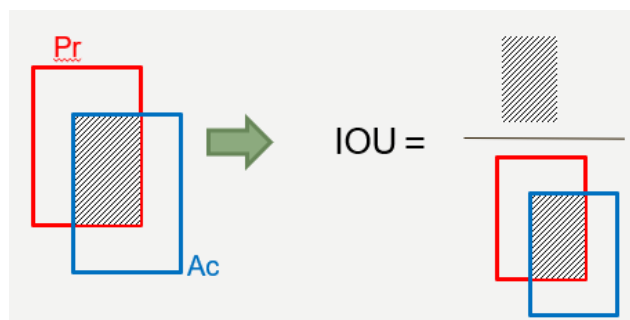


圖 14 IOU 指標計算示意圖

一般而言當該分數高於 50%時，可以視為 Yolov4 有偵測到物件，如圖 15(a)所示；而低於 50%時，則需要人力做進一步檢查，其中可能發生原因為物件於相片中邊緣導致重疊率過低所致，如圖 15(b)所示。基於港區資料之實際測試成果，本研究建議在信賴指數高於 0.7 時，表示模型判斷該物件被完整偵測且可信性較高；若信賴指數介於 0.4~0.7 之間，則應配合其它角度相片做進一步確認，而信賴指數低於 0.4 以下且無其它照片可供檢核時，則建議安排人力前往現場進行檢視。



(a)



(b)

圖 15 信賴指數驗證結果(a)高於 50% (b)低於 50%

四、結論

為了從 UAV 影像獲得更多資訊，諸如物件是否存在或毀損，而使用 YOLOv4 進行連續辨識，其中同一位置物件需透過多張相片辨識與交互確認，以提高偵測準確率，未來可以 UAV 搭配自動化辨識技術輔助進行港區巡查作業。

隨著巡查次數增加於不同時段及環境下進行資料收集，其大量樣本可使巡查流程使用之模型更加優化，也可朝該方向去做探討與改善，並提供相關巡查影片驗證影片辨識結果，有助於改善單張相片可能出現角度與其他原因造成的偶然性錯誤。

此外，因礙於目前影像取像方式，雖已成功偵測大部分物件，但物件於相片中過小或拍攝角度過大會偵測失敗的可能性，未來應針對欲偵測設施大小及其所需精度，彈性調整取像方式，並根據不同種類巡查項目進行影像模型訓練，以確保各設施能於影像中被清楚辨識與偵測。

參考文獻

1. Ammar, A., Koubaa, A., Ahmed, M., & Saad, A. (2019). Aerial images processing for car detection using convolutional neural networks: Comparison between faster r-cnn and yolov3. *arXiv preprint arXiv:1910.07234*.
2. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.
3. 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).
4. Li, M., Zhang, Z., Lei, L., Wang, X., & Guo, X. (2020). Agricultural greenhouses detection in high-resolution satellite images based on convolutional neural networks: Comparison of faster r-cnn, yolo v3 and ssd. *Sensors*, 20(17), 4938.
5. Liu, Z., An, J., and Jing, Y., (2011). A simple and robust feature point matching algorithm based on restricted spatial order constraints for aerial image registration. *IEEE Transactions on Geoscience and Remote Sensing*, 50(2), 514-527.
6. Malik, Z., and Siddiqi, I., (2014). Detection and recognition of traffic signs from road scene images. In *2014 12th International Conference on Frontiers of Information Technology*, 330-335.
7. 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).
8. Tsai, C. H., and Lin, Y. C., (2017). An accelerated image matching technique for UAV orthoimage registration. *ISPRS Journal of Photogrammetry and Remote Sensing*, 128, 130-145.
9. Wang, C. Y., Liao, H. Y. M., Wu, Y. H., Chen, P. Y., Hsieh, J. W., & Yeh, I. H. (2020). CSPNet:

A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 390-391).

10. Wang Wanguo, Tian Bing, Liu Yue.,(2017) Research on recognition of small power components in UAV patrol image based on RCNN [J]. Journal of Geo-Information Science, (2)
11. Zhuo, X., Koch, T., Kurz, F., Fraundorfer, F., and Reinartz, P.,(2017). Automatic UAV image geo-registration by matching UAV images to georeferenced image data. Remote Sensing, 9(4), 376.