

利用 Mask R-CNN 偵測並分類空拍影像中車輛¹

Aerial Image Vehicle Detection and Classification Using Mask R-CNN

王鈞奕 Jyun-Yi Wang², 蘇志文 Chih-Wen Su³,
黃家耀 Walter Ka-Lo Wong⁴, 張開國 Kai-Kuo Chang⁵,
葉祖宏 Tsu-Hung Yeh⁶, 林樹德 Shu-Te Lin⁷,
孔垂昌 Chui-Chang Kung⁸, 溫基信 Chi-Sin Wen⁹

摘要

在智慧型交通運輸系統中，城市交叉路口的車輛檢測和分類一直是一個重要問題。隨著無人機的快速發展，越來越多基於影像的車流分析研究標的，從傳統監視器影像轉為空拍影像。在本研究中，我們利用最新的 Mask R-CNN 深度學習架構，針對空拍影像進行車輛偵測和分類任務。為了能更精確地偵測空拍影像中的車輛區域，本研究利用 Mask R-CNN 抽取出像素級的車輛區域，並透過可旋轉的矩形加以擬合。同時，我們也將車輛區分為小客車、貨卡車、巴

¹ 本文為道路交通車流及事故風險偵測與分析工具之發展應用(計畫編號: MOTC-IOT-106-SEB009)部分研究成果，特此致謝。

² 中原大學資訊工程系碩士。

³ 中原大學資訊工程系專任助理教授(聯絡地址: 32023 桃園市中壢區中正路 200 號, 電話: 03-2654730, E-Mail: lucas@cycu.edu.tw)。

⁴ 國立交通大學運輸與物流管理學系專任副教授。

⁵ 交通部運輸研究所運輸安全組組長。

⁶ 交通部運輸研究所運輸安全組副組長。

⁷ 交通部運輸研究所運輸安全組助理研究員。

⁸ 交通部運輸研究所運輸安全組研究員。

⁹ 訊力科技股份有限公司副總經理(研究案計畫主持人)。

士與其他類型車輛。實驗結果顯示，Mask R-CNN 在平均精度均值的表現上，明顯優於另一種常用於車輛偵測的深度學習架構 Faster R-CNN。

關鍵詞：空拍影像、影像處理、深度學習、車輛偵測與分類

Abstract

In intelligent transportation systems, vehicle detection and classification at urban intersections has always been an important issue. With the rapid development of unmanned aerial vehicles (UAVs), more and more researches have been starting to perform image-based traffic analysis on aerial videos rather than CCTV videos. In this study, one of the recently developed deep learning architectures, namely Mask R-CNN, was adopted for vehicle detection and classification tasks in aerial images. In order to detect the area of a vehicle more precisely, the pixel-level segmentation of each vehicle was extracted and fitted by a rotatable bounding box. In the meantime, this study divided the detected vehicles into four categories: sedan, truck, bus and other. The experimental result shows that the mean Average Precision of Mask R-CNN is significantly better than that of another commonly used deep learning architecture, Faster R-CNN.

Keywords: Aerial Image, Image Processing, Deep Learning, Vehicle Detection and Classification.

一、前言

在複雜的都會區交通環境下，為能有效掌握高事故風險地點之車流軌跡，一般監視器影像中不同角落的車輛方向、大小、形狀都不相同，加上攝影機視角的差異與監控影像品質之問題，造成傳統影像監控技術難以直接應用到不同攝影機所拍攝的影像內容，往往需要事前對系統參數進行人工調整，才能應用不同攝影機所拍攝之影像。

而無人機(Unmanned Aerial Vehicle)由於可懸停於空中，能利用下方酬載攝影機垂直俯拍路口車流影像，因此幾何失真遠較一般道路監視器影像小，同時也迴避傳統電腦視覺領域物體辨識(分別是車輛間遮蔽及不同視角下之車輛辨識問題)難題，進而偵測出道路上車輛之位置。而近幾年來隨著深度學習的崛起，許多科學領域都有相當突破性的發展，特別是在影像偵測和分類方面，十年前被認為是難以突破的問題，如今都有了突破性的發展。因此，本研究希望利用深度學習中的卷積神經網路(Convolutional Neural Networks, CNNs)架構，透過空拍影像進行車輛的偵測與分類工作。

二、文獻回顧

近年來隨著智慧城市的興起，自動化服務以及非監督式系統越來越受到重視。在智慧型運輸系統(Intelligent Transportation System, ITS)中，許多應用也隨之產生，例如：自動駕駛、駕駛輔助系統、即時交通路況分析、意外偵測、自動化交通監控系統等。在這些應用中，車輛偵測不僅是基本的議題，且要在混和車流的都市交通環境之下，利用影像處理技術來偵測車輛更是一大挑戰。

用以偵測車輛之硬體感測器種類繁多，包括：環路線圈偵測器、微波雷達偵測器、超音波偵測器、無線射頻辨識(Radio Frequency Identification, RFID)偵測器、攝影機(套用影像處理)等。環路線圈偵測器是一種侵入式的偵測器，這類的偵測器設置時必須埋在地底下或是鋪在地面，由於設備架設地點的限制，通常僅用於高速公路或較寬敞的道路；微波雷達偵測器，雖然偵測範圍廣且不易受天氣影響，但兩台偵測器間如果距離太近會造成訊號干擾；超音波偵測器則是會因為氣溫變化影響偵測效能，造價較昂貴之外，感測距離也較短；無線射頻辨識偵測器也有訊號干擾的問題，使得讀取率降低。上述偵測器多半適用於小範圍車道，較難以分析廣範圍交岔路口之複雜車流狀況。

近年來隨著無人機低價化與普及化的趨勢，本研究以無人機空拍之方式，能將交岔路口之交通資訊一覽無遺的呈現在影像中，並透過影像處理技術來偵測不同車輛與其他進階資訊，有效地整理分析影像中之大量交通資訊，並從交通數據中檢測交岔路口交通衝突風險的存在。

在 Khana(2017)等人研究中提到無人機在交通相關領域應用與研究，並統整出包含硬體面與軟體面的數個重要步驟，其中包含無人機相關之飛行計畫、飛行方式、影像等資料蒐集方式，加上所須的資料分析技術、資料呈現與詮釋方法，以及針對交通問題的參數最佳化處理等步驟，如圖 1 所示。

近年來隨著無人機的普及化，利用無人機進行車流分析已成為新的趨勢，相對於一般地面監視器的俯拍方式，無人機酬載之數位攝影設備可以從完全垂直於地面之角度進行拍攝，可克服一般監視器因視角而產生車輛變形的問題，每部車輛在影像中可始終維持一致的車頂外觀，且拍攝範圍也可在不利用廣角鏡頭之前題下，取得任意大小的合適拍攝範圍。

在 Wang(2008)等人所提車輛偵測與追蹤方法架構中，系統性地將傳統物件偵測與追蹤技術應用至空拍影像。其車輛偵測與追蹤流程如圖 2 所示，包含了影像對齊校正、車輛偵測、車輛追蹤等步驟。由於空拍影像中的車輛外觀較一般監視器影像單純且容易處理，因此非常適合傳統的影像處理技術。本研究係採用類似的處理架構，取得車輛的位置與軌跡特徵。首先透過無人機取得交岔路口之連續空拍影像，這些影像會經過影像晃動校正(Image Registration)等前處理，取得穩定的連續空拍影像以確保所有影像座落在同一個 2 維座標系統中。在車輛偵測與分類上則引入最新的深度學習架構，

以提升偵測結果的穩定性。由於傳統上的影像特徵，如邊緣(Edge)、輪廓(Contour)等，相當容易受到拍攝環境光線變化、甚至建築物陰影等影響而無法準確取得。但隨著近年來深度學習在影像處理上的成功發展，已經漸漸取代利用傳統影像特徵的偵測方式，本研究希望透過導入深度學習強化車輛偵測的準確度，同時也輔以特徵追蹤技術，修正未能準確偵測的車輛位置，以提供完整精確的車輛軌跡資訊。

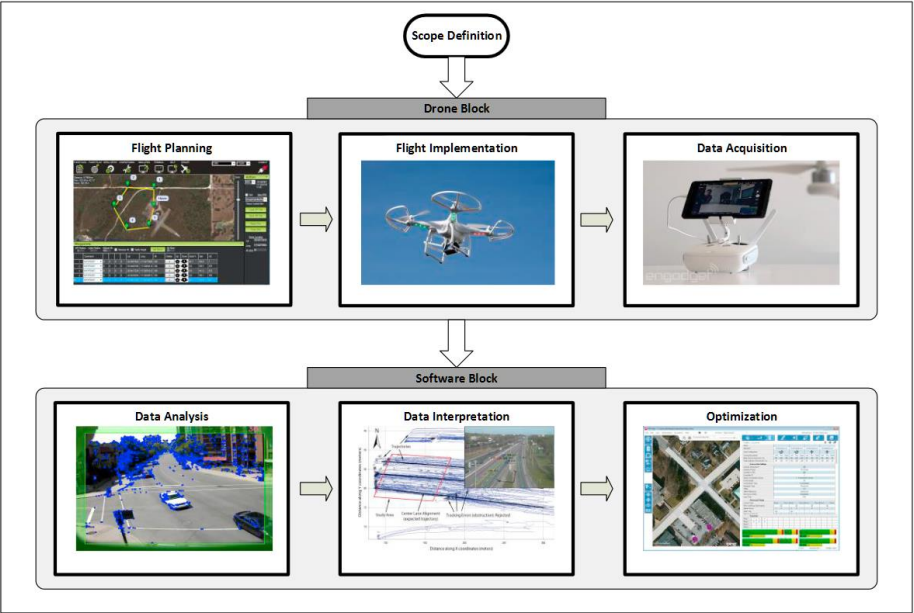


圖 1 無人機應用於交通領域研究之系統架構圖(Khana et al., 2017)

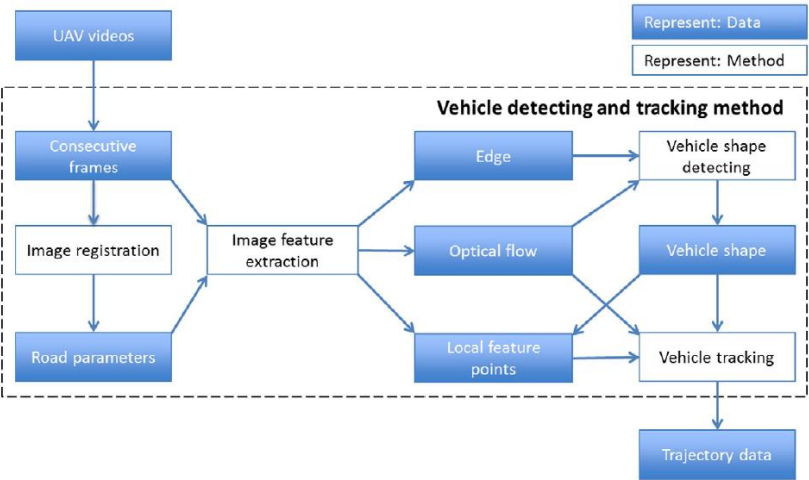


圖 2 車輛偵測與追蹤流程圖(Wang et al., 2008)

隨著資訊技術的蓬勃發展，機器學習已逐漸成為現今車流分析中不可或缺的一環。過去支援向量機(Support Vector Machine, SVM)、方向梯度直

方圖(Histogram of Oriented Gradient, HOG)、Adaboost 與 Harr-like 特徵等分類器與特徵的結合，被大量用於偵測車輛位置與及車種分類兩大問題上。但近幾年來隨著深度學習的崛起，使得許多科學領域都有相當突破性的發展，特別是在影像辨識和分類方面，十年前被認為是難以突破的問題，如今正陸續獲得解決。因此，本研究希望利用深度學習中的卷積神經網路架構，依車種對車輛進行分類工作，卷積神經網路與過去機器學習方法的最大不同之處，在於過去方法皆有賴人類智慧去設計影像特徵，再透過機器學習根據影像特徵進行分類，因此影像特徵設計上的好壞會影響分類結果。然而由人類設計的影像特徵往往較為理想而單純，難以適用所有複雜而多變之實際影像內容。相對地，卷積神經網路將影像特徵的生成涵蓋在其網路架構裡，透過非常大量的影像訓練，在不同尺度(Scale)上產生大量適用於各類影像的局部特徵，這些由百萬張真實照片所訓練歸納得出的局部影像特徵，遠較人類設計的影像特徵更為泛用，也更能呈現多數影像的局部特質。另一方面，網路中的多層結構也保障了影像描述上所需的複雜結構，故影像辨識和分類問題是目前深度學習最成功的領域之一。

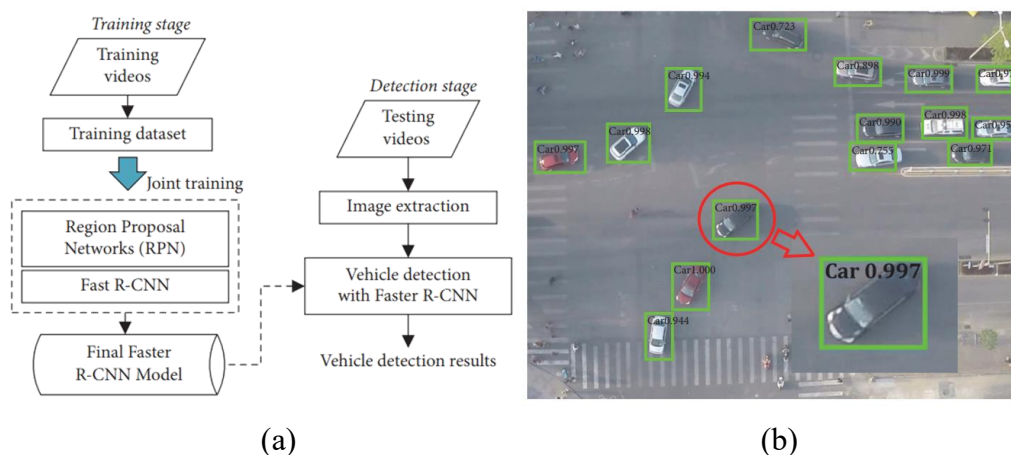


圖 3 Xu 等人所提之(a)車輛偵測流程圖與(b)車輛偵測結果 (Xu et al., 2017)

Xu 等人(2017)在 2016 年將 Faster R-CNN(Regions with Convolutional Neural Network features)(2016)架構用於空拍影像之車輛偵測分類，如圖 3 所示。在單純偵測出車輛不分車種的情況下，準確率可以達到 94.94%。由於上述例子的成功，因此我們認為可以從現有的卷積神經網路架構中挑選出符合需求的架構，來達成路口空拍影像下的車輛偵測與分類工作。

Tang 等人(2017)在 2017 年提出了同樣基於 Faster R-CNN 的車輛偵測方法，透過一種車輛提案網路(Vehicle Proposal Network, VPN)，組合不同層的特徵來生成候選車輛狀區域。然後，使用車輛分類網路(Vehicle Classification Network, VCN)以進一步驗證候選區域，並在八個方向上對車輛進行分類。由於基於卷積神經網路的深度學習方法，通常需要大量的人工

標記的訓練數據，但空拍影像的標記數據往往相當有限。因此，他們將原始的大型空拍影像裁切，並通過以四個角度（即 0° ， 90° ， 180° 和 270° ）旋轉增加訓練資料集的大小。

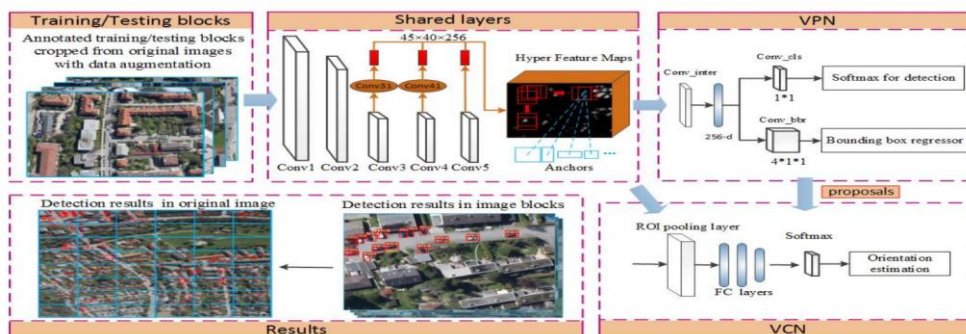


圖 4 Tang 等人設計之架構圖(Tang et al., 2017)

Sommer 等人(2017)在 2017 年將 Faster R-CNN 與 FCN(Long et al., 2015) 做結合，詳如圖 5 所示。透過結合分割與偵測結果來減少偵測時的錯誤率，並且減少 region proposal 的分類數量，首先他們先將影像做透過 RPN 預測車輛偵測框，同時他們基於 FCN 設計出 3 種分割架構 AFCN-32、AFCN-MS 以及 AFCN-D16 將影像做分割，並將分割結果和 Faster R-CNN 結果結合，透過這種方法將錯誤率降低。

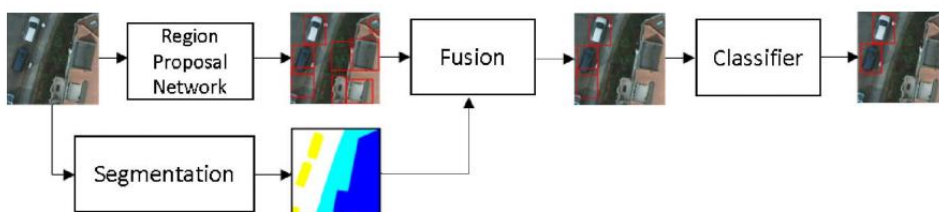


圖 5 Sommer 等人設計之架構圖(Sommer et al., 2017)

從過去文獻我們可以發現，一般 R-CNN 等深度學習架構下，車輛偵測框無法真正緊貼車體外緣。但在實際空拍影像中，車輛在轉彎過程中並不適合用邊界框來描述實際車輛範圍。因此本篇論文將重點探討空拍影像下，能夠準確偵測車輛外緣之卷積神經網路架構，以便未來能更精準地找出實際車輛外緣範圍，提升交通車流分析的準確性。

三、研究方法

在本章節中，將詳細討論本研究在空拍影像中的車輛偵測與分類方法。我們主要將四輪車輛分為 4 個種類，分別為小客車、貨卡車、公車與其他大型車(垃圾車或是救護車)，並且基於卷積神經網路架構下進行偵測與分類處

理。圖 6 為本研究方法的流程圖。由於我們使用的影像為空拍影像，容易受飛行環境影響導致影像位置偏離，於是透過影像穩定化將影像校正。為了省下不必要的運算處理，在實驗中僅針對路口中心四周進行偵測，對無人飛行載具所錄製的影像進行感興趣區域(Region of Interest)的擷取。所有後續實驗偵測、分類與分割皆於 ROI 範圍中進行處理。在車輛偵測與分類方面，我們採用 He 等人(2017)所提之 Mask R-CNN 架構，利用 Mask R-CNN 輸出之遮罩(Mask)找出車輛輪廓。除了獲取分類結果外，更利用其產生之物件分割結果來找出貼合車輛邊緣之矩形擬合(Rectangle Fitting)，取代由物體上下左右所形成之邊界框(Bounding Box)。

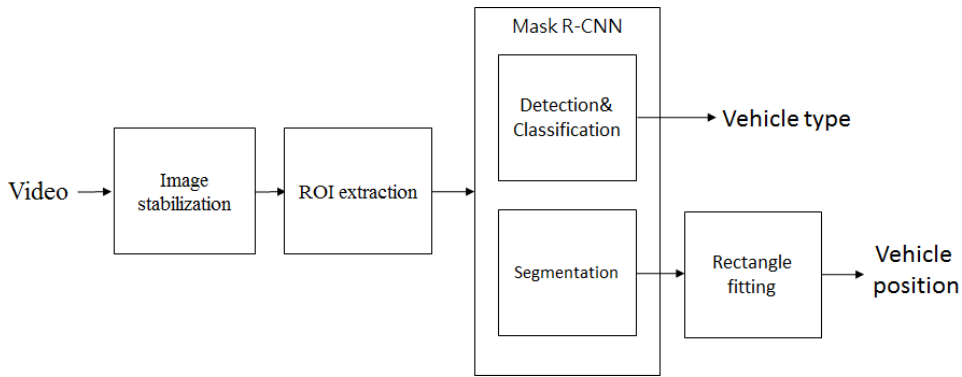


圖 6 本研究方法之流程圖

3.1 空拍影像穩定化

在利用空拍影像進行交通車流分析之前，由於無人機本身會受到陣風影響，導致影像中的街道位置偏離，因此須先針對影像進行對齊校準(Image Alignment)，以確保所有影像中的道路、車輛物件皆維持在同一個平面座標系統。本研究採用由 Evangelidis 與 Psarakis 所提出的 ECC(Enhanced Correlation Coefficient)影像對齊技術(Evangelidis et al., 2008)，透過找尋輸入影像與樣板影像間最大的相關係數，決定出一個 3×3 的轉換矩陣，讓輸入影像中的所有點都能透過這個轉換矩陣，投影到與樣板影像一致的平面座標系統，如圖 7 所示。

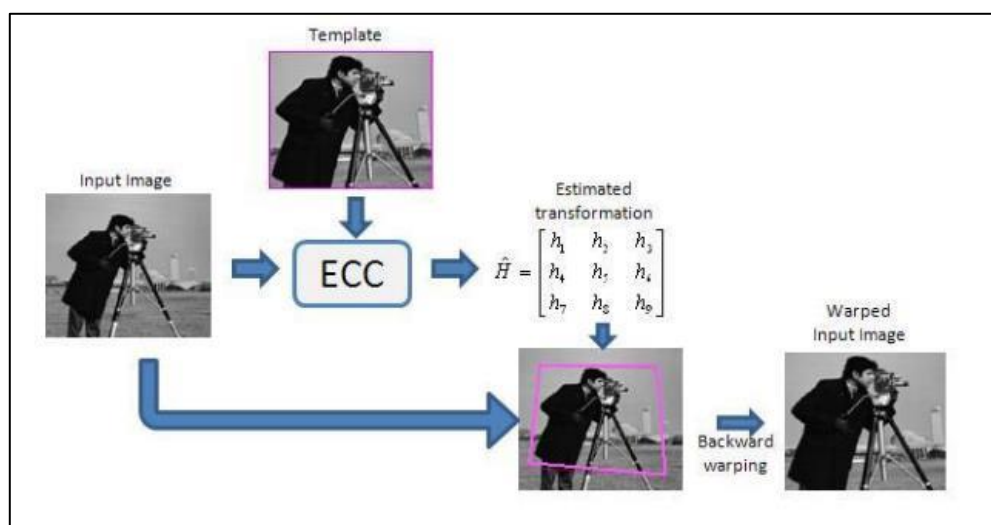


圖 7 ECC 影像對齊法之流程示意圖(Evangelidis et al., 2008)

本研究採用之空拍影片的拍攝高度為 75 公尺、解析度為 1920×1080，攝影鏡頭採非廣角鏡頭以維持影片中道路的直線性質。後續對經過穩定化處理的空拍影片，以十字路口為中心選定 1024×1024 範圍作為 ROI。在一般三線道環境下，此範圍足以涵蓋路口中心往外延伸至車道停止線後約 2 台車距離，如圖 8 所示。



圖 8 ROI 結果圖

3.2 利用 Mask R-CNN 偵測/辨識車輛

在最早的卷積神經網路架構中只能對影像中單一物體進行分類，但本研究需要在單張影像中對多部車輛位置進行偵測，於是使用 R-CNN 方法先

找出影像中車輛後進行分類，並透過線性迴歸找出車輛在影像中的位置。R-CNN 架構在物體偵測上有著相當成功的突破，但基於物體偵測的方法速度仍舊非常緩慢，Ren 等人在 2015 年基於 R-CNN 進行優化，發展出新的架構 Faster R-CNN(Ren et al., 2016)。自此，Faster R-CNN 成為在一般影像中偵測車輛的常見方法之一。隨著無人機的快速發展，越來越多空拍影像被應用於交通分析上，以彌補道路監視器畫面中傾斜視角所帶來的問題。特別是車輛在空拍影像中，可以更直觀地觀察車輛在道路上的行走軌跡。然而，Faster R-CNN 等一般深度學習方法偵測到的物件邊界框，卻無法緊貼於轉彎車輛的邊緣，如圖 3(b)。

有鑒於此，本研究引進 Mask R-CNN 可進行物件分割的特性，將其應用在空拍影像下的車輛邊框偵測。為了解決原本 R-CNN 架構中，物體偵測出的邊框無法貼合目標的問題，Long 等人(Long et al., 2015)在 2015 年發表了一篇基於 CNN 的影像分割架構，以找出物體完整的輪廓。他們將原本卷積神經網路最後幾層的完全連接層替換成卷積層，因此取名做 Fully Convolutional Networks(FCN)，如圖 9 所示。透過將卷積與池化後特徵圖反卷積，將縮小維度後的特徵圖還原到原大小，並對所有像素點進行分類，進而取得物體分割的結果。這個概念後來也被 Mask R-CNN 架構所採用，將分類、偵測和分割三大機制統合進了單一個架構之中。

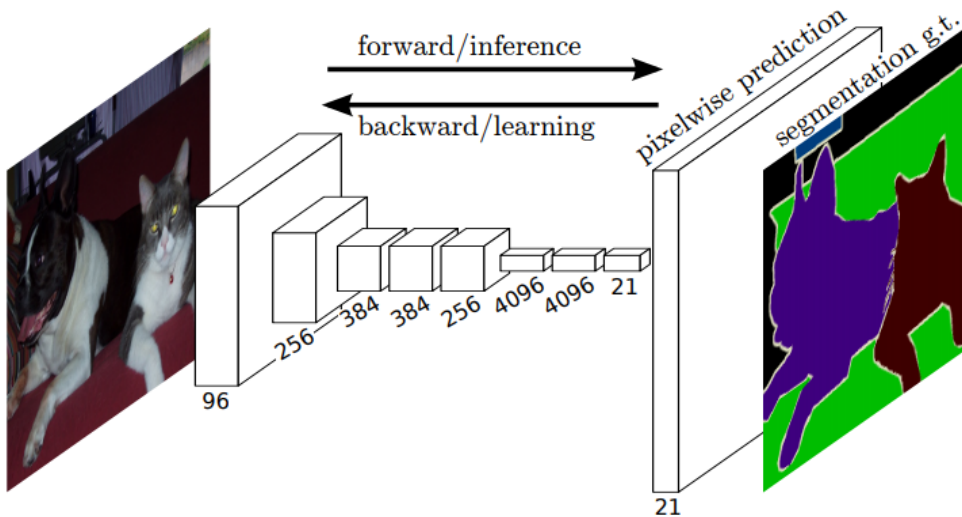


圖 9 FCN 架構圖(Long et al., 2015)

為了取得空拍影像中車輛的準確位置，我們選擇使用 Mask R-CNN 中的分割結果，來找出空拍影像中的車輛遮罩，圖 10(a)為 Faster R-CNN 偵測結果，其中紅框為小客車之偵測邊框、藍框為貨卡車之偵測邊框，可以清楚發現 Faster R-CNN 偵測到的物件邊界框無法緊貼於轉彎車輛邊緣；相對地，圖 10(b)為 Mask R-CNN 偵測結果，可以清楚發現 Mask R-CNN 以不同顏色

遮罩(mask)來表示每部車輛的範圍，並同時完成每部車的分類結果，如圖 10(c)。其中，紅色區域為代表小客車類別之像素點、綠色區域為代表貨卡車類別之像素點，

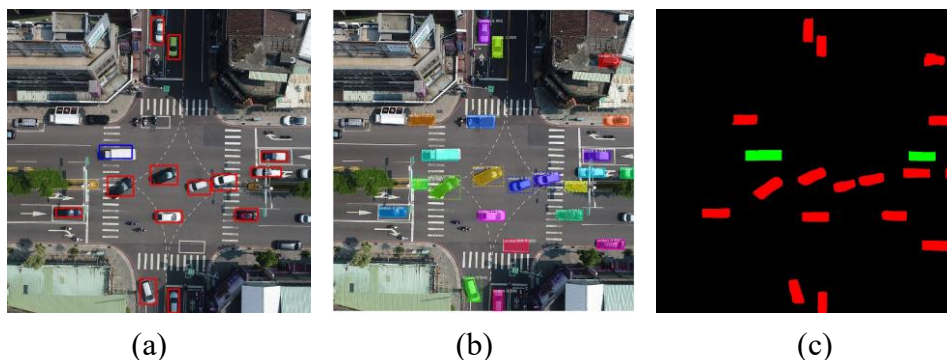


圖 10 (a) Faster R-CNN 偵測/分類結果；(b) Mask R-CNN 偵測結果；
(c) Mask R-CNN 偵測/分類遮罩

3.3 對車輛範圍進行矩形擬合測試環境與資料

由於 Mask R-CNN 的遮罩輪廓資訊量太大，無法有效描述車輛範圍，因此本研究透過對車輛輪廓進行矩形擬合，取出最小的輪廓外接矩形，並以該矩形 4 個角點來更有效地描述車輛範圍。我們利用 Mask R-CNN 的遮罩輪廓資訊，使用旋轉卡尺(Rotating Calipers)演算法(Toussaint, Godfried T., 1983)找出最左下角的點以及矩形的最長邊與最短邊，再找出旋轉中心與旋轉角度。利用這些資訊可以找出矩形的 4 個頂點座標位置，作為對該輪廓進行矩形擬合的結果，使得找出的矩形框能夠貼合不同方向的轉彎車輛，如圖 11 所示。

四、實驗結果與分析

4.1 實驗環境與測試資料

本研究的實驗環境皆於安裝 Windows 8 作業系統、搭載 Intel® Core™ i7-5820k CPU@3.30GHz，記憶體 80G，顯示卡 NVIDIA GTX1080Ti 11G 的電腦環境下執行，針對 1024×1024 大小之測試空拍影像進行。整體車輛偵測速度平均約為 3fps。

實驗的訓練與測試影片是透過無人機在 75 公尺高度所拍攝錄製的影片，一共拍攝了 6 個不同位置的交通路口影片。每部影片都在經過影片穩定化

後，以路口為中心重新裁切為 1024×1024 解析度的影片 fps。我們從 6 個路口的空拍影片中(如圖 12(a))，透過人工挑選上千張車輛數較多的不同時間畫面做為訓練資料。並從 4 個路口的空拍影片中(如圖 12(b))，隨機挑選出 4 段未包含在訓練資料的 3 分鐘長影片段落，以 1fps 頻率取樣出 720 張測試畫面。

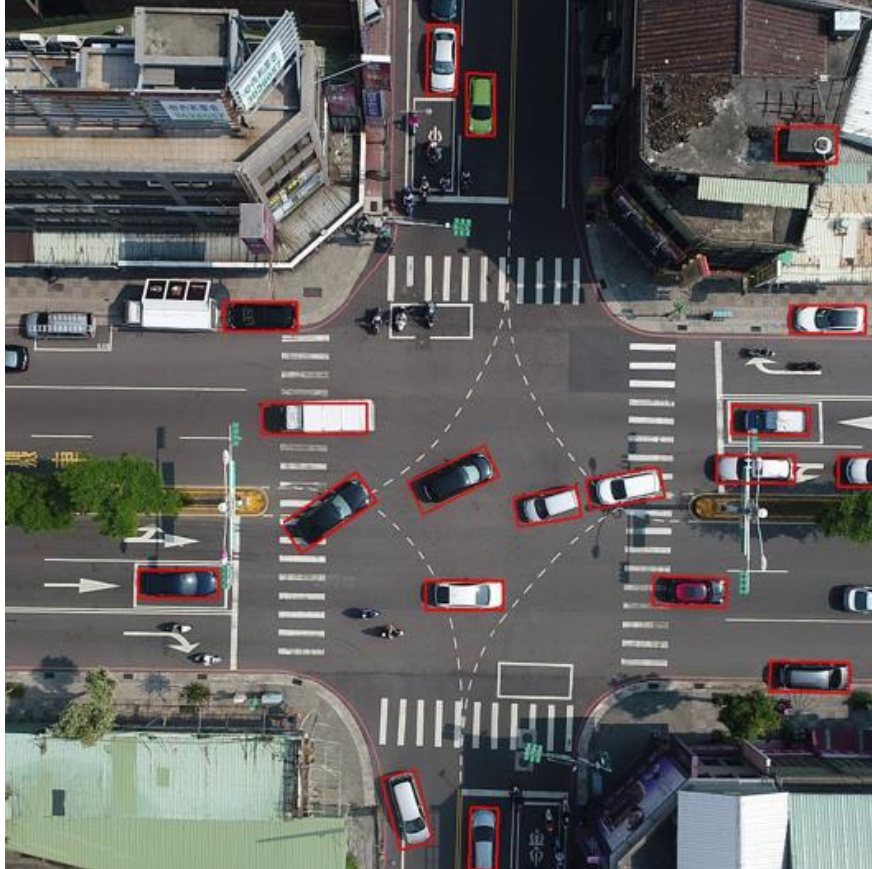


圖 11 矩形擬合結果

測試影片中之車輛數量如表 1 所示，每部影片皆為 3 分鐘，畫面擷取頻率為 1fps，包含 180 張測試畫面。在訓練車輛時，我們將車輛分為小客車(Sedan)、貨卡車(Truck)、巴士(Bus)、其他(Others)四類。其他類別為垃圾車、水泥預拌車等其他巴士以外大型車輛。



(a)



(b)

圖 12 (a) 6 個訓練用路口畫面範例；(b) 4 個訓練用路口畫面範例

表 1 測試空拍影像資訊

	# of Frames	Sedan	Truck	Bus	Others	Total
Video1	180	1078	302	11	0	1391
Video2	180	1168	120	27	0	1315
Video3	180	2039	152	18	0	2209
Video4	180	1557	140	11	6	1714
Total	720	5842	714	67	6	6629

4.2 實驗結果

在物件偵測上對於邊框定位的精度，通常使用 Intersection over Union (IOU)來進行度量，因為偵測結果不可能百分百跟人工標記的位置範圍完全匹配，因此就存在了上述 IOU 的定位精度評價公式，來評定物體偵測出的位置與實際位置的重疊比例。而 R-CNN 在訓練物體偵測時，也會先找出多個可能的物體矩形框，透過計算 IOU 大於一定閾值時來判斷它是否為目標物，如圖 13 所示。

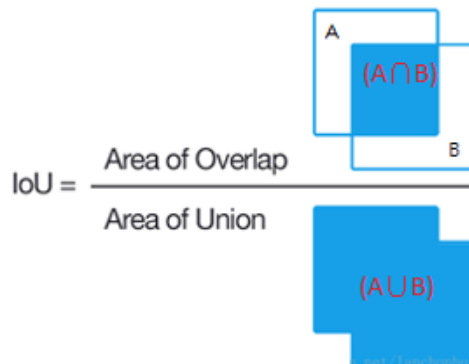


圖 13 IOU 示意圖

利用 IOU 計算出重疊率後，為了判斷與真實車輛位置重疊之偵測框是否被歸類為正確車種，我們採用了平均精度(Average Precision, AP)來當作衡量指標，如圖 14 所示。AP 是用來衡量資料檢索準確度的指標，對一個搜尋系統而言返回的結果必然是有序的，正確的結果要優先出現，於是有了 AP 的概念。其作法是對一個有序的目標對象列表，先求出每個正確分類位置上的車輛位置，然後計算出所有正確分類的平均精度均值(mean Average Precision, mAP)。

Mean Average Precision: example

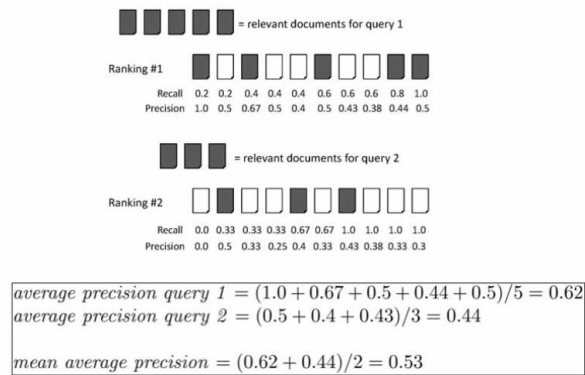


圖 14 AP 計算示意圖 (Lavrenko, 2013)

表 2 為 4 部測試影片的 mAP 結果，在 IOU 為 0.7 的情況下 Mask R-CNN 訓練的 mAP 結果皆能達到 7 成以上，在空拍路口影片 1 和影片 2 mAP 較為低的原因主要在於路口流向較為複雜，提高了誤偵測的發生機率。此外，從表 2 可以看出 Mask R-CNN 的偵測效果比 Faster R-CNN 明顯更高，代表 Mask R-CNN 的偵測結果能較完整的貼合車體，讓整體 mAP 比 Faster R-CNN 提升更高。

表 2 Faster R-CNN 與 Mask R-CNN 之 mAP 比較

	Faster R-CNN	Mask R-CNN
Video1	0.648	0.754
Video2	0.610	0.748
Video3	0.677	0.783
Video4	0.674	0.772
Average	0.652	0.764

接下來我們比較 4 個測試路口中的分類情況，由於在空拍影像中訓練資料各類別所占的比例極不平均，在交通路口小客車所占比例超過七成，在有足夠的訓練資料情況下，小客車能夠得到最好的訓練結果，但這也導致其他類別的訓練不足。

表 3 為各車種的整體分類結果，在此我們只對有成功偵測出的車輛進行分類，並對分類結果做分析。從表 3 結果中我們可以明顯看出在貨卡車的分類結果不佳，主要原因是由於貨卡車外觀會視其車斗所載貨物而有極大差異，加上訓練資料不足導致分類結果不好。然而對於訓練樣本較為充足的小客車，其正確分類的比例高達 99.4%；外觀與其他類車輛差異較大的巴士，亦有 96.8%可正確分類。

表 3 車輛分類之混淆矩陣

Ground truth	Classification				
		Sedan	Truck	Bus	Others
	Sedan	5500	18	14	0
	Truck	199	359	18	0
	Bus	1	1	61	0
	Others	5	0	0	0

五、結 論

本研究中我們利用 Mask R-CNN 對空拍影像進行了車輛偵測與分類，並透過矩形擬合取得更精準的車輛範圍。我們比較了 Mask R-CNN 與 Faster R-CNN 結果，經由實驗得知，Mask R-CNN 有助於在空拍影像中找出車輛轉彎時的完整輪廓，相較於 Faster R-CNN 等邊界框更適合交通分析上的需求。

由於深度學習需要大量的人工標記資料進行訓練，但由於需花費大量人力與時間，導致空拍影像的訓練資料量仍相當有限，未來仍可透過增加各類車輛訓練資料量，並增加在不同路口下的空拍影像資料，來增加分類模型在不同路口環境下的魯棒性(Robustness)。另一方面，如何改善網路架構，來進一步提升準確率或是直接透過深度學習，產生擬合後的矩形四個角點，都是值得後續探討的問題。

參考文獻

- Evangelidis, G. D. and Psarakis, E. Z. (2008), "Parametric Image Alignment using Enhanced Correlation Coefficient Maximization," *IEEE Transaction on Pattern Analysis & Machine Intelligence*, Vol. 30, No. 10, pp. 1858-1865.
- He, K., Gkioxari, G., Dollár P and Girshick, R. (2017), "Mask R-CNN," *IEEE International Conference on Computer Vision (ICCV)*, pp. 2980-2988.
- Khana, M. A., Ectorsa, W., Bellemansa, T., Janssens, D. and Wetsa G. (2017), "UAV-Based Traffic Analysis: A Universal Guiding Framework Based on Literature Survey," *ScienceDirect Transportation Research Procedia*, Vol. 22, pp. 541-550.

- Lavrenko, V. (2013), Evaluation 12: mean average precision, Retrieved Feb 15, 2019,
website:<https://www.youtube.com/watch?v=pM6DJ0ZZee0&feature=youtu.be&t=2m29s>
- Long, O., Shelhamer, E. and Darrell, T. (2015), “Fully Convolutional Networks for Semantic Segmentation,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431-3440.
- Ren, S., He, K., Girshick, R. and Sun, J. (2016), “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1137-1149.
- Sommer, L., Nie, K. and Schumann, A. (2017), “Semantic Labeling for Improved Vehicle Detection in Aerial Imagery,” *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1-6.
- Tang, T., Zhou, S., Deng, Z., Lei, L. and Zou, H. (July 2017), “Fast multidirectional vehicle detection on aerial images using region based convolutional neural networks,” *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1844-1847.
- Toussaint, G. T. (1983), “Solving Geometric Problems with the Rotating Calipers,” *Proceedings of IEEE MELECON'83*, Athens, Greece.
- Wang, X., Ma, K. T., Ng, G. and Grimson, W. E. (2011), “Trajectory Analysis and Semantic Region Modeling Using Nonparametric Hierarchical Bayesian Models,” *International Journal of Computer Vision*, Vol. 95, No. 3, pp 287–312.
- Xu, Y., Yu, G., Wang, Y., Wu, X. and Ma, Y. (Aug. 2017), “Car Detection from Low-Altitude UAV Imagery with the Faster R-CNN,” *Journal of Advanced Transportation*, Vol. 2017, pp. 40-49.

(收稿 107/09/28，第一次修改 108/02/15，接受 108/08/01，定稿 108/10/14)