



人工智慧於交通運輸潛在應用之回顧

張仲宇、鄭又嘉、何語萱、謝宗穎／國立臺灣大學土木工程研究所交通工程組 碩士生
朱致遠、陳柏華／國立臺灣大學土木工程研究所交通工程組 教授

人工智慧的再次興起與科技進步帶動了智慧運輸的發展。產官學界無不努力合作推升國內交通智慧化之程度，國外也不斷產出更精準之模型。如何結合國內外技術應用在國內交通環境是一個重要之在地化課題。本文回顧國內外文獻，首先介紹現有之影像辨識、重疊物件偵測、重識別以及聲音辨識等模式，並提出於交通領域可能之應用參考，其中包含跨攝影機追蹤、高密度車流環境偵測、低能見度環境下之偵測與高噪音機動車輛偵測。提出技術同時也思考應用於國內混合車流環境之可行性。期望透過本文激盪出更多智慧運輸之應用。

關鍵詞：物件偵測、影像偵測、聲音偵測、智慧運輸

The recurring rise of artificial intelligence and its progress drive the development of advanced transportation systems. Both the academia and the industry devoted themselves for smarter transportation in Taiwan. Models advancing the frontier of artificial intelligence has been proposed every year. How to adapt these approaches in Taiwan is a critical challenge. This article reviews recent papers where we first introduce existing image recognition, overlapping object detection, re-identification, and audio recognition models. Then, potential applications are proposed, including cross-camera tracking, high-density traffic tracking, low visibility detection, and noisy vehicle detection. We also consider the feasibility of mixed traffic flows in Taiwan. We hope that this article sparks more ideas on applications of intelligent transportation.

Key Words: Object detection, Image detection, Audio detection, Intelligent transportation.

前言

近年來隨著機器學習 (machine learning) 及人工智慧 (artificial intelligence, AI) 的再次興起，自動化及智慧化等課題成為許多領域發展的焦點。隨著國內外學界及產業界越來越多人才及時間的投入，機器學習及人工智慧技術在近年來突飛猛進。我國政府也投入相當的預算期望提升台灣智慧運輸的發展。交通部自 106 年起推動第一期「智慧運輸系統發展設計畫 (106-109) 年」^[1]，刺激國內相關智慧交通產業之蓬勃發展，109 年更接續提出「智慧運輸系統發展設計畫 (110-113) 年」^[2]，除延續第一期的成果外，更結合 5G、交通安全及人本交通等概念期望將台灣智慧交通發展成垂直產業生態鏈，邁向國際市場。交通部運輸研究所之「交通事件資訊整合服務與精進計畫 (1/2)」^[3] 詳述

了交通資訊平台的開發過程與應用，並提出改善方向。

傳統機器學習與現今人工智慧技術皆以數學模型為基礎，搭配其他學科如統計及機率來解決問題。近年資通訊科技的發展使得資料處理、傳輸與儲存能力大幅提升，偵測硬體設備的提升使得資料品質提升以及來源更加多元，運算設備的處理效能亦不可同日而語，這些發展都一再將人工智慧之能力推升至另一顛峰。為了達成更豐富之應用，其中所使用的理論與模型也越趨成熟。除傳統數值資料外，結合影像、聲音以及光達 (LiDAR) 所收集的點雲 (point cloud) 等資料作運算也成未來發展之趨勢。

傳統交通領域多使用數值資料，如：速度、流量與密度等，以建立模型進行分析。有些研究透過模擬軟體進行策略擬定之評估；卓訓榮等人^[4] 收集國道五

號車流資訊來擬定匝道控制策略，並利用微觀車流模擬軟體來評估與調整策略。亦有研究收集歷史資料，並利用數學模型進行分析，試圖找出顯著因子；鍾易詩等人^[5]收集國道交通事故資料，發現逆高斯脆弱對數羅吉斯特之加速失效模式能最有效捕捉事故特性，並分析變數與事故之關聯；黃柏宗^[6]利用支持向量機（Support Vector Machine, SVM）與隨機森林（Random Forest）分析交通事故資料，並進行碰撞機率之估計。相對於數值資料，影像與音訊能夠捕捉其他的特徵（feature），如：空間分佈資訊、移動軌跡、現場聲音與外觀等；資料的收集比起固定偵測線圈也更有彈性，設備既可裝設在燈桿也可採用移動式部署方式收集，如能與現有資料結合，能提供智慧運輸領域更豐富且完整之資料來源。吳沛儒等人^[7]以電腦視覺之AI相關技術自動辨識道路監視器影像中之車輛，獲取車流參數，也進一步以邏輯判斷偵測出交通事件。

開發出一高精確或高效率之模型一直是國內外資訊工程領域研究競逐的目標，每年國際研討會皆產出眾多新技術，而先進（state-of-the-art）模式皆以該領域的最高準確度為目標，參與舉辦之競賽（challenge）或使用標準化之比較基準（benchmark），將模型推向更高之精準度。然而，將這些模型應用在實際問題時可能需要一定之客製化，例如，將這些論文應用在交通領域，因模型本身不是專為交通應用設計，必定要對原始程式或輸入資料進行調整。牽涉到不同領域都有其核心知識（know-how），應用之前需將模型重新訓練（training）、進行實驗步驟或演算法的調整等，所以最新之技術通常不會馬上被應用到交通領域。而同為運輸領域之研究，也會因為交通環境不同，而需要調整模式。歐美等國家因機車較為少見，其所開發之模型也較少考慮到汽機車混合車流的情形^[8]。魏健宏等人^[9]對類神經網路應用於國內交通運輸作了系統分類整理，也提出可能之發展方向。

本文回顧國內外文獻，範圍包含交通領域期刊之現有應用以及資訊工程相關研討會以及國內期刊。祈能藉由其他領域技術的介紹並結合國內現況與資料激盪出適合的應用方向。以下將分不同主題提出幾個應用方向。第二章首先介紹物件偵測的現有技術，包括影像偵測、重疊物件、重識別與聲音偵測。在第三章提出基於第二章所提之技術於交通之應用。最後在第四章總結本文所提之可能方向。

物件偵測技術

目前物件偵測所使用的資料來源主要為影像、聲音和點雲。影像的來源主要是道路監視器所拍攝的影片、路側燈桿架設之攝影機、空拍機、行車紀錄器以及公開的資料集（dataset）；聲音主要由手機內建麥克風、路側或是燈桿架設之錄音設備偵測；點雲需要使用光達機收集。有了資料後方可訓練模型來學習特徵，進而讓模型從輸入之資料擷取關鍵特徵，進而自動偵測物件。

現有影像偵測技術

交通部運輸研究所之「應用人工智慧（AI）影像辨識技術進行交通數據蒐集與分析之研究^[10]」對於AI影像偵測技術應用在交通管理的現況作了回顧，也詳細描述了將偵測技術應用在國內場景的測試結果及其使用設備。

目前使用影像的物件偵測主要分為兩階段（two-stage）和單階段（one-stage）之模式。兩階段模式以R-CNN（region-based convolutional neural network）一系列為代表，其始祖R-CNN由Girshick等人^[11]於2014年CVPR（IEEE Conference on Computer Vision and Pattern Recognition）研討會提出，主要由三個部分組成，第一部分在一張圖片裡使用selective search方法產生多個類別獨立區域候選（category-independent region proposals），之後將這些邊界框（bounding box）交由第二部分的卷積神經網路（convolutional neural network, CNN）萃取特徵，再被輸入第三部分的線性支持向量機做分類，同時透過迴歸方法調整邊界框位置，最終找出圖片中特定位置的物體與其類別。其後也發展出一系列改進的方法，如Fast R-CNN^[12]將R-CNN每個候選框都必須輸入CNN的方式，改進為整張圖片輸入CNN再讓候選框各自取出特徵，也使用CNN內的softmax取代SVM作分類。Faster R-CNN^[13]透過Region Proposal Networks（RPN）取代selective search找出候選框，提升了搜尋候選框的速度，使得整體運算效率加快。其後也發展出Mask R-CNN^[14]可對圖像進行內容意義之分割（semantic segmentation）和實體分割（instance segmentation），對於物體的分割更加精準，可精細到像素級別。

單階段演算法則是以YOLO（you only look once）

系列模式為代表，由 Redmon 等人^[15]發表於 2016 年 CVPR 研討會。YOLO 將物件偵測視為單一迴歸問題，首先將一圖片切割成個格子 (grid)，對每個格子使用 CNN 預測邊界框以及類別信心值 (confidence)，最後整合成類別機率地圖 (class probability map) 做出最後預測。YOLOv4^[16] 為 YOLO 系列發展至第四代之方法，改進了訓練資料、訓練方法以及模型架構，與前幾代相比不僅提升了速度也提高了精準度，精準度甚至超越一些兩階段方法。

以上所提及兩類模式各有優劣。一般來說，在辨識速度上因為兩階段方法需先找出候選再進行辨識，所以速度會比單階段方法慢；但相對的精準度會比單階段方法高。隨著技術發展，爾後也出現單階段方法精準度贏過較老舊兩階段方法之情形。圖 1 將原圖與 Mask R-CNN、YOLOv3 及 YOLOv4 所產生之偵測結果作比較，Mask R-CNN 能夠較準確的圈出物件的輪廓並給予類別，同類別不同物件也會給予不同顏色表示；比起 YOLOv3，YOLOv4 偵測出更多類別，如交通號誌、植栽等，但也會發生偵測錯誤的情形，將建築物燈光誤認為號誌。圖 2 同樣是對三個模型作比較，左下角的人群中實際上有六人，因重疊緣故，YOLOv3 將六人成功圈選，YOLOv4 卻只偵測到五人。

重疊物件偵測

如「現有影像偵測技術」所述，一般來說基於候選的偵測方法精準度較高，但高度重疊物件的偵測仍然是一大挑戰，其困難之處主要在於高度重疊物件之間可能十分相似，以及高度重疊物件可能被非極大值抑制 (Non-maximum Suppression, NMS) 錯誤抑制，導致其中一個或多個物件未被偵測出。

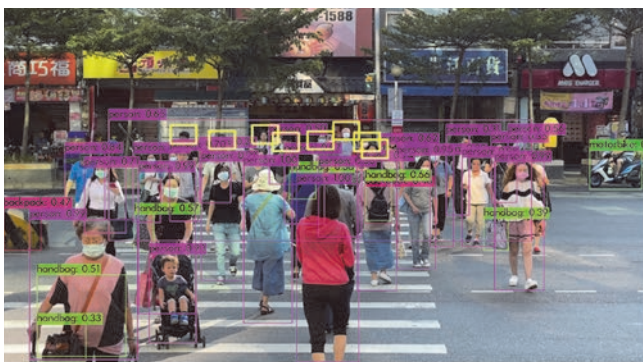
Chu^[17] 等人提出了一個新的偵測方法以解決上述兩個偵測物件重疊時所面臨的問題，且於重疊較少的情況下也可保持良好的偵測效果，此方法適用於所有具有候選框的偵測模型：首先，將候選與真值 (Ground Truth) 的一對一關係改為一對多關係，即由一個候選框預測一個實例 (Instance) 改為預測多個實例；並在原始 NMS 演算法前插入一個檢查機制，判斷兩個邊界框是否為同一個候選所預測出，由同一個候選預測出的所有實例被預期是唯一的，若兩邊界框來自同一個候選，則不抑制該邊界框。最後，為防止預測物件數量增加可能造成的假陽性 (False Positive) 數量上升，可再選擇性地加入一個改進模組。此方法主要可用於改善高物件密度下的偵測效果，因此該篇研究採用 CrowdHuman^[18] 資料集作為訓練資料。CrowdHuman 為適用於行人偵測的資料集，有較高的平均行人密度。圖 3 為一高密度行人場景。



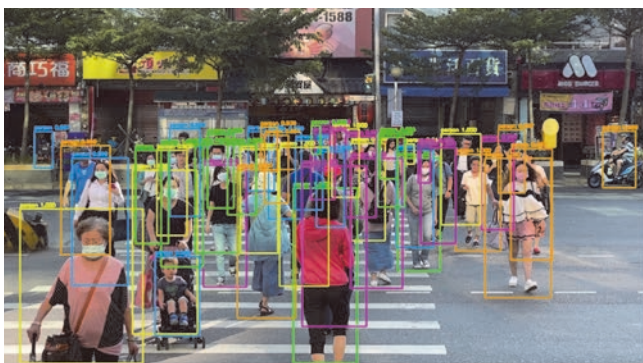
圖 1 左上角為原始圖片；右上為使用 Mask R-CNN 偵測結果；左下為 YOLOv3 偵測結果；右下為 YOLOv4 偵測結果。拍攝於臺北市信義路靠近永康街口



圖 2 左上角為原始圖片；右上為使用 Mask R-CNN 偵測結果；左下為 YOLOv3 偵測結果；右下為 YOLOv4 偵測結果。拍攝於臺南奇美博物館



(a) 使用 YOLOv3 偵測



(b) 使用重疊物件偵測模型

圖 3 YOLOv3 與重疊物件偵測模型於行人偵測之結果比較，上圖黃色方框為 YOLOv3 無法成功偵測行人頭部，為手動標示

重識別

以行人為例，重識別 (re-identification, Re-ID) 為不同攝影機捕捉 (capture) 之影像而建立關聯，可用來判斷不同攝影機捕捉到的是否為同一人並給予編號^[19]。同一攝影機拍攝到的同一人可能只有一個角度，如果結

合不同角度的攝影機，將可取得同一人不同角度的影像與特徵，模型經過這些影像做訓練後，便可學習到不同角度特徵之關聯性，從許多未分類影像找出同一人不同角度的一系列影像。重識別的挑戰之處在於實際場景中由不同攝影機拍攝到同一人影像外觀會有差異，造成不同之處有光線、解析度、阻擋物、角度與動作等，一般會假設同一人的衣著會是相同的，如果遇到跨日或衣著改變則又是另一挑戰。

Zhong 等人^[20]在行人重識別模型中加入三種底層的不變性 (underlying invariance)，讓模型學習到相同行人之間、不同相機間的更多特徵，使得模型的準確度顯著提升。重識別之應用不侷限於行人。傳統重識別提供一段影片作為目標 (query) 以用來從圖片集 (gallery) 中找出相同目標，此設定稱為「影像對影像」(video-to-video, V2V) 但實務上監視影像往往只有一張確定的圖片可當作追蹤目標。Porrello 等人^[21]使用以「圖片對影像」(image-to-video, I2V) 作為訓練目標的模型，結合兩層的「老師-學生架構」(teacher-student framework) 進行強化訓練，使得學生部分有效提升精準度。此模型也使用行人、車輛以及動物的資料及來進行測試。

由 NVIDIA 公司主辦之 AI City Challenge^[22] 其中挑戰之一便是車輛的重識別，該挑戰所提供之資料集包含了真實車輛由不同攝影機不同角度拍攝照片與電腦合成的影像。合成影像為解決資料數量不足的法之一，透過軟體建立車輛模型，再隨機將車輛模型以不同角度與不同背景圖片進行合成，透過此方法有效增加資料數量。此挑戰的評分方式為對每個目標只考慮前一百

個辨識結果的平均精準度取平均 (mean of the average precision, mAP)；如果排序後前一百個結果裡有越多正確圖片在前就表示此目標辨識精準度越高，最後再取所有目標辨識精準度的平均。

聲音偵測

聲音主要由三大要素構成：音色、音調、響度，在解析音訊時，由於聲音在短時間內相對穩定，通常會將原始的信號資料分割成較小的音框 (frame)，依據不同目的，提取合適的特徵，分析音框中音波之波形、頻率和振幅，例如：梅爾倒譜係數 (Mel-scale Frequency Cepstral Coefficients, MFCC) 主要應用於聲音識別領域。此外，感應器偵測到之音訊到達時間差 (Time-Difference-of-arrival, TDOA) 與都卜勒效應 (doppler effect)，皆可獲取聲源的位置與速度資訊。

關鍵的音訊特徵可以提升模型分類的準確度，Wiczorkowska 等人^[23] 使用路旁車輛的行駛聲，以及實驗室車輛測試的錄音為資料，篩選出適合車種分類的重要特徵，提高車種分類的準確率。除此之外，聲音的特徵也可以獲取基本車流資訊，如：Uchino 等人^[24] 於雙向四車道且車流量大的道路環境，架設二個麥克風，記錄行經車輛的聲音時間與強度，用以處理車輛連續與同時經過偵測點之準確率下降問題；Kubera 等人^[25] 在不同氣候條件下，使用立體聲麥克風偵測單一車輛行駛之速度變化，透過分析聲譜圖的特徵，分類目標車輛為加速、減速或固定速度行駛，其模型準確率高達 90%；Gatto 和 Forster^[26] 在觀測道路上方架設偵測器，透過機器學習模型分類整體車流環境聲音的特徵，以預測道路為壅塞或是自由車流狀態。

然而，聲音和影像時常成對出現，現實生活中，透過視覺與聽覺的結合，我們能更準確地辨認與定位物體，對於人工智慧之模型訓練上，亦是如此。以往在人工智慧相關應用的資料收集時，多著重影像處理，較少研究將聲音的資訊納入考量。因此，若能善加運用影像與聲音的特徵，便能更完整地提供所需資訊。Zhao 等人^[27] 建立一整合影像分析、聲音分析與聲音合成之模型，使用樂器演奏影片作為訓練資料，可為聲音做定位，找出影片中每個像素 (pixel) 發出的聲音，結果顯示可定位不同樂器合奏在影像中的發聲位置。Qian 等人^[28] 提出融合影像和音訊的二階段的模型，分析多物件多聲源的影片，包含結合影像與聲音的特徵，分類影片中的物件，以及配對各物件並定位音源，最後以音源熱

力圖之視覺化方式呈現。此外，Tian 等人^[29] 對同一資料集進行物件分類，該研究分別使用僅有影像、聲音，以及影像與聲音特徵皆有的模型進行訓練，其結果顯示影像與聲音特徵的結合能提升分類的準確率，並且聲音的特徵還能夠輔助定位影像中的物件。

吳沛儒等人^[7] 在其運用影像辨識車流的研究中提到，造成模型誤判的原因有機車體積較小較難偵測、角度造成遮擋、水中倒影、白天及雨天準確率較低與特種車輛準確度差。前四項皆是因為影像的限制所造成，因為攝影機的角度會出現死角與遮擋，環境光線與道路擁擠程度會使模型判斷失準；最後一項為訓練資料所導致，給予模型的訓練資料本身特種車輛佔比低，造成模型對於這些車輛得辨識能力較差。跟國外所開發模型相比，國內因機車占比高的混合車流特性也會是關鍵之分析因素之一。如果加上聲音作為偵測時的判斷來源之一，可在一定程度上提升精準度，使模型更加可靠。

聲音的偵測較不易受環境限制，且設備成本與分析時所需資源相對較低，但容易受到噪音干擾；因此，在交通領域上，聲音可以應用於難以設置攝影機之道路環境，如：山區、崖邊道路等；抑或是用來輔助原有的影像分析，像是協助能見度低的環境偵測車輛、還原肇事現場等，運用聲音的特徵獲取車流相關資訊，並在危險交通環境中給予示警，有機會提升用路安全。

小結

基於影像的偵測，如：一階段和兩階段方法，可以幫助我們從現有設備獲取的影片中偵測物體，也已經可以達到即時運算偵測。有了影像偵測技術作為基礎，進而延伸出其他方法來幫助我們解決問題。重疊物件偵測技術用來解決非極大值抑制的錯誤抑制問題，使得改善高密度偵測場景的準確度。重識別技術結合影像辨識與特徵萃取，可從大量未分類圖片中找出屬於同一物體之影像。聲音偵測技術的發展既可作為影像辨識的輔助，提升準確度，也透過其獨特的特徵發展出一套不同於影像的辨識領域。以上技術可針對不同交通應用做結合，相輔相成，使智慧運輸更加可靠、應用層面更廣。

物件偵測於交通之應用

本章節延續前面所介紹之物件偵測技術，結合目前國內現有技術與應用，提出未來在國內智慧運輸之可能應用方向提供參考。首先介紹跨攝影機的物件追蹤，第二小節為高密度車流環境下之偵測，接著介紹

在夜晚或能見度低的環境下之物件偵測，第四節介紹高噪音機動車輛偵測，最後做本章小結。

跨攝影機追蹤

重識別技術特別適合應用在跨攝影機的追蹤與辨認。只要目標物離開單一攝影機，外觀就會因角度以及光線等產生改變，可透過重識別技術中的特徵擷取方法以及訓練來辨認不同角度的模型特色，用以協助達成跨攝影機追蹤。如果攝影機能夠同時收集聲音，將能利用聲音作輔助，進而提升辨識及追蹤之精確度。本小節以交通執法以及交通流量收集為例，提出跨攝影機追蹤的可能應用。

交通執法

目前國內交通科技執法的主要應用有：利用影像辨識對於車輛違停進行偵測，只要車輛出現在預先設置的範圍裡，便依照預先邏輯進行警告或辨識車牌自動告發；又或是利用攝影機偵測車牌作為區間測速、闖紅燈之舉發依據，缺點是攝影機角度必須能夠清楚擷取車牌或是設置專用的攝影機。警政署目前使用的 M-Police^[30] 系統可對民眾進行臉部辨識，在進行臨檢或盤查時可快速比對民眾身份，也可即時辨認出通緝犯。當目的為追蹤逃逸嫌犯時，因為攝影機解析度不足與角度問題便無法利用遠距離的監視器影像來辨識臉部，也因為跨監視器問題必須人工調閱監視器進行路線追蹤。此時重識別技術與聲音辨識便可利用少數曾出現的影像來自動搜索其他監視器影像，自動列出一系列影像，重現逃逸路線與目的地，將能大幅增加員警的搜索效率。

交通流量收集

現行高速公路透過無線射頻技術（radio frequency identification, RFID）感應車上的 eTag 作為收費憑藉，將感應門架設置在交流道附近以取得進出高速公路的路徑，同時收集車輛旅次資料。相對於高速公路的封閉性，一般道路之路口數量極多，若要將 eTag 偵測器或傳統車輛偵測器普及於所有路口、路段將面對較高之成本。目前作法主要都將 eTag 偵測器設置在市區內高架道路、橋樑等出入口，作為收集旅次時間及速度之用。如能利用現有道路監視器之影像資料，結合重識別技術，將可大大增加偵測範圍，且道路監視影像不限於路口，可獲得更精細資料。

國外大部分訓練影像、聲音都僅有行人或是小客車，建議可使用較貼近國內環境的訓練資料，如：混合了大小車、機車與機車騎士的影像、聲音資料，來提昇

模型的可靠度和精準度。但過度使用科技也可能引發民眾對於資訊安全、隱私權以及可靠度的疑慮，如何在智慧化與民眾權益之間取得平衡為需要優先考慮的問題。

高密度車流環境偵測

一般非擁擠情況的場景對於前章所提之影像偵測技術大部分皆可勝任，然而擁擠的場景使得物件重疊而降低準確度。使用較低攝影機角度之圖片進行汽機車偵測時容易因重疊而出現偏誤。

Wong^[31] 等人在都市幹道混合交通流的交通特徵研究中提到台灣人均機車持有率居世界之冠；而由於機車體積較小，靈活性高，機車駕駛人通常具有行車紀律（Lane Discipline）較弱的傾向。由於台灣機車數量眾多，且機車駕駛人多習慣穿梭於汽車間，在車流密度較高的環境下，若使用影像作為車輛偵測資料來源，機車受到遮擋的情形較嚴重，容易導致機車遺漏偵測，因此重疊物件偵測技術的改進有助於提升都市道路機車偵測的品質。

圖 4 為本研究將「重疊物件偵測」介紹之重疊物件偵測模型用於機車偵測的成果，(a)、(b) 分別為高密度機車場景與低密度機車場景。此測試成果與行人偵測成果同樣使用 CrowdHuman 資料集作為訓練資料，因此實際為檢測機車駕駛人。由於高密度機車流在許多國家並不常見，建議使用國內的高密度機車影像作為訓練資料，以提升模型在國內交通環境的適用性。



(a) 高密度機車場景



(b) 低密度機車場景

圖 4 重疊物件偵測模型於機車偵測之測試成果

低能見度環境下之偵測

現今人工智慧於影像辨識發展成熟，國內對於影像解析交通環境的投入越來越多，目前普遍仰賴空拍機以及攝影機進行資料蒐集，並在能見度高的環境中，擁有良好的績效。然而，相較於影像辨識與其他偵測設備，聲音辨析擁有設備成本與運算資源較低的優勢，並且聲音對於外在環境極敏感，可偵測周遭細微的變化，同時也能收集大範圍的音訊，用以分析巨觀環境；除此之外，聲音可以根據不同的問題，萃取相對應的特徵，再與影像同步，且不易受能見度影響^[32]。Gan 等人^[33]使用影像偵測影片中車輛位置結合聲音進行訓練，可利用音訊檔和為標記影片作為模型輸入，輸出利用聲音定位圈選出車輛位置

的影像。文中提到使用聲音作為定位可解決影像光線不足場景的限制，但仍有可辨識車輛數較少的技術限制。Vasudevan 等人^[34]運用四對立體聲麥克風、廣角攝影機收集全景影像與聲音的特徵，結合影像之深度資訊，以定位與分類影像中的機車、汽車、電車；此外，此模型在雨天、霧天、夜晚的場景，皆有良好的表現。本研究運用 Tian 等人^[29]影像結合音訊之定位模型，分別套用國內夜晚與雨天情景，偵測機車與汽車，其結果如圖 5 到圖 7。

至今同時運用影像與聲音特徵於交通領域的文獻較少，故本研究認為聲音的偵測可以彌補其他感測器的限制，或是以聲音輔佐影像之分析，皆能為智慧運輸帶來貢獻。



圖 5 低能見度環境之車輛偵測結果 (2020/9/13 午，雨天)



圖 6 低能見度環境之車輛偵測結果 (2020/9/13 晚)



圖 7 低能見度環境之機車偵測結果 (2020/9/16 晚)

高噪音機動車輛偵測

前述之影像結合音訊定位模型除了可用於輔助其他類型感測器以提升偵測精度外，也適合與車輛噪音量測技術相結合，以下就噪音量測科技執法與車輛噪音影響因子研究的角度提出可能的應用。

噪音量測科技執法

以往國內機動車輛噪音稽查以路邊攔檢為主，其缺點為耗費人力資源，且存在攔檢噪音非屬車輛實際行進時噪音之爭議^[35]。目前環保署預告即將實施的機動車輛聲音照相科技執法計畫中，使用陣列式聲音感應器、攝影機、風速計及噪音計組成影像輔助噪音量測的設備，陣列式聲音感應器用於聲源定位，攝影機記錄車輛車牌^[36]；由於設備昂貴，不易全台普及使用。而影像結合音訊定位模型只需以普通攝影機的單聲道或雙聲道影片作為輸入資料，若能將之應用於高噪音車輛偵測，則可使用攝影機記錄的資料同時達成聲源定位及車牌辨識，降低設備成本，有助於將噪音量測科技執法擴大執行。

道路設計與車輛噪音關係探討

城市道路設計對於交通噪音的嚴重程度有所影響，Lu^[37]等人探討了道路特徵對交通噪音的影響，如車道數、路段長度以及路口是否有交通號誌等，在路段中段設置噪音測量點，以路段為單位研究道路與車輛噪音的關係。過往研究多集中於路網設計的層面，若可運用影像結合音訊定位，或有助於探討道路線型設計對車輛噪音大小的影響，偵測路段中可能造成較大噪音的位置。

小結

智慧運輸之應用，會遭遇資料收集、模型轉移適配以及結果與理想差距等問題。運用交通領域之核心知識將問題分別拆解後，較可能提供交通領域之獨特應用；從個別問題找尋最適模型出發，可能依照不同模式之組合與融合而找出新的應用模式與方向，惟須考慮應用之實際性及在地化考量，如國外開發之模型較少考慮混合車流問題，故須經過一番調整方能移植與應用。透過加值方法來提升準確度與可靠度為一重要議題。第二章所提之模型也不僅限於本章所提及之應用，期望本章能啟發更多可能的智慧化交通應用。

結論

隨著科技不斷進步，更精準及可靠的模型紛紛被提出，從模型的出現到被轉化為實際應用之間需要不小的努力。政府、產業界與學術界的合作以及適合之技術引進與開發，讓用路人的每個旅次環境更安全更有效率。本文對國外近年發表之尖端技術進行介紹，並提出應用方向作為未來可能的參考。

本文所提及影像辨識一階段與兩階段模式僅分別以 R-CNN 系列與 YOLO 系列模型做代表，但實際技術種類繁多，並不僅限於此。聲音辨識技術的出現大大改變了「眼見為憑」的傳統影像資料來源，以聲音作物體定位，也能辨識出車流的速度與擁擠程度，聲音與影像的互相配合之下，大大提升了人工智慧模型應用在交通的可能性，也一定程度降低了偵測器的架設成本與設置困難問題。重疊物件偵測為解決了傳統偵測方法的難題，可能解決國內高密度車流因重疊而無

法準確辨識的挑戰。重識別技術則為之後跨攝影機追蹤應用立下基石，讓科技執法與流量收集等應用更上一層樓。雨天、夜晚及濃霧等低能見度場景的辨識可經由聲音與影像的相互配合來逐漸克服。結合聲音與影像的偵測技術也有助於偵測高噪音車輛。

建議在引進新技術時需考慮原始模型訓練所使用之資料及其原始應用場景。考量國內特殊混合車流之特性，必要時進行調整與自行客製化及開發，使成果更能貼近本國需求。期望透過本文介紹能激發更多可能之智慧運輸應用。

參考文獻

- 交通部 (2016)。智慧運輸系統發展建設計畫 (106-109) 年。
- 交通部 (2020)。智慧運輸系統發展建設計畫 (110-113) 年。
- 交通部運輸研究所 (2019)。交通事件資訊整合服務與精進計畫 (1/2)。
- 卓訓榮、黃銘崇、李怡穎 (2016)。高速公路整合控制之啟動時機研究。運輸學刊，第 28 卷，第 2 期，頁 149-186。
- 鍾易詩、邱裕鈞、謝志偉、張開國、葉祖宏、田養民、陳凱斌 (2014)。國道高速公路交通事故持續時間分析與推估：脆弱性存活模型之應用。運輸學刊，第 26 卷，第 4 期，頁 555-578。
- 黃柏宗 (2019)。道路交通之事故辨識與碰撞機率估計：SVM 與 Random Forests 之應用。成功大學交通管理科學研究所碩士論文。
- 吳沛儒、陳其華、蘇昭銘、吳東凌、黃啟倡、鍾俊魁、何毓芬 (2019)。人工智慧之交通事件影像偵測模式與實域驗證。運輸計劃季刊，第 48 卷，第 3 期，頁 159-178。
- 邱裕鈞、謝志偉 (2012)。混合車流格位傳遞模式之建立與驗證。運輸學刊，第 24 卷，第 2 期，頁 245-276。
- 魏健宏、陳奕志 (2001)。類神經網路模式在國內交通運輸研究之成果評析。運輸計劃季刊，第 30 卷，第 2 期，頁 323-347。
- 交通部運輸研究所 (2020)。應用人工智慧 (AI) 影像辨識技術進行交通數據蒐集與分析之研究。
- Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). *Rich feature hierarchies for accurate object detection and semantic segmentation*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- Girshick, R. (2015). *Fast R-CNN*. IEEE International Conference on Computer Vision (ICCV), pp. 1440-1448.
- Ren, S., He, K., Girshick, R. and Sun, J. (2015). *Faster R-CNN: towards real-time object detection with region proposal networks*. Advances in Neural Information Processing Systems 28 (NIPS 2015).
- He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2017). *Mask R-CNN*. IEEE International Conference on Computer Vision (ICCV), pp. 2961-2969.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi A. (2016). *You only look once: Unified, real-time object detection*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). *YOLOv4: Optimal Speed and Accuracy of Object Detection*. arXiv:2004.10934.
- Chu, X., Zheng, A., Zhang, X., and Sun, J. (2020). *Detection in crowded scenes: One proposal, multiple predictions*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12214-12223.
- Shao, S., Zhao, Z., Li, B., Xiao, T., Yu, G., Zhang, X., and Sun, J. (2018). *CrowdHuman: A Benchmark for Detecting Human in a Crowd*. arXiv:1805.00123.
- Bedagkar-Gala, A. and K.Shah, S.K. (2014). A survey of approaches and trends in person re-identification. *Image and Vision Computing*, 32(4), pp. 270-286.
- Zhong, Z., Zheng, L., Luo, Z., Li, S., and Yang, Y. (2019). *Invariance matters: Exemplar memory for domain adaptive person re-identification*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 598-607.
- Porrello, A., Bergamini, L., and Calderara, S. (2020). *Robust re-identification by multiple views knowledge distillation*. The European Conference on Computer Vision (ECCV).
- Naphade, M., Wang, S., Anastasiu, D.C., Tang, Z., Chang, M., Yang, X., Zheng, L., Sharma, A., Chellappa, R., and Chakraborty, P. (2020). *The 4th AI city challenge*. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 2665-2674.
- Wieczorkowska, A., Kubera, E., Słowik, T., and Skrzypiec, K. (2018). Spectral features for audio based vehicle and engine classification. *Journal of Intelligent Information Systems*, 50(2), pp. 265-290.
- Uchino, M., Dawton, B., Hori, Y., Ishida, S., Tagashira, S., Arakawa, Y., and Fukuda, A. (2020). *Initial design of two-stage acoustic vehicle detection system for high traffic roads*. IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pp. 1-6.
- Kubera, E., Wieczorkowska, A., Kuranc, A., and Słowik, T. (2019). Discovering speed changes of vehicles from audio data. *Sensors*, 19(14).
- Gatto, R.C. and Forster, C.H.Q. (2020). Audio-based machine learning model for traffic congestion detection. *IEEE Transactions on Intelligent Transportation Systems*, 22(11), pp. 7200-7207
- Zhao, H., Gan, C., Rouditchenko, A., Vondrick, C., McDermott, J., and Torralba, A. (2018). *The sound of pixels*. The European Conference on Computer Vision (ECCV).
- Qian, R., Hu, D., Dinkel, H., Wu, M., Xu, N., and Lin, W. (2020). *Multiple sound sources localization from coarse to fine*. Proceedings of the European Conference on Computer Vision (ECCV).
- Tian, Y., Shi, J., Li, B., Duan, Z., and Xu, C. (2018). *Audio-visual event localization in unconstrained videos*. Proceedings of the European Conference on Computer Vision (ECCV), pp. 247-263.
- 內政部警政署 (2020)。內政部警政署全球資訊網。https://www.npa.gov.tw/NPAGip/wSite/
- Wong, K.I., Lee, T.-C., and Chen, Y.-Y. (2016). Traffic characteristics of mixed traffic flows in urban arterials. *Asian Transport Studies*, 4(2), pp. 379-391.
- Bai, Y., Lu, L., Cheng, J., Liu, J., Chen, Y., and Yu, J. (2020). Acoustic-based sensing and applications: A survey. *Computer Networks*, 181.
- Gan, C., Zhao, H., Chen, P., Cox, D., and Torralba, A. (2018). *Self-supervised moving vehicle tracking with stereo sound*. IEEE/CVF International Conference on Computer Vision (ICCV).
- Vasudevan, A. B., Dai, D. and Van Gool, L. (2020). *Semantic object prediction and spatial sound super-resolution with binaural sounds*. arXiv: 2003.04210.
- 行政院環境保護署空保處 (2018)。107 年度「偵測照相系統應用於使中機動車輛噪音管制評估之研究」專案研究計畫。
- 行政院環境保護署 (2020)。機動車輛噪音量測方法—以影像輔助方式草案。
- Lu, X., Kang, J., Zhu, P., Cai, J., Guo, F., and Zhang, Y. (2019). Influence of urban road characteristics on traffic noise. *Transportation Research Part D: Transport and Environment*, 75, pp. 136-155. 