

# 中華民國第 65 屆中小學科學展覽會

## 作品說明書

國中組 生活與應用科學科(一)

第一名

032819

十字路口人車碰撞預測之研究

學校名稱： 國立臺灣師範大學附屬高級中學(附設國  
中)

作者：

國二 林佑謙

指導老師：

黃政嘉

李啟龍

關鍵詞： YOLO、LSTM、PET

## 得獎感言

### 科展心得

這次的科展，我有幸得名。除了要感謝評審的提拔。還需要感謝父母、老師的支持。

我從小就喜歡許多發明，從剛開始的硬體電風扇的接觸，到發球機的建造以及現在可以用Python訓練模型，這都要感謝父母的大力支援。

在科展題目還未完全確定方向時，我阿嬤被摩托車撞倒並且骨折，知道時，讓我十分傷心。我想要解決這個問題，並且改善用路人的安全。

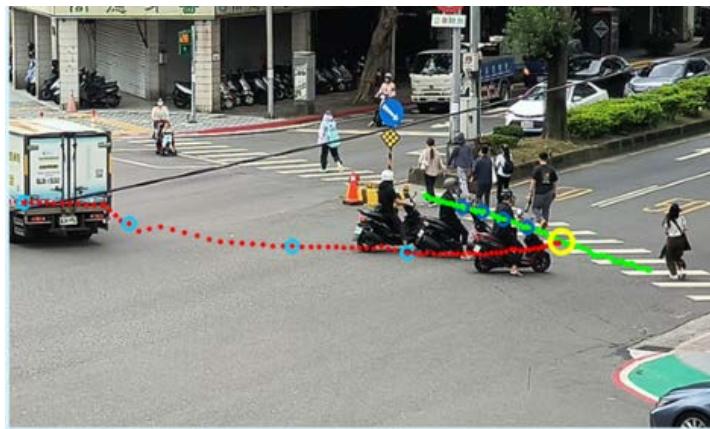
我先透過去十字路口拍攝影片進行，還發現公車司機因為駕駛過於疲勞而忽略報警器的種種案例，再到訓練YOLOV11模型、LSTM模型。一路上遇到了許多困難，在評審來的那一刻，也化為讓我可以與評審侃侃而談的燃料。

這次的科展讓我學到了許多東西，是如何與人打交道、是如何與生命、生命溝通、是如何跟電腦學習。

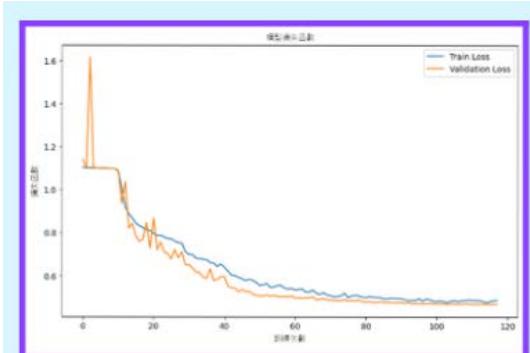
感謝這次科展所有的工作人員、評審、老師、家長的支持，才能有圓滿且具有價值的回憶。



混淆矩阵圖



實際路口拍攝



訓練準確率圖

準確率圖

## 摘要

本研究探討如何利用電腦視覺技術預測十字路口行人與車輛碰撞風險，提升交通安全，採用 YOLO 物件偵測與運動追蹤技術，分析監控影像，透過 LSTM 進行時間序列預測，評估可能的碰撞風險。研究先進行影像標註與資料處理，提升自訂物件偵測能力：偵測行人與車輛和追蹤其移動軌跡，建立LSTM預測分類模型，學習運動模式以預測碰撞風險。研究設計先用小球碰撞實驗模擬真實交通場景，訓練模型預測碰撞風險，並透過驗證分析調整參數，再進行實際路口影像處理。結果顯示能預測未來 2 秒碰撞風險的準確率達 73% 以上，還可以持續改進，提升精準度。未來，本研究可結合智慧交通管理與車聯網，提升道路安全並減少碰撞事故。

## 壹、前言

### 一、研究動機

台灣每年因為車禍死亡的人數竟然高達 3000 人，平均每天至少有一名行人因為交通事故喪生，這真是讓人感到震驚和難過。每當在新聞上看到行人被撞而死亡報導，心裡就湧上一股沉重的無力感。馬路上依然充滿危機，肇事的新聞還是層出不窮（馮惠宜，2024）。甚至連國外媒體都稱台灣為「行人地獄」，這不只是讓人覺得丟臉，更讓人感到憤怒和無奈。我一直在想，為什麼這個問題一直沒有改善？為什麼我們只是單純地走在路上，都還要提心吊膽，害怕下一秒可能會被車撞？

政府常說要加強交通安全，還會舉辦各種宣導活動，台北市把2024年定為「交通安全年」，市政府做了很多改善交通的措施，調整行人專用號誌，行穿線退縮和設立庇護島，讓大家過馬路更方便、安全。

科技不斷持續進步，從簡單影像處理，發展到現在的深度學習、自自動生成圖片和影片，電腦視覺技術的發展速度超乎想像。它已經滲透到我們的日常生活，從手機相機的 AI 優化，到自動駕駛、智慧醫療，每個領域都能看到它的影子。隨著 AI 的進步，電腦視覺將會變得更強，可能達到像人類一樣的視覺理解能力。身為學生，想透過自行研究的方式，學習電腦視覺技術，找到解決這個問題可能的方法。

## 二、研究目的

- (一) 使用電腦視覺技術（YOLO），進行行人與車輛的偵測與分類，建立穩定的物件辨識模型。
- (二) 運用影像運動追蹤與透視轉換技術，建立十字路口中車與人的運動軌跡資料。
- (三) 以後侵入時間（Post-Encroachment Time，PET）為風險指標，進行資料標註與分類，構建可監測潛在衝突的數據集。
- (四) 透過 LSTM 模型預測未來 2 秒內是否發生人車碰撞事件，提升道路使用者安全。
- (五) 驗證模型效能並提出可行改進建議，作為未來智慧交通應用基礎。

## 三、文獻回顧

### (一) YOLO

在電腦視覺的物件偵測中，傳統方法通常採取兩階段學習（Two-Stage Learning）的方式，首先選取可能的物件區域，然後對這些區域進行進一步的分類與辨識。然而，這種方法的缺點在於，當偵測出的候選物件過多時，計算成本會大幅增加，導致運算時間加長，並對硬體資源要求更高。這不僅提升了系統的建置費用，也降低了智慧型手機對該技術的實用性。

為了解決這個問題，一階段學習（One-Stage Learning）的架構應運而生。這個方法透過單一神經網路，同時完成物件定位與辨識，使運算速度大幅提升。儘管其辨識精度相較於兩階段方法稍有下降，整體辨識能力仍然保持在可接受範圍內。其中，YOLO（你只能看一次，You Only Look Once）便是一階段學習方法中最具代表性的模型之一。

YOLO 的核心概念是將輸入影像劃分為多個網格，每個網格負責預測數個物件的邊界框以及對應的類別機率。每個邊界框會產生五個關鍵數值：(x 座標, y 座標, 寬 w, 高 h, 信心分數)，其中信心分數計算方式：

信心分數 = P 網格內是否有物件的機率 X

IoU 預測框與實際框重疊程度

表示當網格內無物件時，信心分數為零：若網格內存在物件，則信心分數會反映預測邊界框與實際物件之間的重疊程度（IOU, Intersection over Union）。YOLO 的設計不僅提升了運算效率，使其適用於即時應用場景，提升物件偵測的實用性。

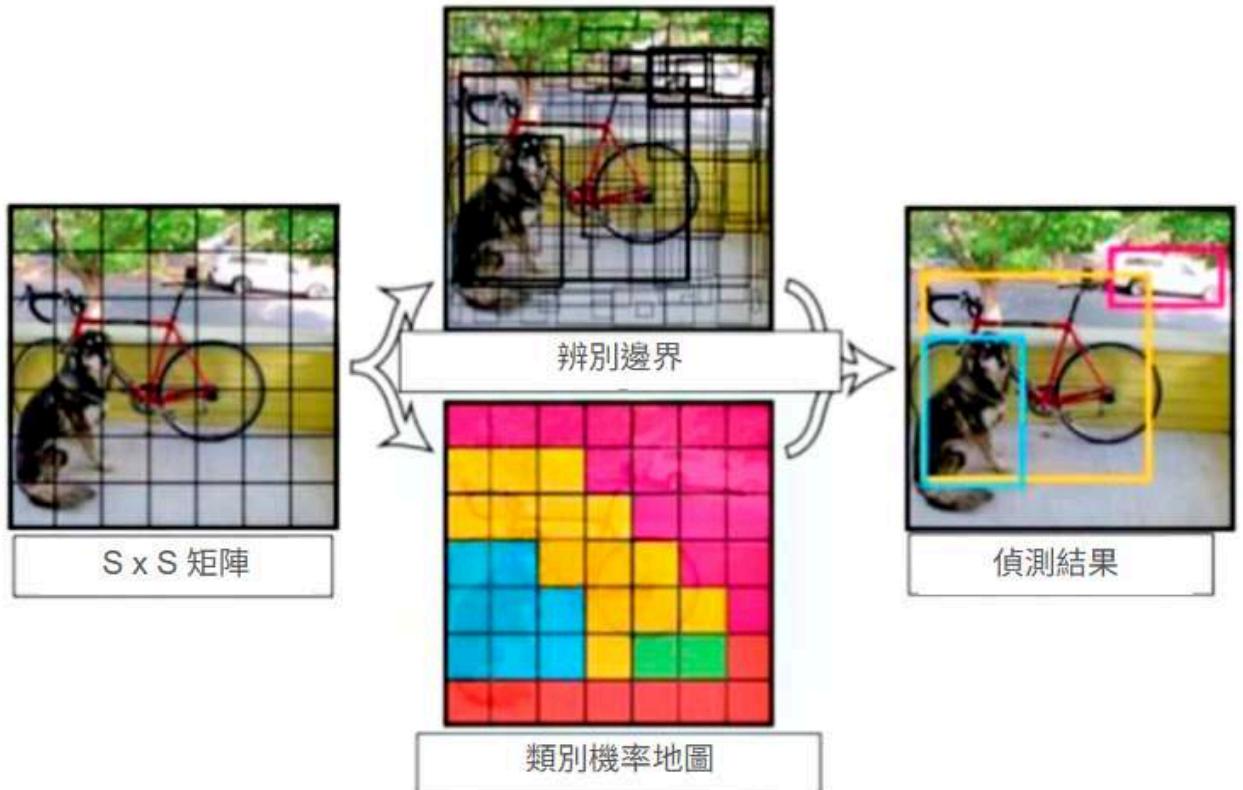


圖1-1：YOLO 可以一次辨識多種和多件物品和位置。圖片改編自Nikolic, N (2023)。

## (二) 長短期記憶網路 (LSTM)

長短期記憶網路 (LSTM) 是基於時間序列的模型 (Hochreiter & Schmidhuber, 1997)，它是一種改進版的循環神經網路 (RNN)，解決傳統 RNN 在訓練過程中容易出現的梯度消失與梯度爆炸問題。LSTM 能夠有效學習長期依賴資訊。主要內容存放在記憶單元，這是一條能夠長時間保存資訊的路徑，並透過遺忘門、輸入門與輸出門來控制資訊的存取。遺忘門決定哪些舊資訊應該被丟棄，輸入門負責選擇應該加入記憶的新資訊，而輸出門則決定最終輸出的內容。這樣的結構使得 LSTM 在處理長時間序列數據時表現更好，適用於自然語言處理、語音識別與時間序列預測等應用。相較於傳統 RNN，

LSTM 能夠有效記住遠距離的資訊，並透過閘門機制靈活調整記憶，避免梯度消失問題，進而提升模型的學習能力與準確性。

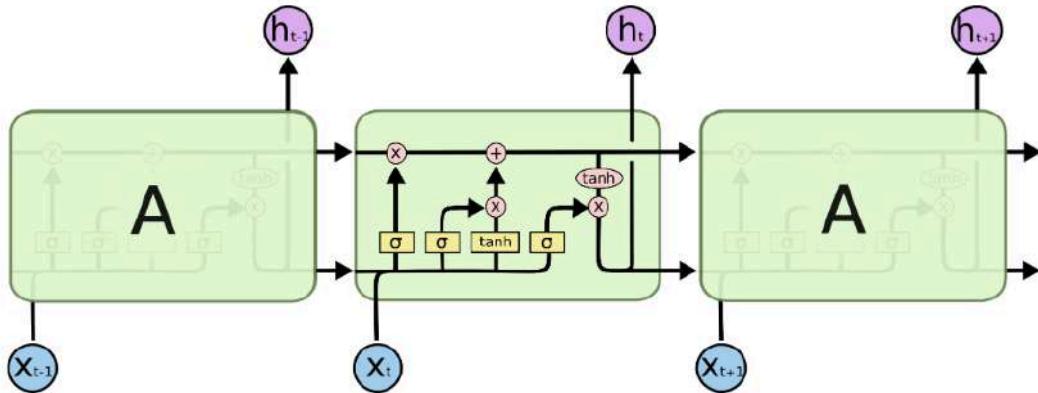


圖1-2：LSTM 輸出和輸入之間的關係。圖片來源：Understanding LSTM Networks (Christopher O. , 2015) ，

### （三）影像運動追蹤 BoT-SORT

BoT-SORT 是一種多個目標追蹤技術 (Aharon, N. et al. , 2022) ，它在監視攝影機、交通監測和運動比賽分析等應用特別出色。它比先前的方法更能追蹤被遮擋的物體、保持移動物體編號 ID不亂跳、長時間穩定追蹤，同時還保留了 ByteTrack 的分配 ID方式，先分配高度信心結果，再考慮低度信心的物件。這個系統先用 YOLO 偵測目標，再用改良版的卡爾曼濾波器來預測移動，在快動作場景下依然能保持追蹤準確性，如比賽中的球或球員、或高速公路車輛監控，同時透過修正相機晃動來讓追蹤更準確。接著分成兩個階段來配對物件——先用 IoU 和 ReID 確保高度信心分配，再用 ByteTrack 方式補充其他目標，最後還會利用線性內插來維持 ID連續性。簡單來說，BoT-SORT就是讓電腦變得更會「認人」和「記住人」，即使人在畫面裡消失一陣子，也能再次找回來。

### （四）路口碰撞研究

使用機器學習預防和分析路口碰撞方面，有不少研究，例如《數據不平衡下以機器學習方法預測交通事故嚴重性之分析》（胡大瀛和李岳洪，2022）研究中發現，訓練模型需要大量數據，但交通事故資料常存在數據不平衡的問題。因此收集臺南市的公開資料，並使用過採樣（SMOTE）（Chawla, N. V. , 2002）和欠採樣來處理數據不平衡的情況，得到不錯的結果。

應用空拍影像與交通衝突分析技術，辨識路口問題並提出改善建議以提升安全。溫基信，2021）。研究駕駛行為與事故風險預測（張斯涵，2022），透過統計預測與機器學習技術，從駕駛行為數據中預測事故風險。研究發現，駕駛行為特徵與事故風險之間存在顯著關聯，機器學習模型能有效預測事故風險。在強化學習在交通信號控制中的應用（Du, W. et al. , 2022），採用強化學習方法來達成實現無碰撞的控制交通信號。該方法將道路安全標準納入強化學習模型，確保路口安全性，同時提高交通流量效率。

強化學習應用於路口談判行為學習：研究利用深度Q學習讓自動駕駛車輛在路口與其他車輛進行談判行為學習。結果顯示，所學得的策略能在98%的情況下安全通過路口，同時避免過於被動的行為。深度學習預測車輛碰撞風險研究中（謝辰陽，2020），使用行車紀錄器影片，並透過卷積神經網路（CNN）和長短期記憶網路（LSTM）來建構模型，預測未來3秒內是否會發生車禍。CNN可以分析影像的空間特徵，而LSTM則能學習時間序列模式，讓預測更準確。

## 貳、研究設備及器材

### 一、軟體環境：

1. Google Colab ( GPU T4 )
2. Visual Studio Code + Anaconda
3. Python + Pandas + Keras + OpenCV

### 二、硬體環境：

1. Google Colab上的虛擬機
2. ASUS 桌機 ( GPU RTX4080 )

## 參、研究過程

### 一、研究架構

從確定研究主題到分析改進，整個過程包括幾個階段。首先，明確研究問題，並學習與電腦視覺相關的知識，了解別人對路口安全的研究成果以確定方向。接著，收集路口影片作為影像處理的基礎，並使用 Python 程式進行資料處理與分析。如果涉及人工智能，需用整理後的資料訓練模型以解決問題。最後，分析結果並改進研究，逐步將腦海中的想法轉化為具體成果。

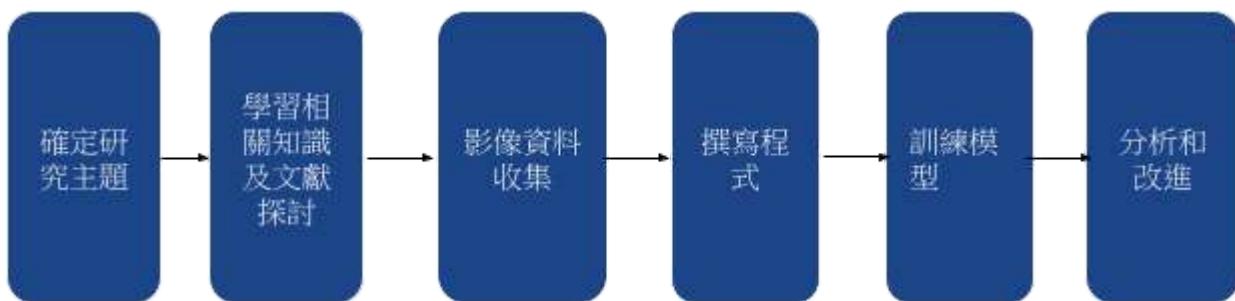


圖3-1：研究架構。圖片來源：研究者自製。

### 二、研究流程

研究流程圖說明分析十字路口影像的流程。首先，會錄製十字路口的影片，並進行放大與裁切，用以確保影像清晰適合分析。接著人工標註影片中的物件，包含車輛與行人，並調整影像亮度對比，使其適合訓練電腦。透過 YOLO 影像視覺技術，讓電腦學會辨識影像中的物件。當訓練完成後，電腦就能自動分析新錄製影片，偵測並追蹤物件的移動路線，投射到地圖上的座標，並整理數據以幫助後續預測碰撞的模型學習。最後檢視分析結果，評估準確率與 F1 分數，並回測影片。若準確度不足，則需調整訓練方式，讓系統能更準確辨識十字路口影像。

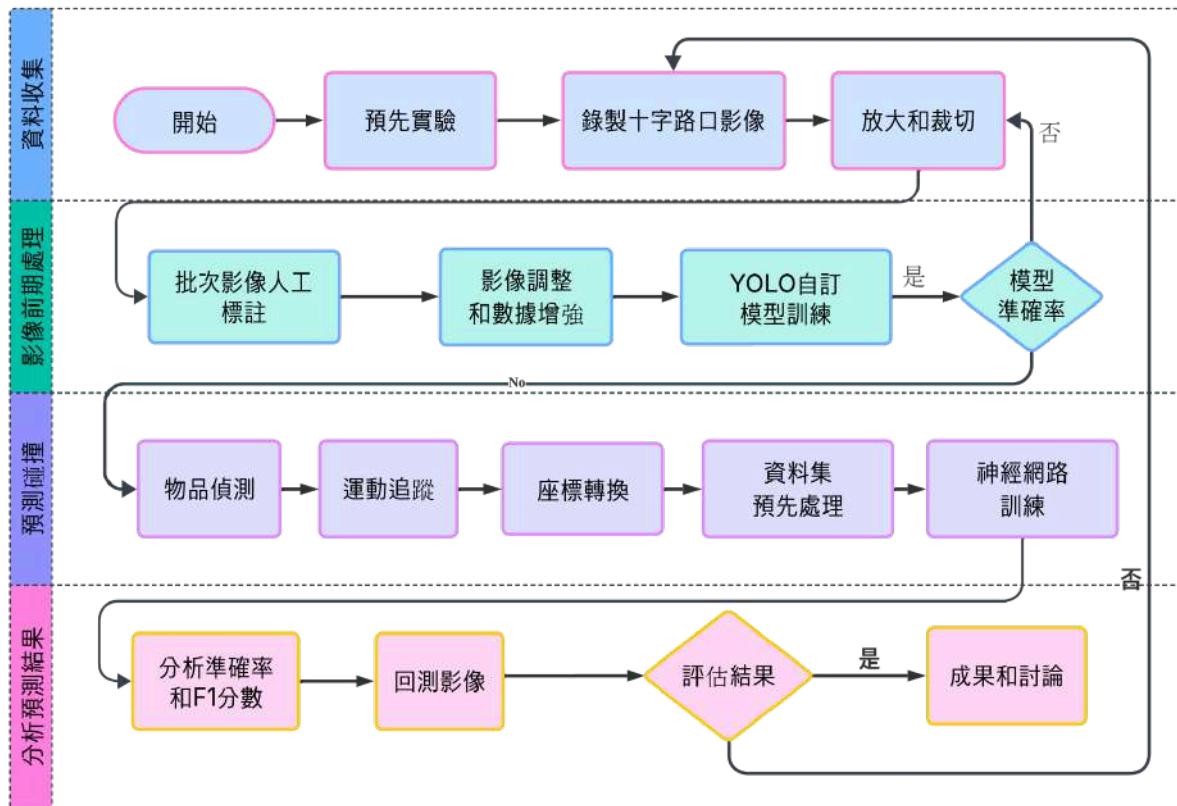


圖3-2：研究流程圖。圖片來源：研究者自製。

### (一) 預先實驗

為了模擬真實世界中物體的運動和交互作用，設計了一個簡單的模擬實驗。實驗中，在 500x500 的虛擬畫面上模擬兩顆小球的運動：一顆「快速球」和一顆「慢速球」。小球的初始位置、顏色和速度都是隨機生成的，如同現實世界中車輛和行人的不同速度和位置。當小球撞到畫面的邊界（牆壁）時，會像真實世界中的物體一樣反彈回來，繼續移動。特別關注小球可能發生碰撞或交叉的時刻。在這些時刻，我會記錄下碰撞前3秒的運動軌跡，作為訓練機器學習模型的資料。這個模擬實驗可以幫助學習和預測物體的運動軌跡，並了解在不同情況下可能發生的碰撞事件。

準備小球碰撞的訓練資料，輸入資料碰撞前1到3秒間的四筆資料，所以是時間點是1、1.5、2、2.5秒，內容為兩個球的座標(ball1\_x, ball1\_y, ball2\_x, ball2\_y)，輸出是碰撞的情形，輸出2表示非常可能發生碰撞，靠近的時間差為0.5秒內，輸出1表示接近但未發生，靠近的時間差為1秒內，輸出0，表示離很遠。三種資料各取 3000 筆建立資料集，所以共有9000筆測試資料。80%為訓練數據 ( $X_{train}, y_{train}$ )，20%為測試數據 ( $X_{test}, y_{test}$ )。

再訓練LSTM神經網路來分類這些碰撞事件。最後訓練出一個能夠自動偵測碰撞風險的神經網路模型，準確率可以有80%，也就是可以預測一秒後的可能碰撞事件。

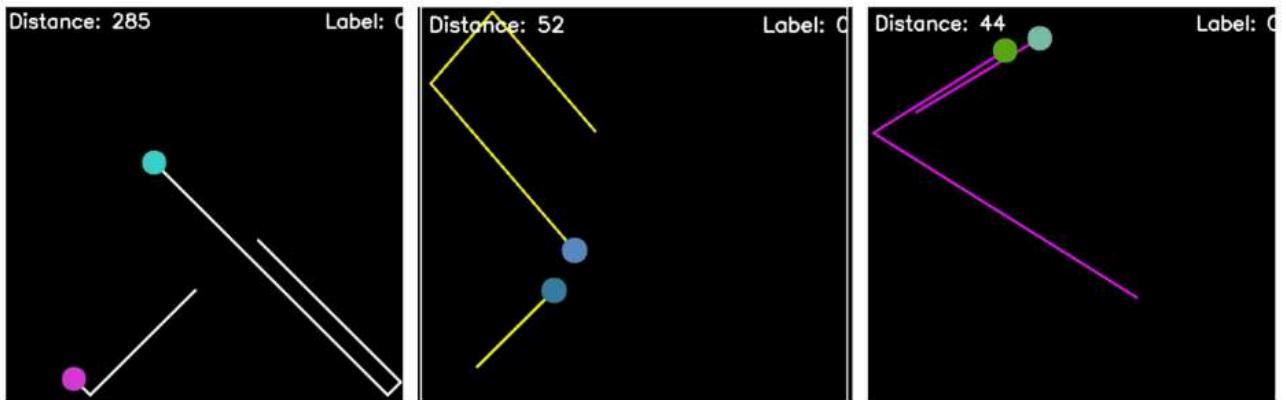


圖3-4：三種碰撞事件，左到右分別為安全、接近和非常可能。圖片來源：研究者自製。

預測小球碰撞模型訓練時準確率的折線圖，橫軸是訓練的訓練次數，縱軸是準確率。藍色曲線表示訓練準確率（Train Accuracy），橙色曲線表示驗證準確率（Validation Accuracy）。可以觀察到剛開始時，準確率隨著訓練進行快速提升，顯示模型正在有效學習。在約 20 次訓練後，準確率提升速度減緩並趨於平穩。驗證準確率與訓練準確率趨於一致：訓練大約 40 次後，兩條曲線逐漸靠近且重疊，這是好的現象，表明模型的過擬合現象（Overfitting）較小，具有良好的分類能力。訓練末期，訓練準確率和驗證準確率都接近 0.8，表示模型已達到較高的表現。驗證準確率穩定性：

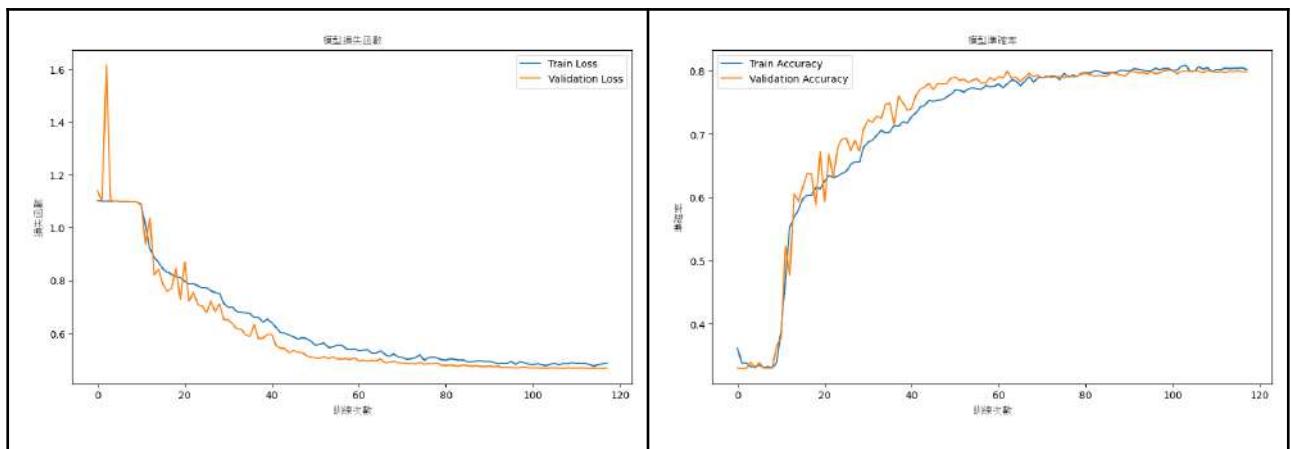


圖3-4：小球訓練時的損失函數和準確率。圖片來源：研究者自製。

## (二) 錄製十字路口影像

利用台北市公開資訊的即時交通資訊網，獲取光復南路與忠孝東路口的即時影像。由於每次影像播放的時間限制為 150 秒，因此採用人工，連續擷取影像並依序錄製，最終累積約 3 小時的交通影像，以便進行人和車輛安全分析。



圖3-5：錄制路口的位置是在大巨蛋的監視攝影機。圖片來源：臺北市即時交通資訊網

### (三) 批次人工標註影像

在錄製影像後，使用 Roboflow 網站的功能，學習來標註資料並匯出 YOLO 格式。開始時需要先在 Roboflow 建立一個專案，選擇「物件檢測」這個類型，然後把要標註的影像上傳上去。如果這些影像還沒標註，先批次手動標註一小部分，這些初步標註可以用來訓練一個基礎模型。接下來，匯出這些標註成 YOLO 格式，然後用 YOLO 模型進行訓練，得到一個初始的模型權重。接著就讓這個模型幫忙去自動標註剩下的影像。主動學習會挑出一些模型不太確定的部分（像是信心值很低的影像物體），再檢查和修正，這樣可以確保標註的準確性。在 Roboflow 裡，標註的介面很好用，可以很快就把錯誤的標註修正好。修正後的資料可以再匯出成 YOLO 格式，然後重新訓練模型，這樣做幾次，模型就會越來越準確。匯出資料集時，Roboflow 會幫你把影像和標註分開存好，標註檔會用 YOLO 格式記錄每個物件的類別和位置（像是 `<class_id> <x_center> <y_center> <width> <height>`），人工標註流程就是自動 + 人工合作，學習標註起來蠻容易的。

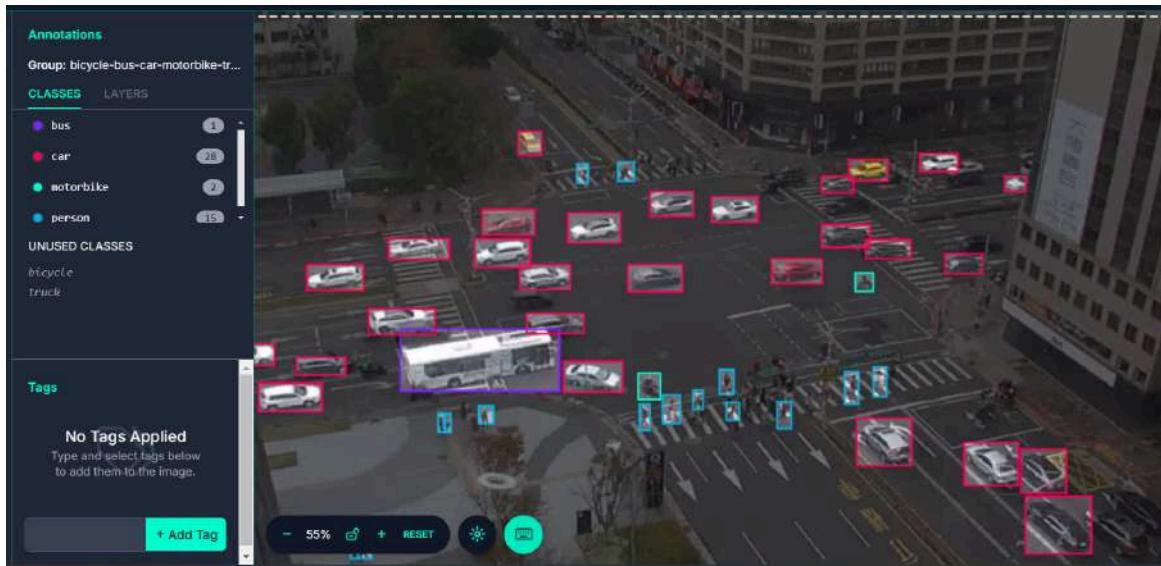


圖3-6：圖片標籤標註頁面。圖片來源：Roboflow網站

#### (四) 影像調整和數據增強

影像進行標準化處理以提升訓練效率與模型性能，調整大小至 1024x1024，使用自動均衡化對比度，再用數據增強隨機旋轉和隨機調整亮度等方式，提升資料多樣性來增強模型對路口不同天氣亮度辨識路上人和車的能力。

The screenshot shows the Roboflow data augmentation settings. It is divided into two main sections: 'Preprocessing' (step 3) and 'Augmentation' (step 4).

- Preprocessing (Step 3):**
  - Preprocessing:** What can preprocessing do? Decrease training time and increase performance by applying image transformations to all images in this dataset.
  - Resize:** Stretch to 1024x1024 (Edit button)
  - Auto-Adjust Contrast:** Using Adaptive Equalization (Edit button)
- Augmentation (Step 4):**
  - Augmentation:** What can augmentation do? Create new training examples for your model to learn from by generating augmented versions of each image in your training set.
  - Rotation:** Between -15° and +15° (Edit button)
  - Brightness:** Between -15% and +15% (Edit button)

圖3-7：資料集數據增強調整時的設定畫面。圖片來源：Roboflow網站

#### (五) YOLO自訂模型訓練

將數據集分為三部分，分別用於訓練、驗證和測試模型。因數據增強成為3倍，所以影像數量842張，訓練集（77%）用在模型學習特徵。驗證集（6%），用在調整訓練時的參數。測試集（15%）用在模型最後的評估。從roboflow下載時用YOLO V8的格式。

訓練過程如圖鄉3-8，展示了模型訓練過程中的損失與評估指標的變化，分為訓練和驗證的各項數據。以下是詳細 box\_loss 是邊界框的定位誤差，cls\_loss 衡量模型在分類目標上的錯誤程度，dfl\_loss 是與邊界框回歸相關的損失，評估邊界框預測的精細化程度，曲線下降，表明模型的框架定位更加精準。最後是在第51次得到最佳mAP50是0.607。

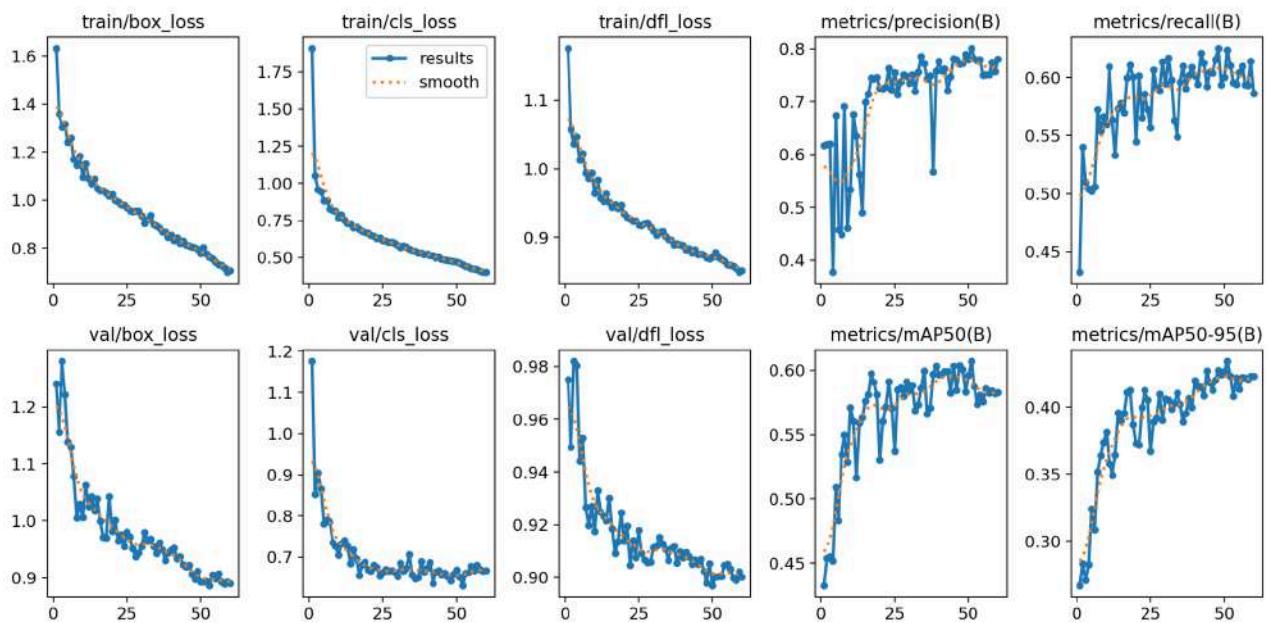


圖3-8：YOLO自訂訓練的過程，包括多種損失和評估指標。圖片來源：研究者自製。

訓練後產生混淆矩陣 (Confusion Matrix)，用於評估分類模型的預測性能。橫軸表示實際標籤 (True)，縱軸表示模型的預測標籤 (Predicted)。每個方格中的數字代表該類別的預測數量。深藍色方格（例如在 "car-car"）表示正確分類的數量最多，模型在這類的預測表現最好。淺藍色方格表示分類錯誤，錯誤的次數取決於顏色的深淺。右側的色階條 (colorbar) 表示數量的大小，顏色越深，數量越高。圖中可以模型在分類各種物體和背景方面的表現。模型在分類汽車方面表現最好，但在物體（特別是人）和背景之間，以及背景與其他物體之間存在顯著的混淆。根據顯示的數據，模型在自行車和卡車類別上的性能似乎非常弱，可能是資料集中這些物件較少的原因。

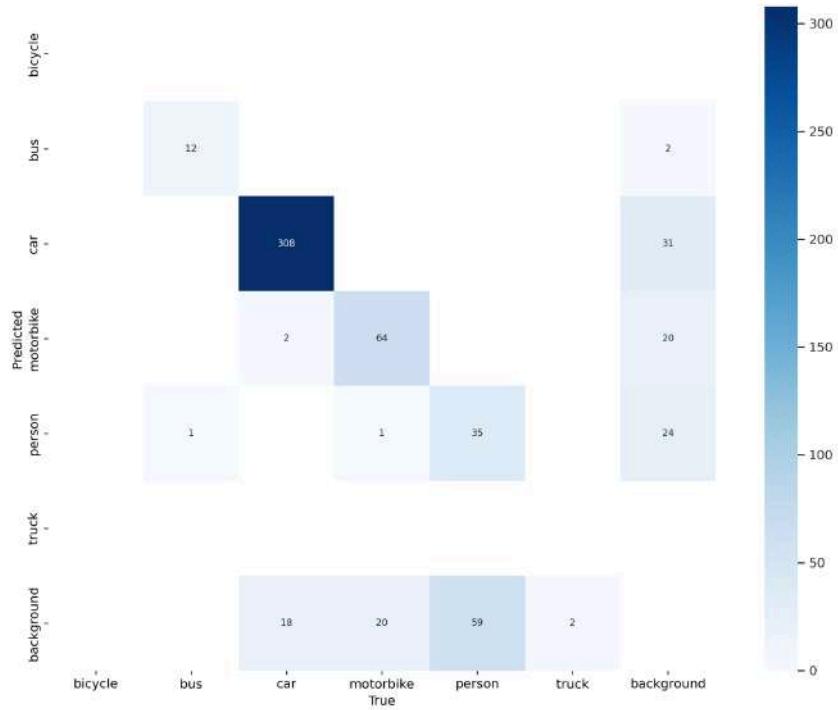


圖3-9：YOLO自訂訓練後的混淆矩陣。圖片來源：研究者自製。

## (六) 物品偵測

訓練出YOLO自訂義模型後，去測試大巨蛋旁交通監控畫面，上面標註了物件偵測結果，例如車輛（car）、摩托車（motorbike）、行人（person），並附帶信心分數（例如 0.89、0.46 等），還有一個 Bus（公車）類型標籤出現在右上區域。可以看出來行人在畫面中非常的小，所以又退回到YOLO自訂訓練好多次，才讓行人的辨識改進許多。

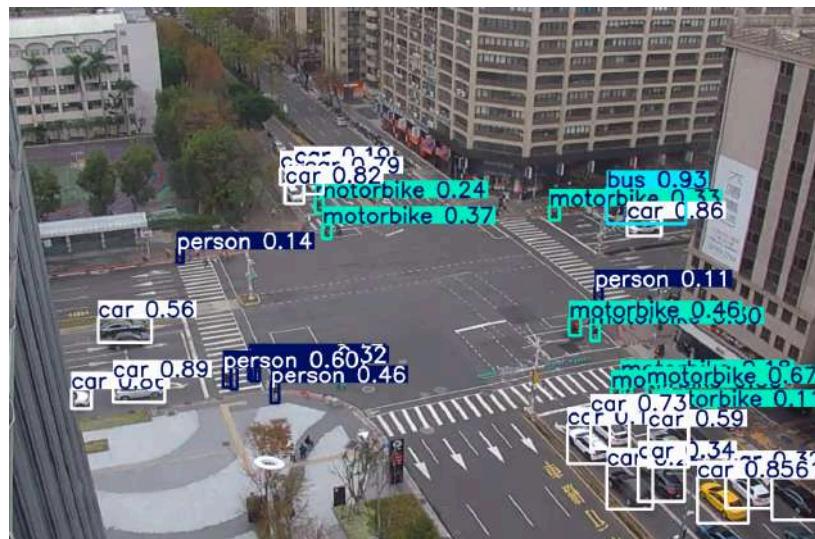


圖3-10：路口物品辨識和偵測。圖片來源：研究者自製。

## (七) 運動追蹤

YOLO第8版中有包括運動追蹤的功能，使用的就是前面有討論過運動追蹤BoT-SORT，主要就是會將路口的每一個可以辨識的物品加上編號id，同時加上它的信心程度分數（如圖3-11），在ReId方面有調整相關參數，讓影片中的物件可以較容易維持編號，影片播放時，編號會維持有一定表現。

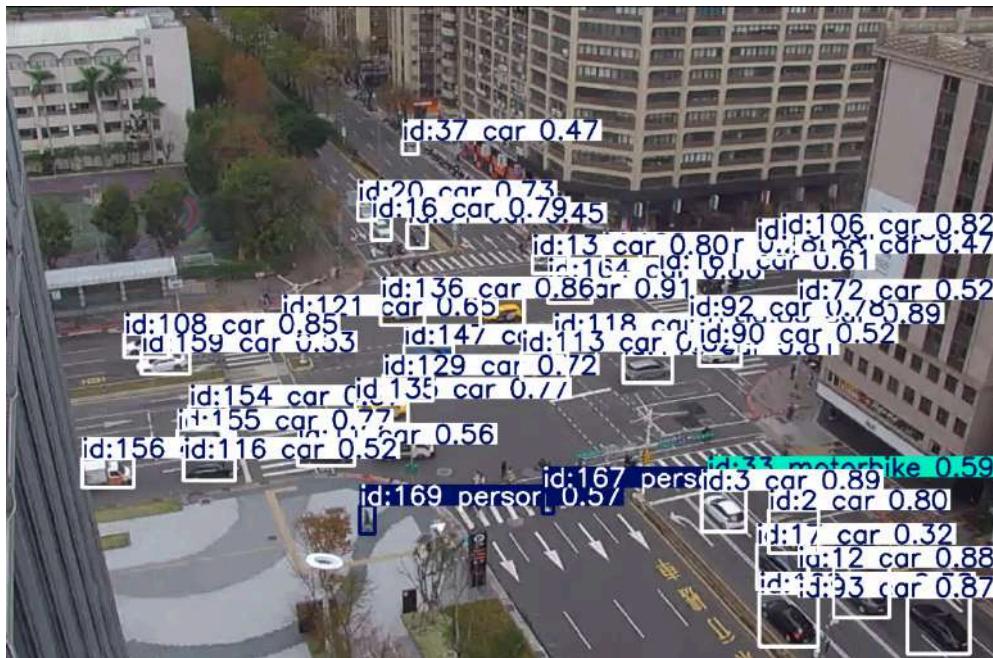


圖3-11：路口運動追蹤。圖片來源：研究者自製。

## (八) 座標轉換

透視轉換（Perspective Transform）是一種幾何變換，用來改變影像的透視角度。這種的座標轉換主要功能在矯正影像變形、鳥瞰視角轉換、文件掃描變正等應用中很常見。透視轉換，需要 4 組座標點才能計算（如圖3-12所示）。處理完座標轉換後，影像上的目標（如人或車的路線）可以被清晰地提取並標記。例如，在交通分析應用中，經過透視矯正後，從監控影像中提取的人與車的移動軌跡可以被精準地繪製到 Google 地圖上，方便進行更直觀的可視化與分析（如圖 3-13 所示）。這樣提高了影像的可讀性，也為後續找出問題的資料集提供視覺化的幫助。

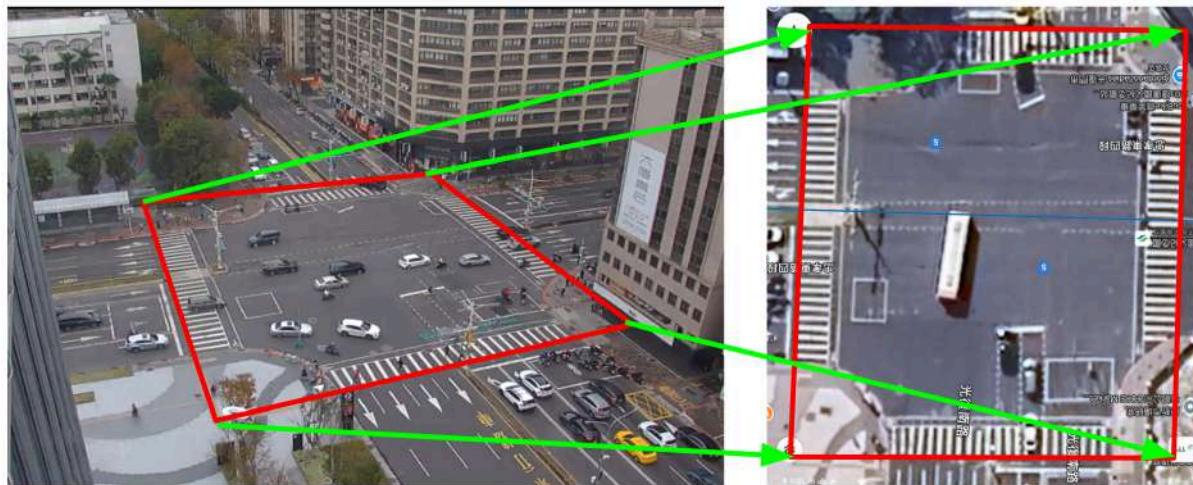


圖3-12：實際路口的座標轉換。圖片來源：研究者自製。

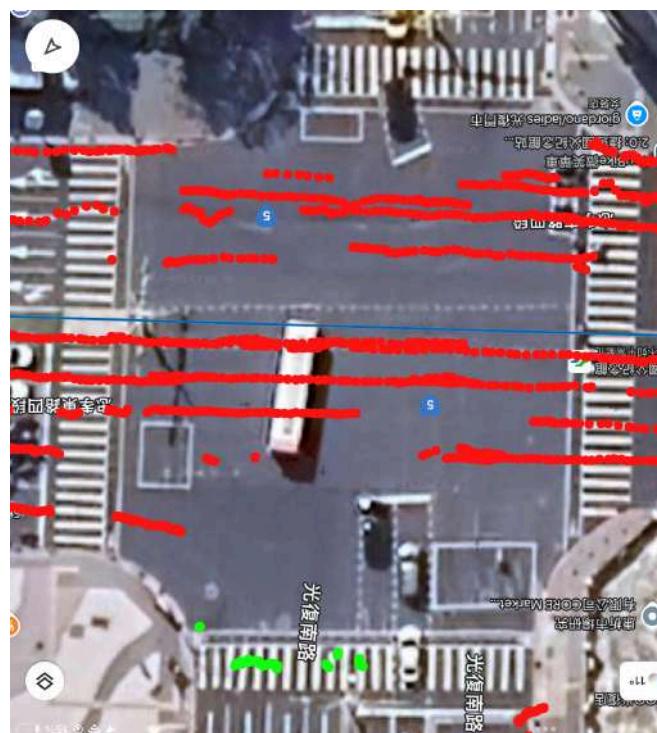


圖3-13：車子和行人軌跡圖。圖片來源：研究者自製。

## （九）資料集預先處理

在十字路口安全研究中，侵入後時間（Post-Encroachment Time，PET）是一種衡量行人與車輛之間潛在衝突風險的指標，並被廣泛用於道路交通安全研究。它的定義說明如下：

$$PET = t_2 - t_1$$

$t_1$ ：第一個道路使用者（像是行人或車子）離開可能發生衝突區域的時間。

$t_2$ ：第二個道路使用者到達同一可能發生衝突區域的時間。

PET 表示兩者在該區域的時間間隔。

根據文獻（Kathuria and Vedagiri, 2020），PET 值的分類標準如下： $PET \leq 3$  秒：屬於嚴重衝突，表示高風險的行人-車輛衝突。 $3 < PET \leq 6$  秒：屬於輕微衝突，有一定的風險但較低。 $PET > 6$  秒：屬於安全情形，基本無潛在危險。



圖3-14：車子從行人後方穿過情況，行人在 $t_1$ 先離開，車輛在 $t_2$ 時後來到達相同地方。

圖片來源：研究者自製。

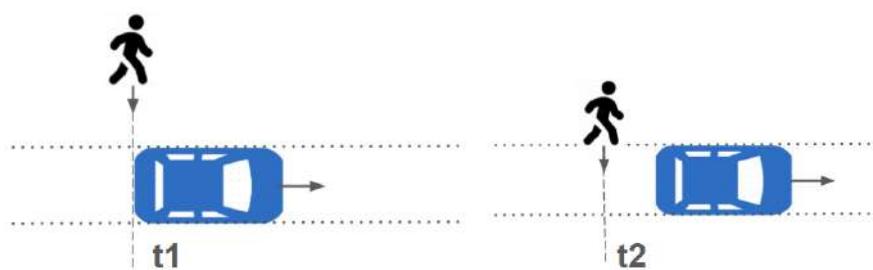


圖3-15：車子從行人前方穿過情況，車輛在 $t_1$ 先離開，行人在 $t_2$ 時後來到達相同地方。

圖片來源：研究者自製。

從錄製影片中，取出路線資料，分類成為安全 ( $6 < PET$ )、輕微衝突 ( $3 < PET < 6$  秒)、人車衝突 ( $PET \leq 3$  秒)，準備資料集數量為 210 : 20 : 9。由於數據分佈明顯不平衡，為了避免模型在訓練時偏向於數量較多的類別，採用了SMOTE數據增強技術。通過空間附近資料小範圍隨機組合成新樣本，使三種類別的數據量達到相同，平衡數據集，提高模型的訓練效果。這樣的處理方式能有效減少模型對特定類別的偏差，確保對

所有類別的預測能力更加均衡。輸入的資料和小球相似，是（person\_x, person\_y, car\_x, car\_y, pet\_type），而是事件發生前兩秒的路線資料4筆，時間間隔0.5秒。

## (十) 神經網路訓練

採用和小球碰撞相似的神經網路模型結構，不過找參數上有做調整。

LSTM 層：第一層 LSTM 使用 return\_sequences=True，因為輸出的序列需要傳遞到下一層 LSTM。第二層 LSTM 將序列輸出壓縮為單一向量。

Batch Normalization：每個主要層後面加上一個 Batch Normalization，穩定訓練過程。

Dropout層：部分層後面使用 Dropout，防止過度擬合（Overfitting），設定丟棄率為 0.2。

Dense 層：最後幾層是全連接層，最終輸出一個 3 類分類結果，激活函數為 softmax  
編譯模型:優化器選擇 Adam，損失函數選擇 CategoricalCrossentropy（適用於多類別分類）。

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 4, 64)	17,664
batch_normalization (BatchNormalization)	(None, 4, 64)	256
lstm_1 (LSTM)	(None, 64)	33,024
batch_normalization_1 (BatchNormalization)	(None, 64)	256
dropout (Dropout)	(None, 64)	0
dense (Dense)	(None, 32)	2,080
batch_normalization_2 (BatchNormalization)	(None, 32)	128
dropout_1 (Dropout)	(None, 32)	0
dense_1 (Dense)	(None, 16)	528
batch_normalization_3 (BatchNormalization)	(None, 16)	64
dense_2 (Dense)	(None, 3)	51

Total params: 54,051 (211.14 KB)  
Trainable params: 53,699 (209.76 KB)  
Non-trainable params: 352 (1.38 KB)

圖3-16：LSTM模型架構圖。圖片來源：研究者自製。

圖3-17展示了模型訓練過程中的損失函數與準確率的變化情況。左圖顯示模型在訓練集和驗證集上的損失值隨著訓練次數的減少而逐漸降低，反映出模型的誤差不斷縮小；右圖則顯示準確率逐漸提高，模型的預測能力在訓練集與驗證集上均有所提升。兩者的趨勢表明模型在訓練過程中效果逐步改善並趨於穩定，最後在訓練次數50時，達到73%的準確率。

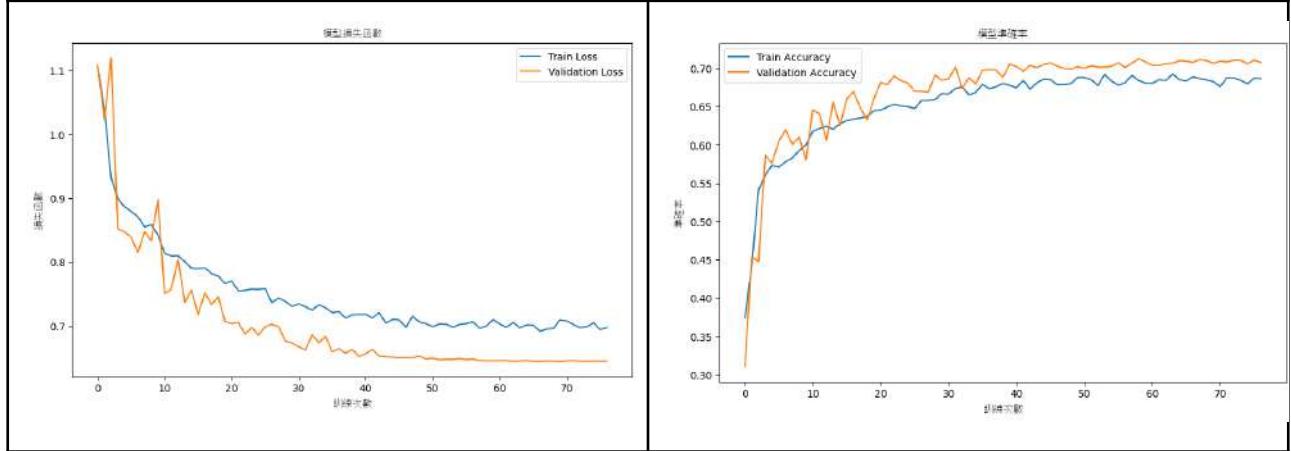


圖3-17：訓練時的損失函數和準確率。圖片來源：研究者自製。

## (十一) 分析結果、準確率和F1分數

### 1. Accuracy (準確率)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

其中：

TP (True Positive) : 真正，被正確分類為正類的樣本

TN (True Negative) : 真負，被正確分類為負類的樣本

FP (False Positive) : 偽正，實際為負類但錯誤分類為正類的樣本

FN (False Negative) : 偽負，實際為正類但錯誤分類為負類的樣本

### 2. F1-score (F1 分數)

定義：F1-score 是 Precision (精確率) 和 Recall (召回率) 的調和平均數：

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

其中：

Precision（精確率）：被分類為正類的樣本中，實際為正類的比例

$$Precision = \frac{TP}{TP + FP}$$

Recall（召回率）：實際為正類的樣本中，被正確分類為正類的比例

$$Recall = \frac{TP}{TP + FN}$$

模型的表現用準確率與 F1 分數來評估。在這次測試中，模型的 準確率為 73%，表示模型在所有預測中有 73% 是正確的。這表明模型在整體預測上具有一定的可靠性，但仍有 27% 的預測結果不準確，未來仍需要進一步改進。此外，模型的 F1 分數為 0.707，這是一個綜合指標，平衡了準確率（Precision）和召回率（Recall）的表現。F1 分數特別適用於資料類別不平衡的情況，當某些類別的樣本數量遠少於其他類別時，僅關注準確率可能導致誤導，而 F1 分數可以更全面地反映模型的分類能力。

F1分數表明模型在處理正類和負類的分類上有一定的平衡性，但還有改進空間。未來的優化方向可以包括：調整分類門檻以平衡 Precision 和 Recall，改進訓練數據的質量和分佈，或採用更複雜的模型架構來提升表現。總的來說，這些評估指標表明模型的基礎性能已達到一定水準，但仍需進一步調整以應對更多挑戰性的應用情形。

## 肆、研究結果

用電腦視覺技術預測十字路口人車碰撞風險，提升交通安全。主要研究結果如下：

一、物件辨識模型的建立與應用：成功使用YOLO電腦視覺技術，針對十字路口監控影像進行了行人與車輛的偵測與分類。透過對842張影像進行訓練（77%訓練集，6%驗證集，15%測試集，經數據增強後），YOLOv8自訂模型在物件辨識上達到了基礎的穩定性，其最佳mAP50（平均精度均值，IoU閾值0.5）為0.607。

二、運動軌跡資料的建立：研究運用了影像運動追蹤技術(BoT-SORT)和透視座標轉換，成功從監控影像中提取了十字路口行人與車輛的運動軌跡，調整參數讓ReId維持編號更穩定。透視轉換將影像座標轉換為鳥瞰視角下的地圖座標，使得後續計算侵入後時間(PET)更為準確，從而建立了可用於分析的運動軌跡資料集。

三、碰撞風險指標與數據集的構建：用後侵入時間(Post-Encroachment Time, PET)作為人車潛在衝突的風險指標。根據PET值將事件分為三類：安全( $PET > 6$ 秒)、輕微衝突( $3\text{秒} < PET \leq 6$ 秒)、人車衝突( $PET \leq 3$ 秒)(頁17)。基於此分類，對提取的軌跡資料進行了標註與分類，構建了一個包含210筆安全事件、20筆輕微衝突事件、9筆人車衝突事件的原始數據集，後續進行了數據增強以平衡樣本，用於訓練LSTM模型。

四、LSTM模型預測效能：研究透過長短期記憶網路(LSTM)模型，對未來2秒內是否發生人車碰撞事件進行預測。LSTM模型在測試集上的預測準確率達到了73%。這表明模型在預測十字路口潛在碰撞風險方面具有初步能力(圖4-1)。同時，模型的F1分數為0.707(頁20)，此指標綜合考量了模型的精確率(Precision)和召回率(Recall)，顯示模型在處理正負樣本不平衡的碰撞預測問題上達到了一定的平衡性。

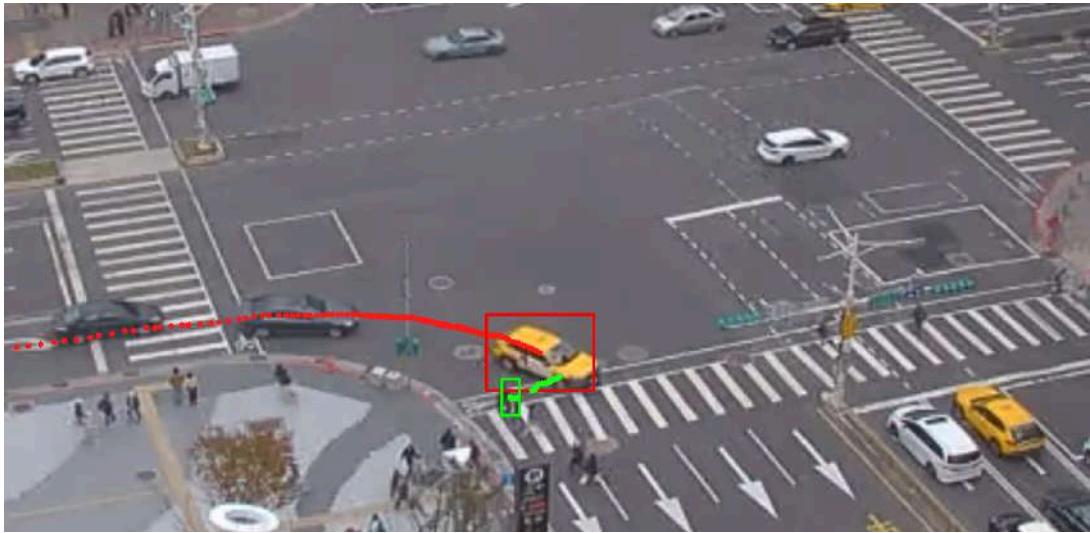


圖4-1：LSTM模型預測人車衝突情形（ $PET \leq 3$ 秒），攝影機視角。圖片來源：研究者自製。

五、模型效能驗證與改進建議：73%的準確率和0.707的F1分數驗證了所提出模型的初步效能。未來仍需要進一步改進。模型在處理正類和負類的分類上有一定的平衡性，但仍有改進空間。改善影像來源的穩定性與解析度、擴充及優化訓練數據集、調整模型架構或嘗試其他深度學習模型等，這些都可作為未來智慧交通應用的基礎。

## 伍、討論

一、錄製台北市路口影像過程相當具有挑戰性，主要受到影像穩定性影響，在高流量時段，即時影像串流系統負載增加，可能導致影像丟失或幀率下降，甚至發生5秒或更多的影像跳過現象，使得關鍵時間點的行人與車輛運動資訊遺失。這些問題進一步影響後續數據處理與機器學習模型的訓練，影像跳幀影響運動軌跡計算：追蹤演算法依賴連續影像來確保ID穩定，若有跳幀情況，可能導致物體ID重新分配或丟失，錄製的影像品質降低影響YOLO偵測準確度：影像壓縮或解析度下降可能造成物件邊界模糊，使行人和車輛的偵測效果下降，為了解決這些問題，未來考慮直接在路口架設運動相機來錄製影片，這樣可以確保影像更清楚和穩定，也能讓碰撞預測的結果更準確。

二、研究過程中發現，從高空拍攝的畫面裡，行人的面積太小了，讓電腦很難準確偵測和追蹤行人的移動。主要原因是攝影機得太高，讓行人在影像裡變得很小不明顯，原本的影像解析度只有 1280x720，畫面細節不夠清楚。而且，在重複錄製的過程中，解析度又進一步下降，讓行人的輪廓變得更模糊。天氣變化對車輛偵測影響不大，但對行人影響較大，特別是在低光環境、雨天或行人穿著變化時，陰天或低光環境（如黃昏、夜間）時，行人輪廓變得模糊，尤其是深色衣服的行人更難偵測。拿東西時也容易辨成摩托車，造成後面處理人車分類時的困難。

三、影像分類中，辨識單車、摩托車以及相關物體常常會遇到挑戰，主要是因為這些物體在外觀、形狀以及部分細節上具有高度相似性。例如，單車與摩托車在結構上都包含車架和車輪，當影像的解析度較低或光線不足時，這種相似性可能會導致分類模型誤判。此外，當人攜帶大件物品時，物品與人的組合形狀可能與摩托車的輪廓類似，增加了模型判斷的難度。

四、研究過程雖然辛苦，但也讓我獲得許多寶貴經驗。從錄製影像、標註行人和車輛，到整理數據、訓練模型，每個步驟都需要大量時間和耐心。標註影像時，遇到光線不好，特別難標，訓練 YOLO 時也常遇到辨識錯誤或準確度不足，需要不斷調整參數。雖然過程很累，但學習 YOLO 偵測讓我覺得很有趣，看到模型準確偵測行人和車輛時，真的很有成就感！這次研究不僅加深了我對電腦視覺與 AI 的理解，也讓我更有興趣繼續探索這個領域！

## 陸、結論

一、本研究使用長短期記憶（LSTM）神經網絡來預測十字路口的行人與車輛衝突，並採用電腦視覺技術從監視器片數據提取行人與車輛的特徵。研究結果顯示，LSTM 模型在十字路口的測試中有 73% 的預測準確率，證明有初步能力預測潛在行人與車輛衝突。該模型可進一步整合至車聯網環境，作為行人碰撞預警系統的一部分，以提升道路安全性，並可透過即時數據分析為駕駛人提供 2 秒前的警告。

二、雖已取得初步成果，但在行人辨識精度、模型對多變天氣與光照條件的適應性、以及追蹤算法的穩定性等方面仍有進一步提升的空間。未來研究可朝著改善鏡頭解析度和實地錄製、擴充更多較好品質的訓練數據集、比較不同的深度學習模型、以及探索提取更強的穩定性特徵與追蹤技術等方向努力。

三、未來可以應用於實際場景，在智慧交通安全應用上具有潛力，電腦視覺與碰撞預測模型可將車聯網中的路側單元 (RSU) 升級為具備主動感知與即時分析的智慧終端，為路口提供「眼睛」與「大腦」。透過聲光警示或資訊推播，提醒危險中的行人或駕駛員。RSU 還可透過 V2I 及 V2P 通訊廣播碰撞風險資訊，輔助路網中連線車輛決策並實現車人雙向預警，促進車聯網環境中技術的整合。

四、總結本研究貢獻，證明了結合電腦視覺與 LSTM 模型的框架，能夠有效預測十字路口行人與車輛的潛在衝突。這對於未來開發智慧交通安全預警系統、減少交通事故、保障行人安全，具有價值與實際應用前景。

## 柒、參考文獻資料

馮惠宜（2024 年 9 月 23 日）。台中公車撞死人路口有路殺陷阱 交通局認：常發生車禍。工商時報。檢自：<https://bit.ly/40OV31g>

Nikolic, N., Davidovic Manojlović, M., & Mészáros, M. (2023). The usage of geospatial tools in traffic sign detection. Zbornik radova Departmana za geografiju, turizam i hotelijerstvo, 52(2), 111 – 122.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735 – 1780. <https://doi.org/10.1162/neco.1997.9.8.1735>

Christopher O, (2015), Understanding LSTM Networks.  
<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Aharon, N., Orfaig, R., & Bobrovsky, B.-Z. (2022). BoT-SORT: Robust Associations Multi-Pedestrian Tracking. <https://arxiv.org/abs/2206.14651>

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321 – 357.

胡大瀛、李岳洪（2022）。數據不平衡下以機器學習方法預測交通事故嚴重性之分析。運輸計劃季刊第五十一卷 第四期，台北市。

溫基信（2021）。路口俯視攝影技術於交通衝突分析之案例應用與比較。交通部運輸研究所 (報告編號 110-067-3496)。

張斯涵（2022）。利用統計預測與機器學習從駕駛行為預測事故風險之研究。國立陽明交通大學運輸與物流管理學系：碩士論文，新竹市。

Du, W., Ye, J., Gu, J., Li, J., Wei, H., & Wang, G. (2022). SafeLight: A reinforcement learning method toward collision-free traffic signal control. arXiv preprint arXiv:2211.10871.

<https://arxiv.org/abs/2211.10871>

Tram, T., Jansson, A., Grönberg, R., Ali, M., & Sjöberg, J. (2018). Learning negotiating behavior between cars in intersections using deep Q-learning. arXiv preprint arXiv:1810.10469.

<https://arxiv.org/abs/1810.10469>

謝辰陽（2020）。基於結合CNN與LSTM神經網路之車輛碰撞風險預測。國立成功大學交通管理科學系：碩士論文，臺南市。

Kathuria, A., Vedagiri, P., 2020. Evaluating pedestrian vehicle interaction dynamics at un-signalized intersections: a proactive approach for safety analysis. Accid. Anal. Prev. 134, 105316.

## 【評語】032819

本作品探討如何利用電腦視覺技術預測十字路口行人與車輛的運動軌跡及碰撞風險，具有提升交通安全的實務應用價值。作者在研究中先進行連續影像中物件標註與資料處理以建構同一個行人或車輛的移動軌跡，接著把這些軌跡當成 LSTM 模型的輸入來預測行人與車輛相撞的三種不同危險等級，可用來做即時路口碰撞預測。此作品目前所用的訓練資料（特別是實際碰撞類型）很少，如此會限制訓練出來模型的泛化能力。另外，作者在進行效能評估實驗時，變因控制可再細化：比如不同時間、天氣、行人密度對碰撞預測正確性的影響目前未進行比較。此作品未來可考慮如何部署於真實世界的路口或車輛上，例如：如何使用車聯網或 5G 網路來即時傳輸碰撞警告訊息給即將相撞的車輛駕駛人和行人。

作品海報



# 十字路口人車 碰撞預測之研究

# 摘要

本研究利用電腦視覺技術預測十字路口的行人與車輛碰撞風險，以提升交通安全。研究結合 YOLO 進行即時物件偵測與運動追蹤，運用 PET 風險指標來分類衝突事件，運用 LSTM 建立時間序列模型，預測碰撞可能性。實驗結果顯示，模型可達成碰撞預測準確率達 78%。未來可應用於智慧交通系統與車聯網，強化主動安全與事故預防能力。

## 介紹

### 研究動機

台灣每年因交通事故死亡人數超過 3000 人，平均每天至少有一名行人喪生，被國際稱為「行人地獄」，令人痛心與憤怒。身為學生，希望透過學習和運用電腦視覺技術，研究行人與車輛碰撞預測，為提升交通安全貢獻一份力量。

### 研究目的

- (一) 運用電腦視覺辨識 (YOLO) 十字路口移動的車輛和行走的行人
- (二) 運用影像運動追蹤與透視轉換技術，建立十字路口中車與人的運動軌跡資料。
- (三) 以後侵入時間 (PET) 為風險指標，進行資料標註，建立潛在衝突數據集。
- (四) 透過 LSTM 模型預測未來發生人車碰撞事件的可能性，提升道路使用者安全。
- (五) 驗證模型效能並提出可行改進建議，作為未來智慧交通應用基礎。

### 文獻回顧

#### YOLO

YOLO 用一階段物件偵測法，透過單一神經網路同時預測物件位置與類別。影像被分成網格，每格預測多個物件的邊界框及其 x、y、寬、高與信心分數。

#### 長短期記憶網路 (LSTM)

LSTM 是一種時間序列模型，透過記憶單元及遺忘門、輸入門與輸出門三大機制，有效保留長期依賴資訊。可應用於語言處理、語音識別與預測任務，在處理長序列資料時，展現出更佳的穩定性與準確性。

#### 影像運動追蹤 BoT-SORT

BoT-SORT 是一種多目標追蹤技術，常用於監視、交通與運動分析。它先處理高信心目標，再考慮低信心項，結合改良卡爾曼濾波器進行移動預測，能在快速場景中穩定追蹤、減少 ID 變動並提升準確度。

#### 路口碰撞相關研究

強化學習廣泛應用於交通領域。Tommy T. 等人 (2018) 將其用於交通號誌控制，兼顧安全與效率；也應用於自駕車在路口的談判決策。謝辰陽 (2020) 則結合 CNN 與 LSTM，透過行車紀錄器預測 3 秒內車禍風險，提升警示準確性。

## 研究過程與方法

### 研究架構

本研究從主題確立開始，學習電腦視覺技術並參考相關研究成果。透過收集路口影像並以 Python 進行處理與分析，若涉及 AI，則用整理後資料訓練模型，最後分析結果並持續改進，將想法轉為具體成果。

確定  
主題

文獻  
探討

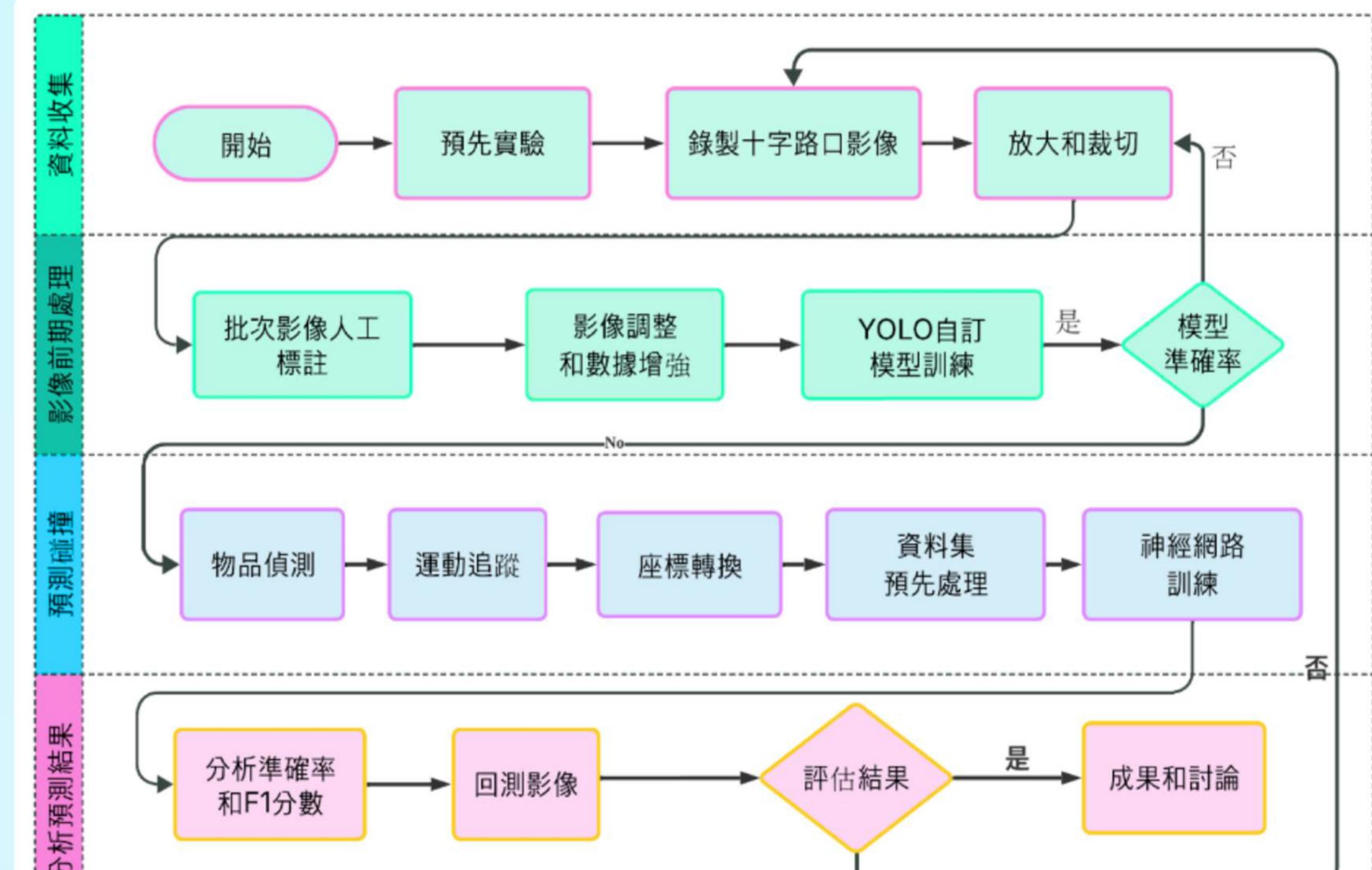
影像  
收集

撰寫  
程式

訓練  
模型

分析  
改進

研究架構圖



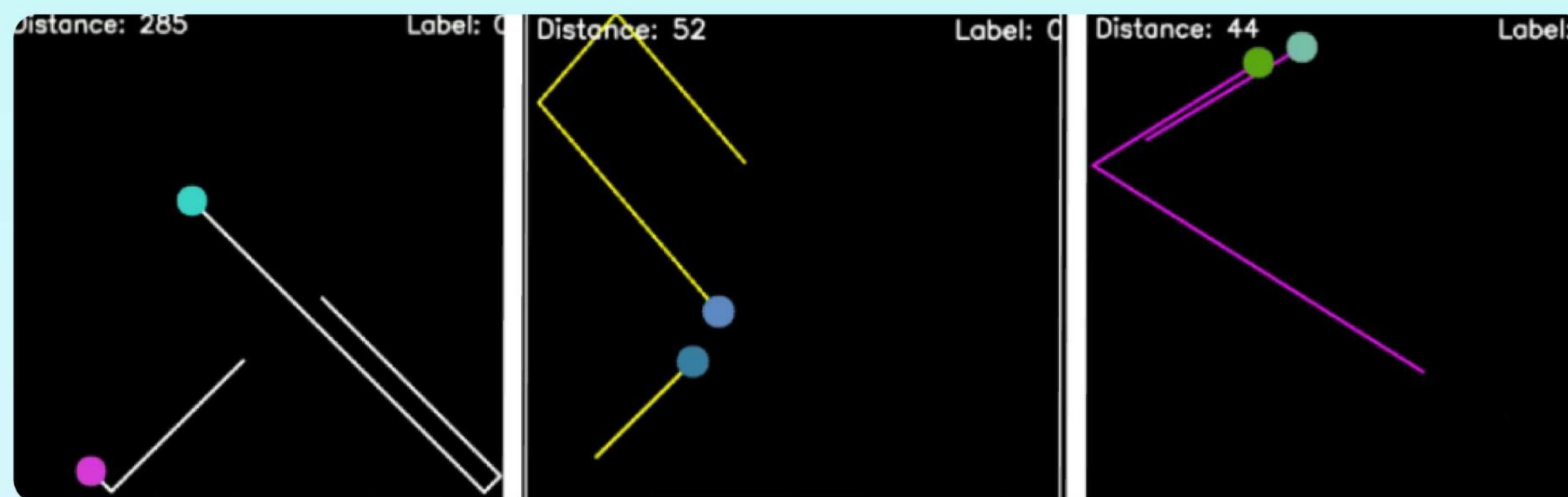
研究流程圖

## 研究過程

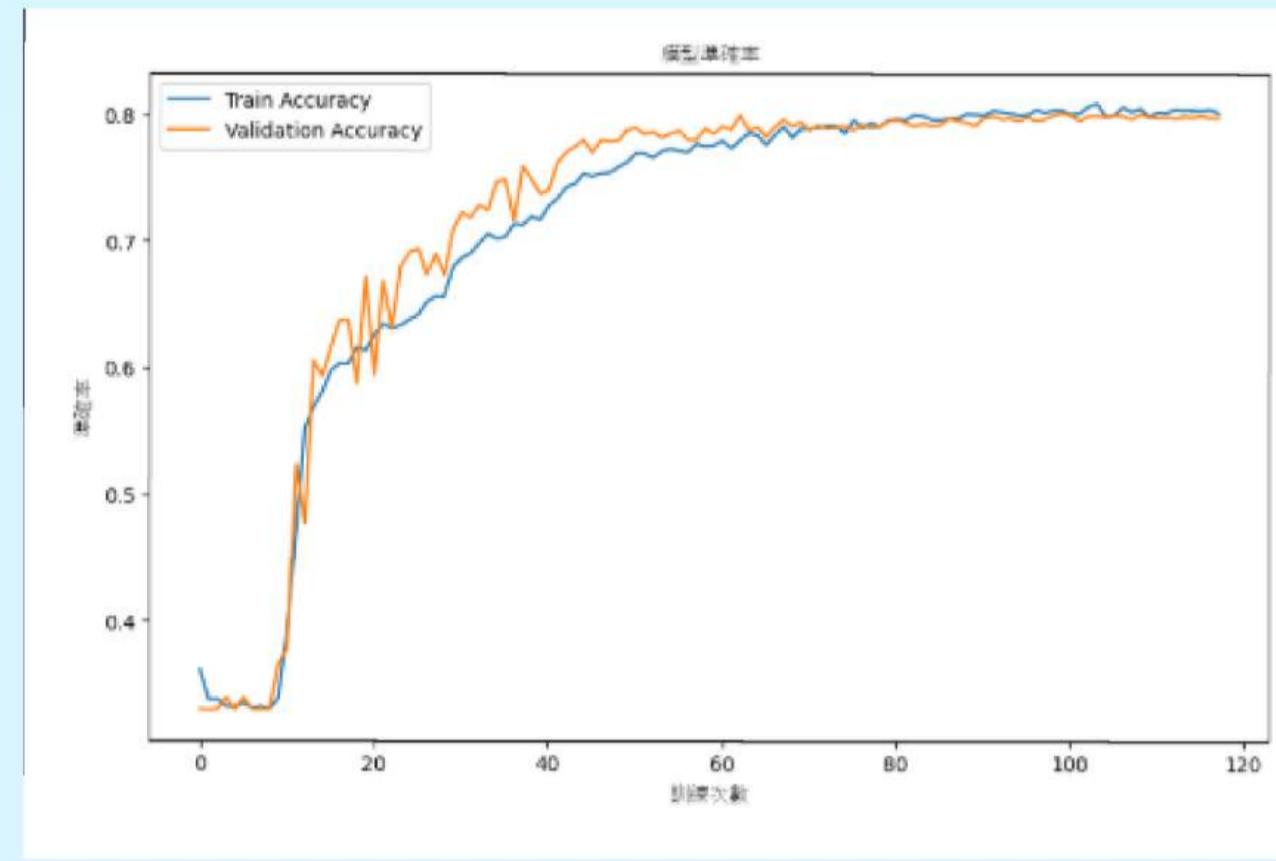
### 預先實驗

設計快慢兩顆小球（模擬車與行人）運動，發生接近或碰撞事件時，記錄前1~3秒軌跡資料作為LSTM輸入（4筆，每筆為兩球座標）。輸出值分類為0~2，共3類：

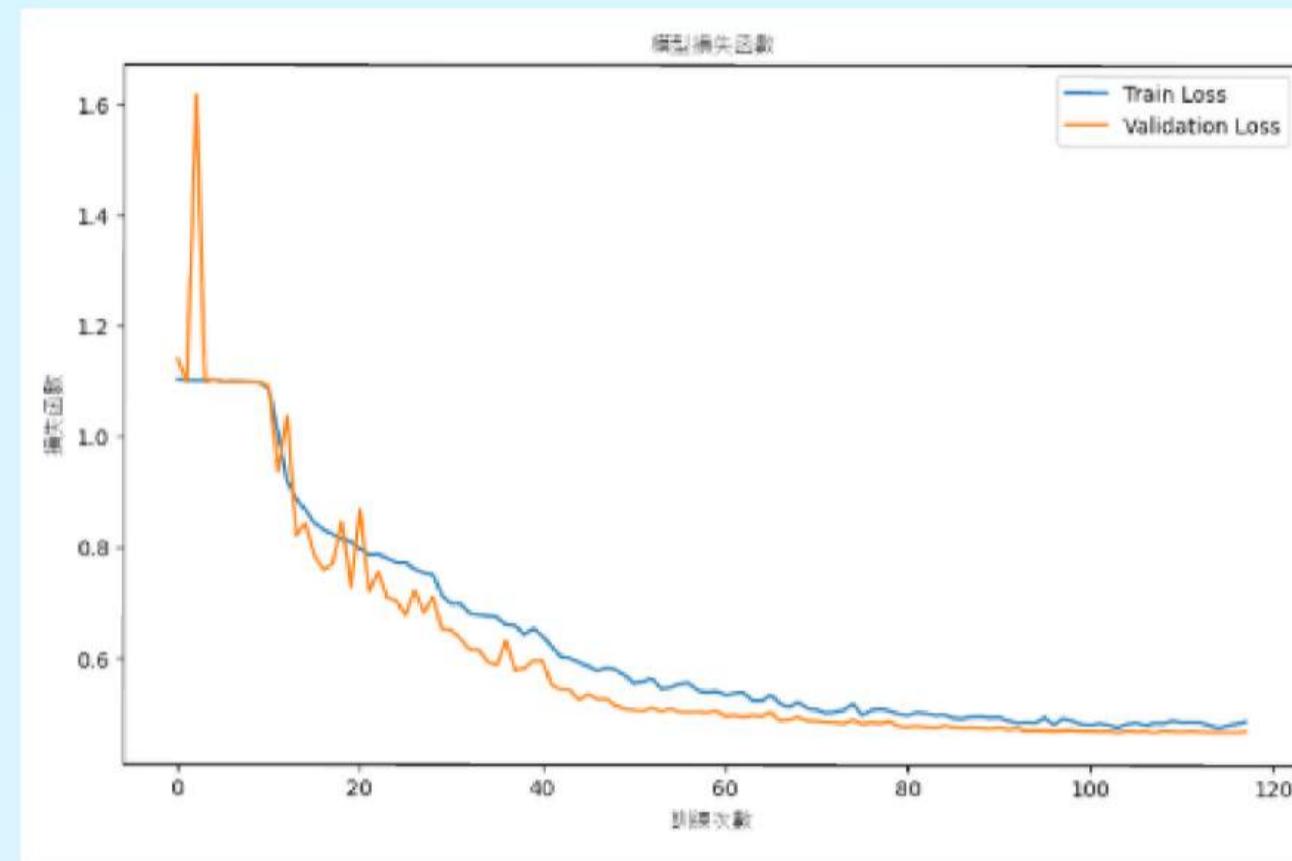
- 2->1秒內極可能碰撞，
- 1->1秒內接近但未碰撞，
- 0->距離未構成威脅。



小球事件分類圖



訓練時損失函數圖

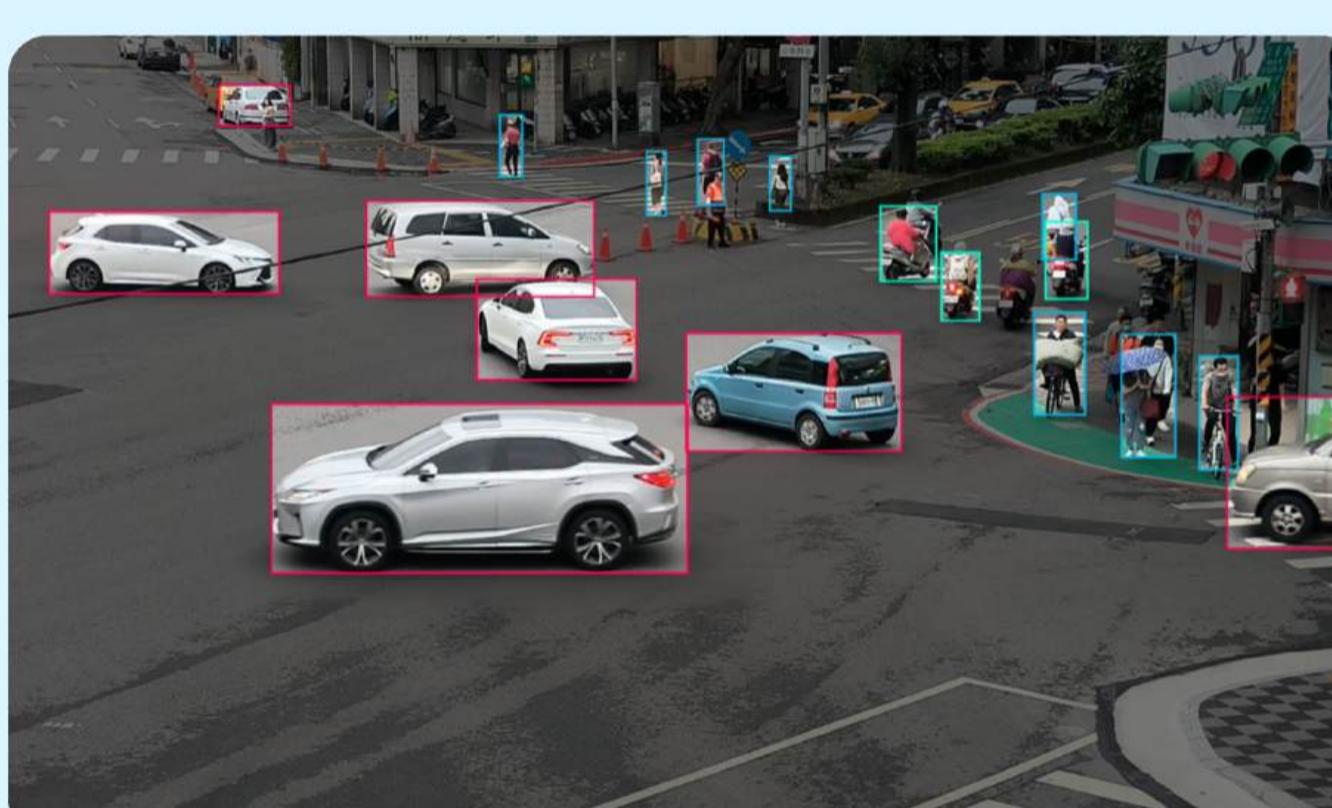


訓練準確率圖

### 影像錄影和標註

原先使用台北市即時交通資訊網影像，影像品質不夠好，改採實際錄製永吉路和松隆路口西向的車流影像，影像包括白天和晚上，約40小時，標註人車資料269筆後，匯出YOLO格式，訓練碰撞風險預測模型。

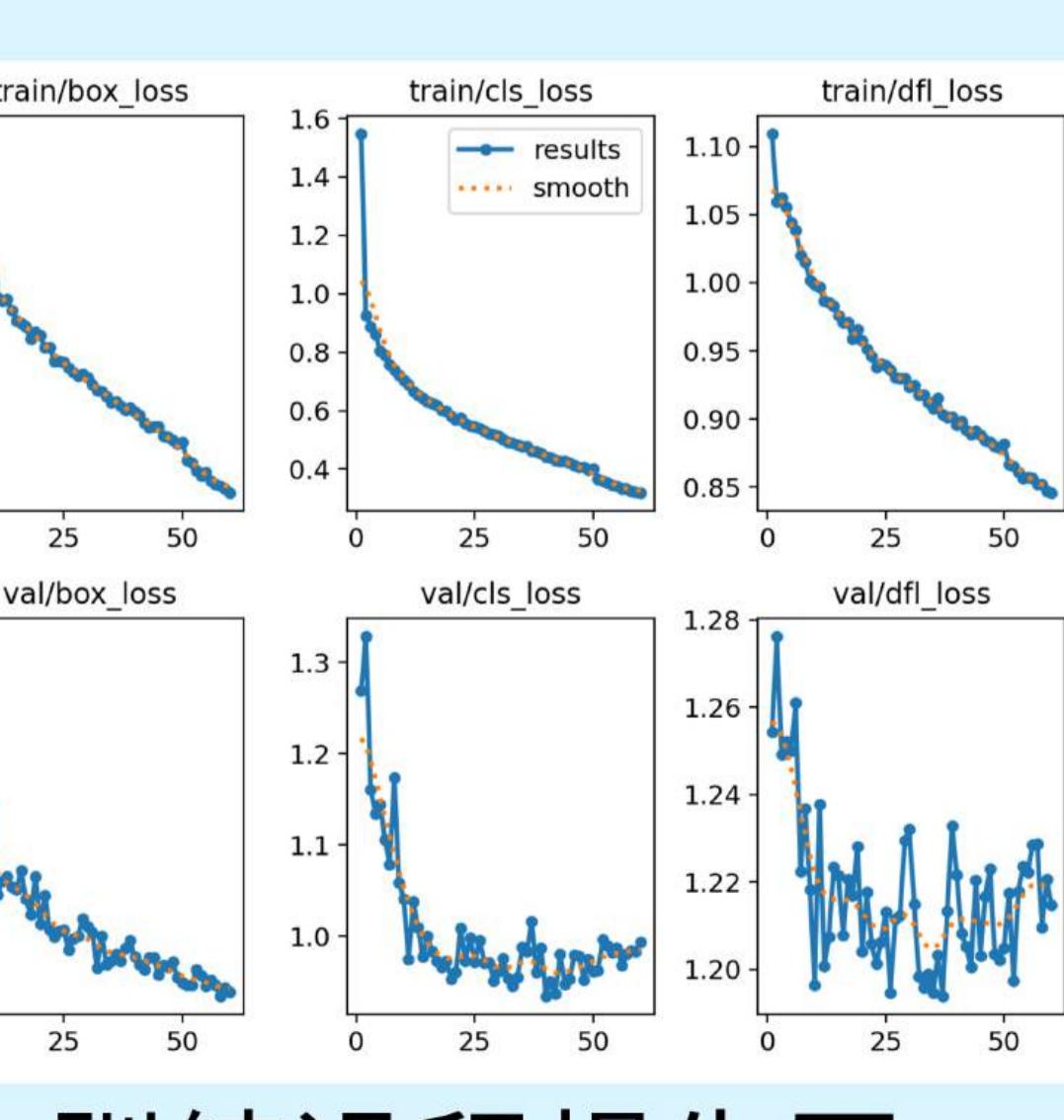
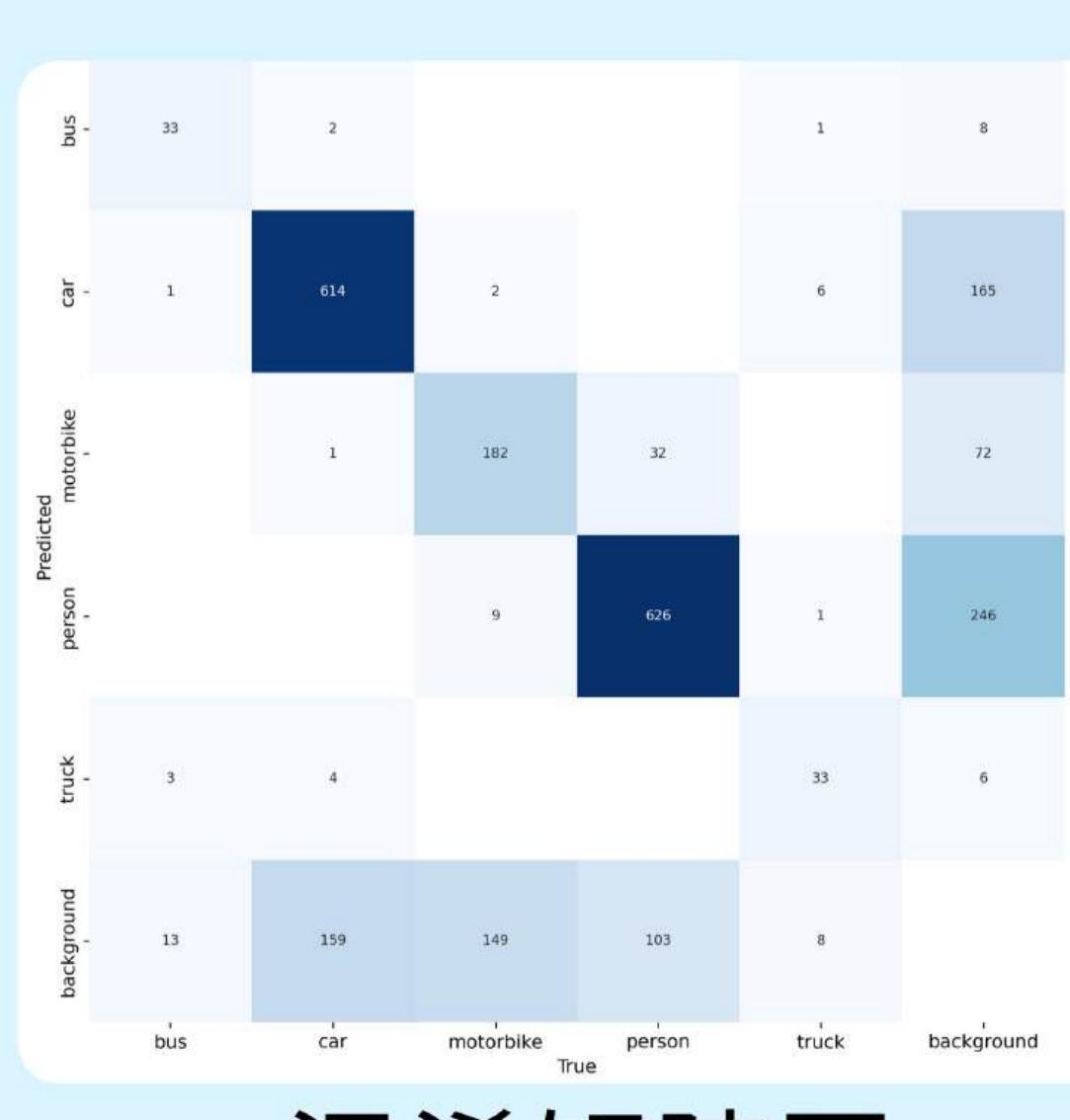
影像用數據增強隨機旋轉和隨機調整亮度，提升多樣性來增強模型對路口不同天氣亮度辨識能力。



人工標註影像圖

### YOLO自訂模型訓練

使用685張圖片做訓練資料，使用數據增強方法，數量加倍。我們把資料分成三部分：77%拿來訓練模型、6%用來在訓練時調整參數、15%用來最後測試準確率。訓練過程中會看到三條線：box\_loss 代表框住物體的位置誤差、cls\_loss 是分類錯誤、dfl\_loss 是判斷框有多準。這些線越低代表模型表現越好。用混淆矩陣來看模型預測表現，mAP@50是72.2%。



混淆矩陣圖

### 人車偵測和追蹤

訓練YOLO自訂模型後，測試永吉路天橋旁交通監控畫面，標註了物件偵測結果，例如車輛、摩托車、行人，並附帶信心分數，反覆回到YOLO自訂訓練10次，才讓行人的辨識改進。YOLOv11中包括運動追蹤功能，修改BoTSORT參數，ReID編號，同時加上它的信心程度分數，播放影片編號會維持，不易重新編號。



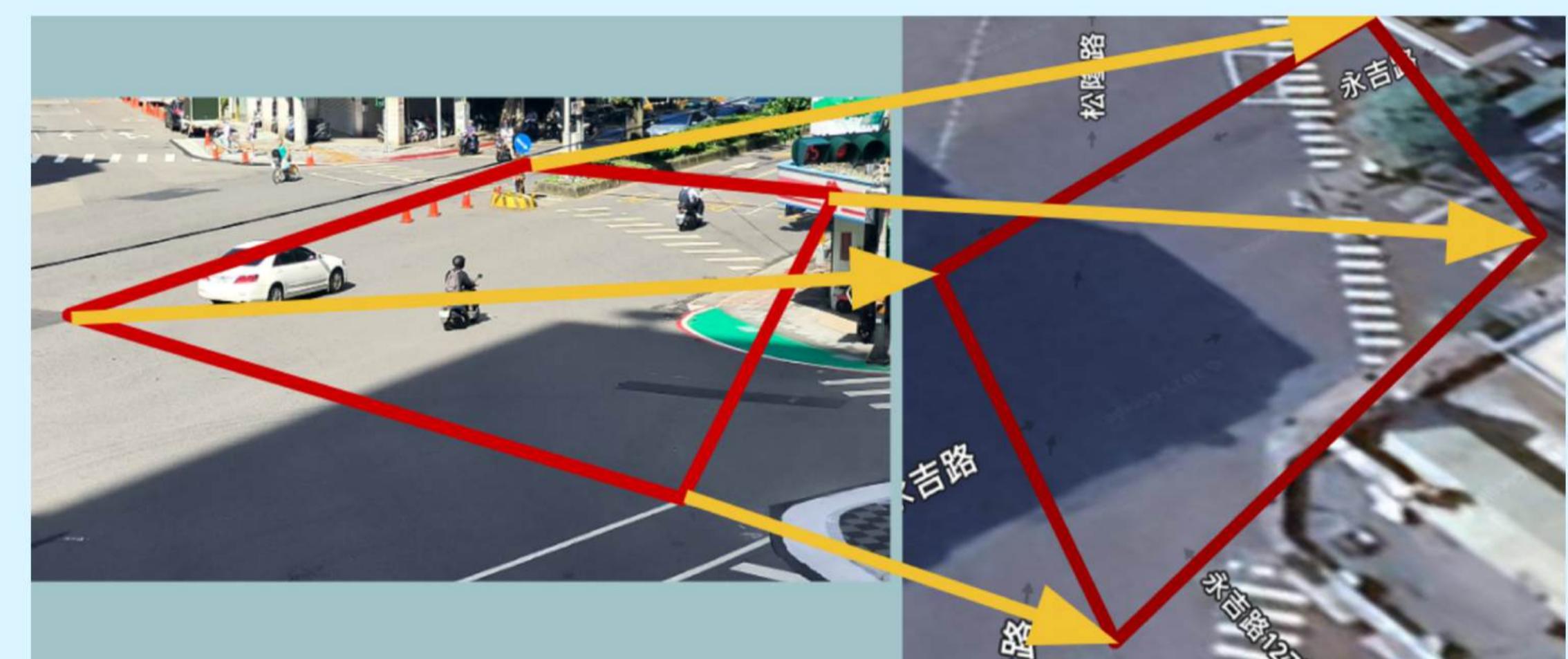
路口物件辨識圖



運動ID追蹤圖

### 座標轉換

鏡頭中近快遠慢，採用透視轉換，改變透視角度。需要4組座標點才能計算，訓練一組YOLO模組，辨識路口四個標識點，自動轉換。經透視矯正後，從監控影像中提取的人與車的移動軌跡可以精準地繪製到Google地圖上，進行更直觀的可視化與分析，提高碰撞預測的正確性。



YOLO路標辨識到路口的正視座標轉換圖

### 資料集處理

在路口安全研究中，PET（Post-Encroachment Time）是衡量行人與車輛潛在衝突風險的重要指標。定義為  $PET = t_2 - t_1$ ， $t_1$  為第一位使用者離開衝突區域的時間， $t_2$  為第二位使用者進入該區域的時間。依據 PET 值分類： $\leq 3$  秒嚴重衝突、 $3\sim 6$  秒輕微衝突、 $> 6$  秒安全情形。本研究自錄影中擷取軌跡資料，共分為三類：安全688筆、輕微衝突212筆、嚴重衝突68筆。由於資料嚴重不平衡，透過小範圍隨機生成新樣本進行數據增強（SMOTE）平衡各類別數量，提升模型訓練品質與分類能力。



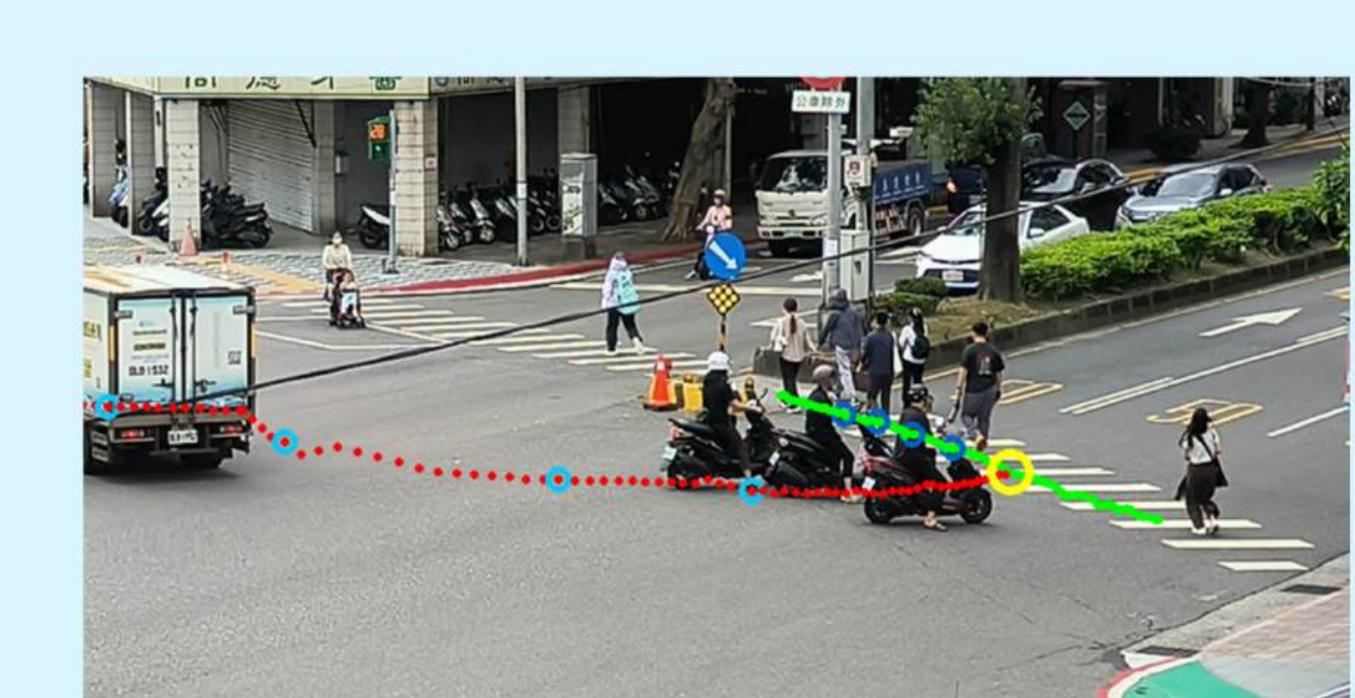
車子從行人後方穿過



車子從行人前方穿過



資料增強平衡直方圖



人車各4組座標資料

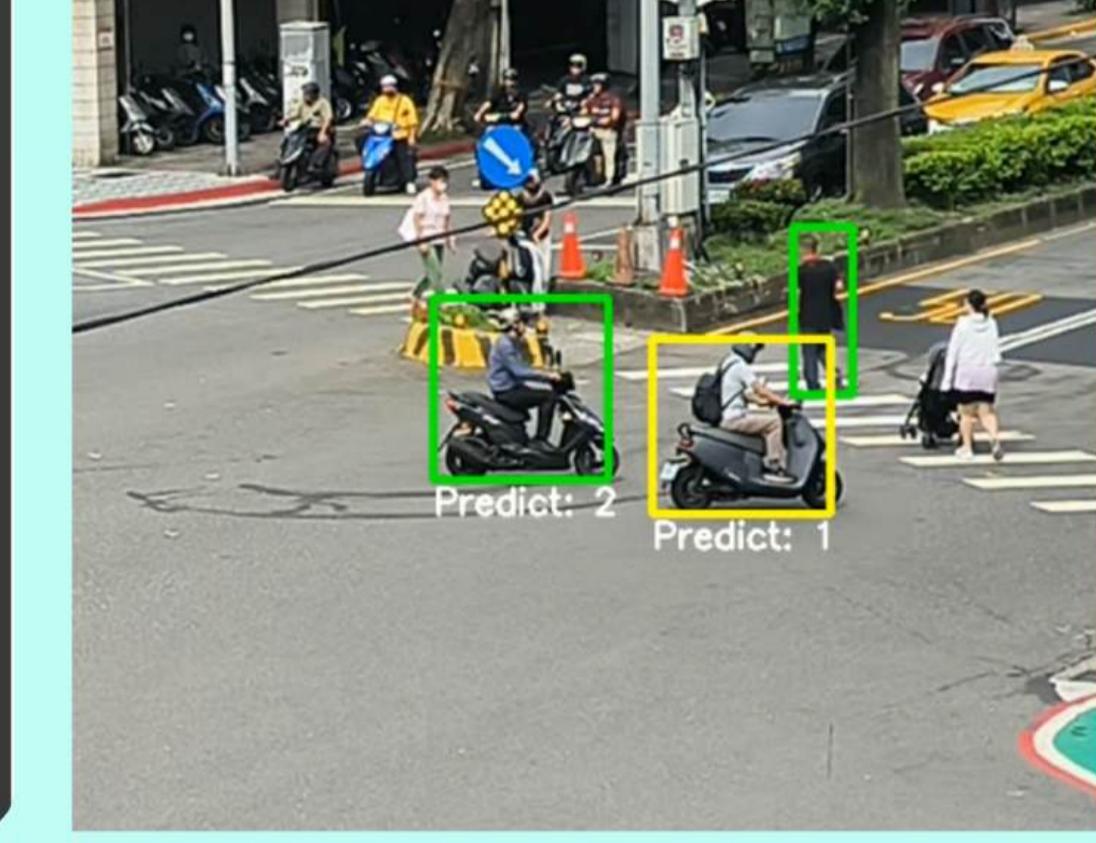
## 神經網路模型訓練

- 採用雙層 LSTM 架構：第一層輸出完整時間序列。第二層將序列壓縮為單一特徵向量。
- Batch Normalization (批次正規化)：每個主要層後加入此層，以提升訓練穩定性與加速收斂。
- Dropout 層：丟棄率為 0.1，有助於減少過度擬合現象。
- 全連接層 (Dense Layer)：負責將壓縮後的向量輸出轉為三類分類結果。
- 用網格搜尋方法，找到最佳參數。

## 準確率和F1分數

模型訓練後準確率為 78%，召回率 80.5%，精確率 80.6%，F1 分數為 80.5%。F1 分數能平衡精確率與召回率，顯示模型在分類正負樣本上具一定平衡性，整體表現穩定。

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 4, 64)	17,664
batch normalization (BatchNormalization)	(None, 4, 64)	256
lstm_1 (LSTM)	(None, 64)	33,624
batch normalization 1 (BatchNormalization)	(None, 64)	256
dropout (Dropout)	(None, 64)	0
dense (Dense)	(None, 32)	2,080
batch normalization 2 (BatchNormalization)	(None, 32)	128
dropout_1 (Dropout)	(None, 32)	0
dense_1 (Dense)	(None, 16)	528
batch normalization 3 (BatchNormalization)	(None, 16)	64
dense_2 (Dense)	(None, 3)	51



LSTM模型架構圖

模型預測結果

## 研究結果

一、建立物件辨識模型：在永吉市民路口重新錄製影像後，用YOLO v11電腦視覺技術，針對十字路口監控影像進行人與車輛的分類。透過對269張影像進行訓練，訓練後達到了mAP 72.2%辨識穩定性。

二、運動軌跡資料的建立：影像運動追蹤技術 (BoT-SORT) 和透視座標轉換，成功從錄製影像中提取了路口行人與車輛的運動軌跡，調整參數讓ReID維持編號更穩定。訓練路標模型再自動轉換影像座標到為鳥瞰視角下的地圖座標。

三、碰撞風險指標與數據集構建：用後侵入時間作為人車潛在衝突的風險指標，將事件分為三類：安全 ( $PET > 6$ 秒)、輕微衝突 ( $3 < PET \leq 6$ 秒)、人車衝突 ( $PET \leq 3$ 秒)。對軌跡資料進行分類，構建包含688筆安全事件、212筆輕微衝突事件、68筆人車衝突事件的數據集，後續進行數據增強 (SMOTE) 以平衡樣本，再訓練LSTM模型。

四、LSTM模型預測效能：研究透過長短期記憶網路 (LSTM) 模型，對未來2秒內是否發生人車碰撞事件進行預測。LSTM模型在測試集上的預測準確率達到了78%。這表明模型在預測十字路口潛在碰撞風險方面具有初步能力。同時模型的F1分數為80.5%，顯示處理正負樣本不平衡的預測上達到不錯的平衡性。

五、模型效能驗證與改進建議：本研究所提出模型初步達到不錯的效能。未來可以從加強摩托車和人的分類、調整模型架構或嘗試其他深度學習模型、跨路口共用模型的研究等方向更深入研究和改進。

## 結論

本研究使用長短期記憶 (LSTM) 神經網絡來預測十字口的行人與車輛衝突，採用電腦視覺技術從監視器片數據提取行人與車輛的特徵。研究結果顯示，LSTM模型在十字路口的測試中有78%的預測準確率，證明有能力預測潛在行人與車輛衝突。可進一步整合至車聯網環境，作為行人碰撞預警系統的一部分，以提升道路安全性，可透過即時數據分析為駕駛人提供2秒前的警告。未來可以結合更多交通感測技術，來提升預測效果。總體來說，研究證明LSTM模型能有效預測行人與車輛的潛在衝突，對於未來智慧交通系統和行人安全預警系統具有重要的應用價值。

## 參考文獻

- 馮惠宜 (2024年9月23日)。台中公車撞死人路口有路殺陷阱 交通局認：常發生車禍。工商時報。檢自：<https://bit.ly/40OV31g>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735
- Aharon, N., Orfaig, R., & Bobrovsky, B.-Z. (2022). BoT-SORT: Robust Associations Multi-Pedestrian Tracking. <https://arxiv.org/abs/2206.14651>
- 胡大瀛、李岳洪 (2022)。數據不平衡下以機器學習方法預測交通事故嚴重性之分析。運輸計劃季刊第五十一卷第四期，台北市。
- 溫基信 (2021)。路口俯視攝影技術於交通衝突分析之案例應用與比較。交通部運輸研究所 (報告編號 110-067-3496)。
- 張斯涵 (2022)。利用統計預測與機器學習從駕駛行為預測事故風險之研究。國立陽明交通大學運輸與物流管理學系：碩士論文，新竹市。
- 謝辰陽 (2020)。基於結合CNN與LSTM神經網路之車輛碰撞風險預測。國立成功大學交通管理科學系：碩士論文，臺南市。