

日期: 4/1

講者: 中央研究院 資訊科學研究所 特聘研究員 廖弘源 博士

題目: From YOLOv4, YOLOv7, to YOLOv9

關鍵字: 人工智慧(AI)、YOLO、智慧城市、物件偵測、ImageNet、AlexNet

心得報告:

本週有幸邀請到中央研究院資訊科學研究所的廖弘源博士，和我們分享從 YOLOv4 和 YOLOv7 最後到 YOLOv9 的設計架構及其應用，首先和我們說明，科學家 (Scientist) 與工程師 (Engineer) 之間的差異，科學家使用**系統性的方法**來解決**長期的問題**，而工程師使用**現有的方法**來解決**當前的問題**。我們須了解其兩者間的差異，才能更好地進入研究的狀態。

接著介紹了由科技部(MOST)贊助的 AI 專案，主題為: Smart City Traffic Flow Solutions，希望通過 AI 技術來優化智慧城市中的交通流，其核心目標為以下三點:

1. 將學術研發能力導入產業。
2. 將人工智慧(AI)技術商業化，讓產品進入國際市場。
3. 讓世界看見台灣的科技能力，提升國際能見度。

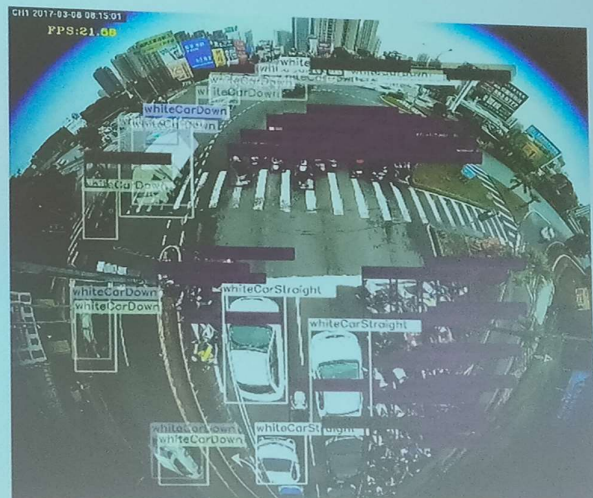
團隊間的共識如下:

1. 選擇來自於產業的真實問題，而不是閉門造車。
2. 解決問題的過程導入科學精神，而不僅僅是依靠經驗。

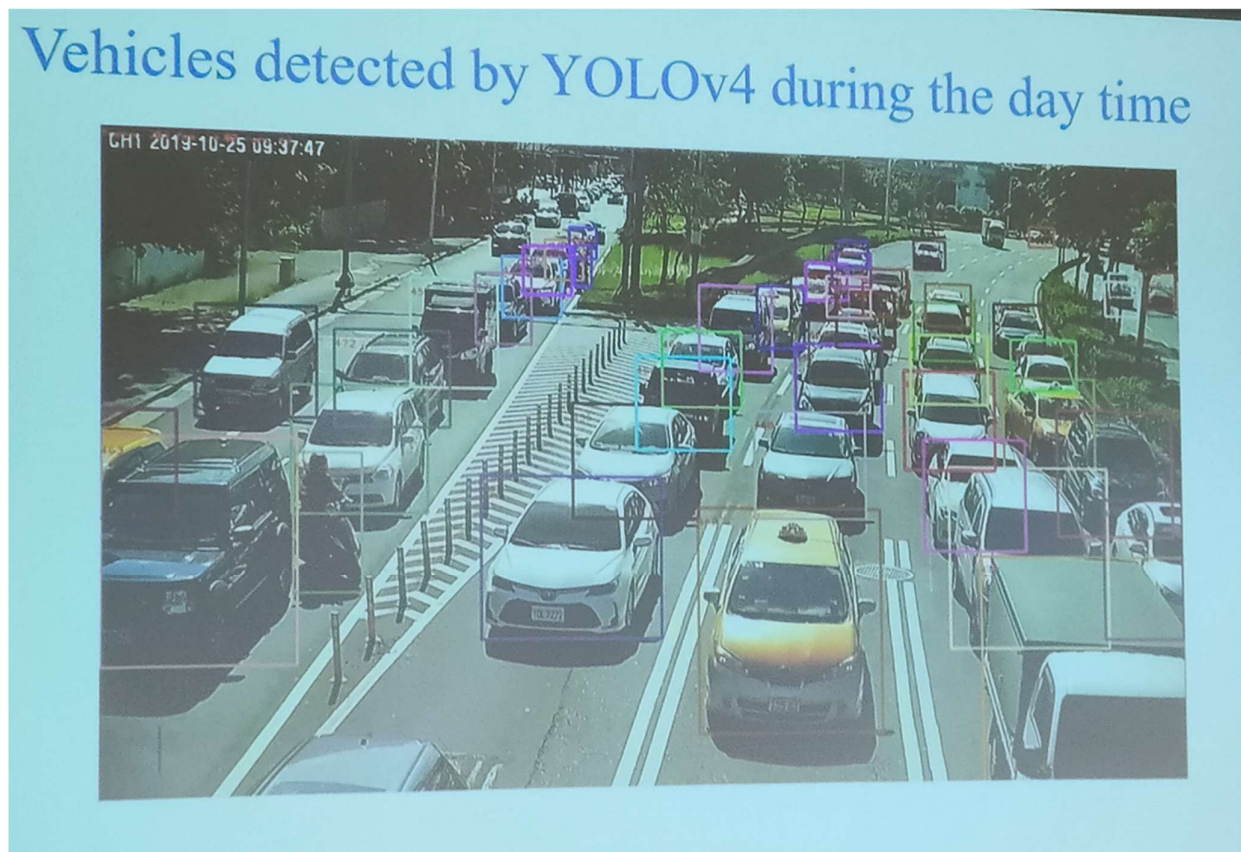
將下來為我們展示了其團隊在 2018/6 ~ 2018/11 的 realtime 成果，僅僅使用了 Nvidia Jetson TX2 和魚眼鏡頭進行攝影就能辨識基本的車輛。

Demo video

- Nvidia Jetson TX2 real time
- Accomplished during 2018.6 ~ 2018.11

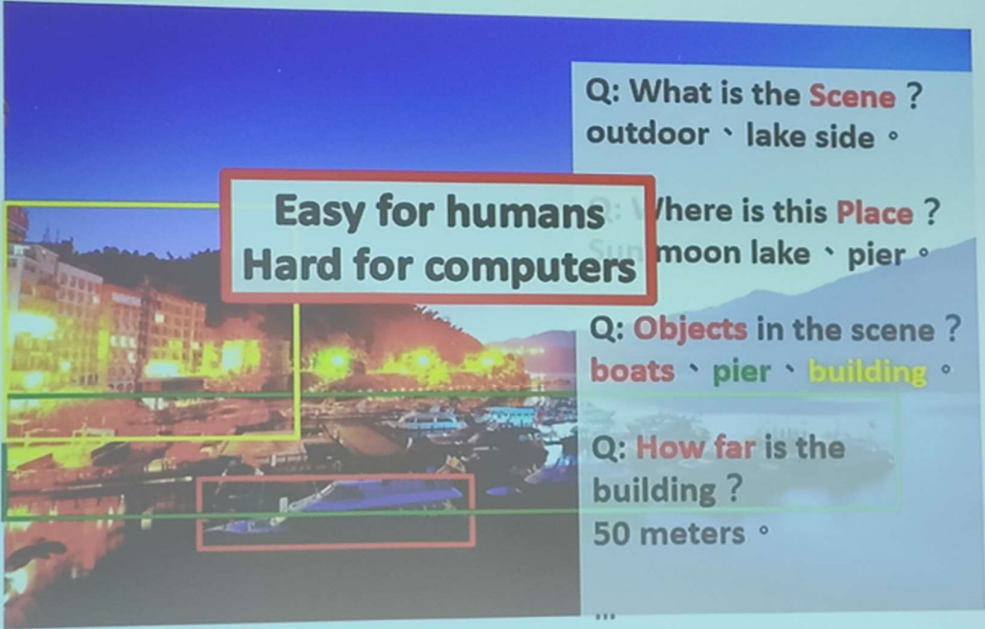


後來使用了 YOLOv4 進行車輛的辨識:



接著向我們介紹從前要處理一張未知影像到底有多麼的困難:

2010 以前處理一張未知影像到底有多難?



Easy for humans
Hard for computers

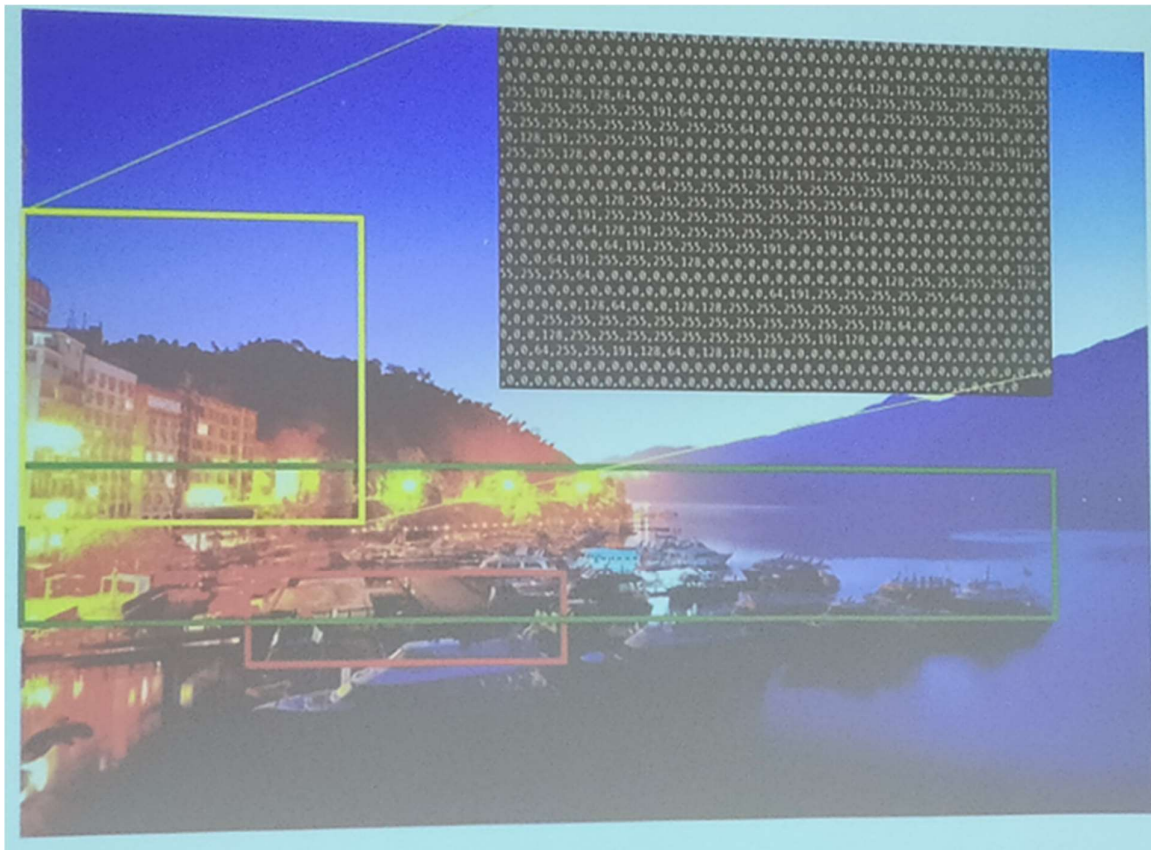
Q: What is the **Scene** ?
outdoor 、 lake side °

Q: Where is this **Place** ?
Sun moon lake 、 pier °

Q: **Objects** in the scene ?
boats 、 **pier** 、 **building** °

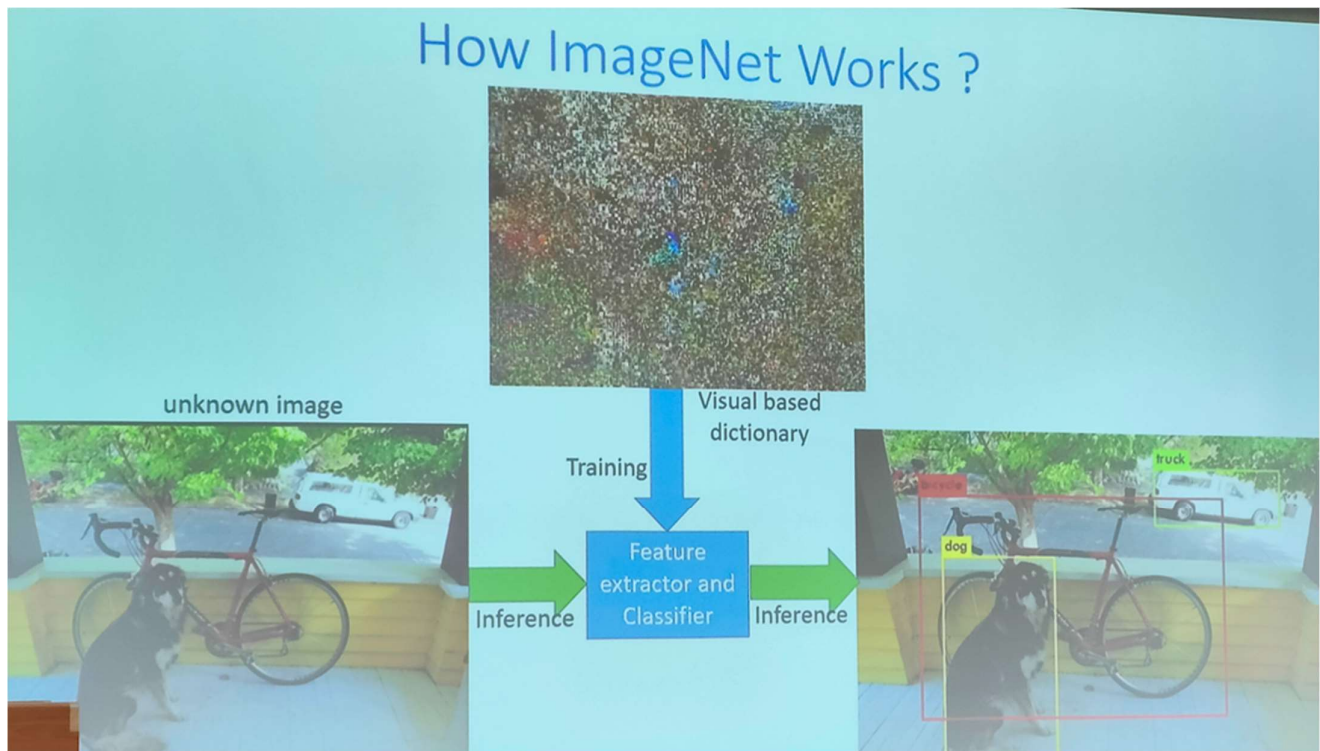
Q: **How far** is the building ?
50 meters °

此張照片對人類來說非常地好判讀，但對於電腦來說，其照片皆為編碼：



ImageNet 的運作原理：

透過 Image 預先標註的圖片(內含 21000 種類別)，用此訓練模型，使其學會各物件間不同的特徵，使其能將未知的圖片偵測其物件。



電腦視覺相關的技術發展歷史：

2010 ImageNet 誕生，是一個大型視覺資料庫，用於視覺目標辨識軟體研究[1]。

2012 AlexNet 誕生，其被認為是電腦視覺最有影響力的論文之一，它加速了深度學習的發展[2]。

2013 R-CNN 誕生，主要用於電腦視覺的機器學習模型，特別是用於物件偵測和定位[3]。

2015 ~ 2022 YOLOv1~v7 陸續誕生，物件偵測的技術逐漸成熟。

R-CNN 其分為三個版本：

1. R-CNN (2013/11)
2. Fast R-CNN (2015/4)
3. Faster R-CNN (2015/6)

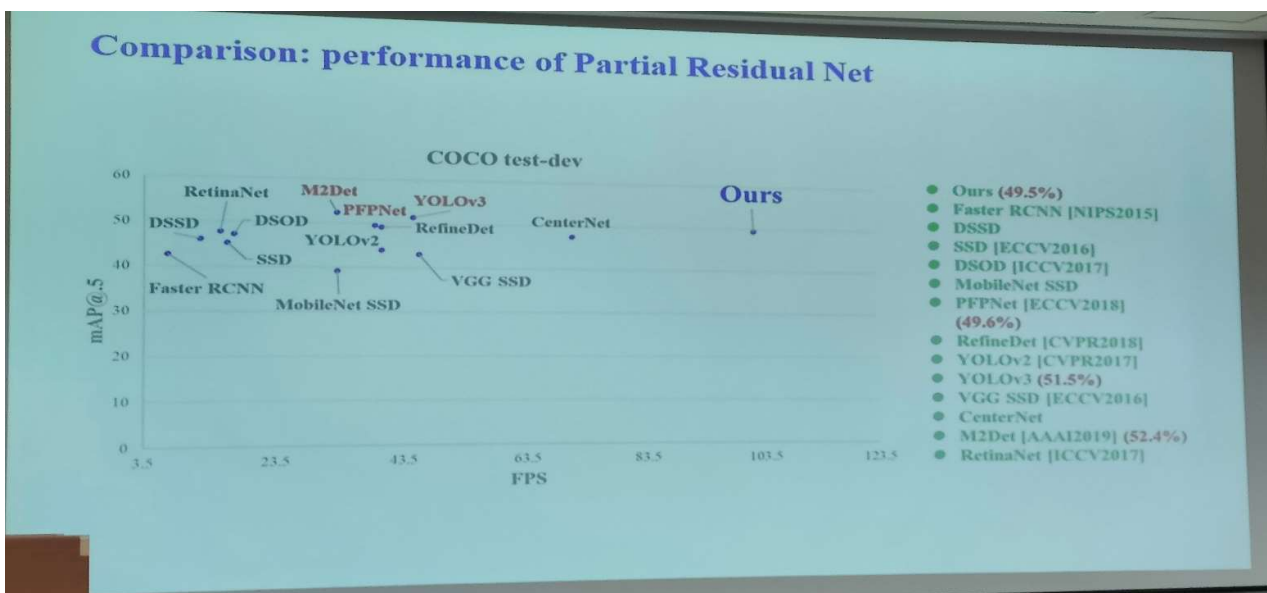
R-CNN 其選取物件流程如下：

1. 對圖像區域透過 Selective Search 產生候選區域。 e.g., 可能有物件的區域。
2. 使用 CNN 模型進行特徵選取。
3. 使用 SVM 進行分類 e.g., 是否為物件。

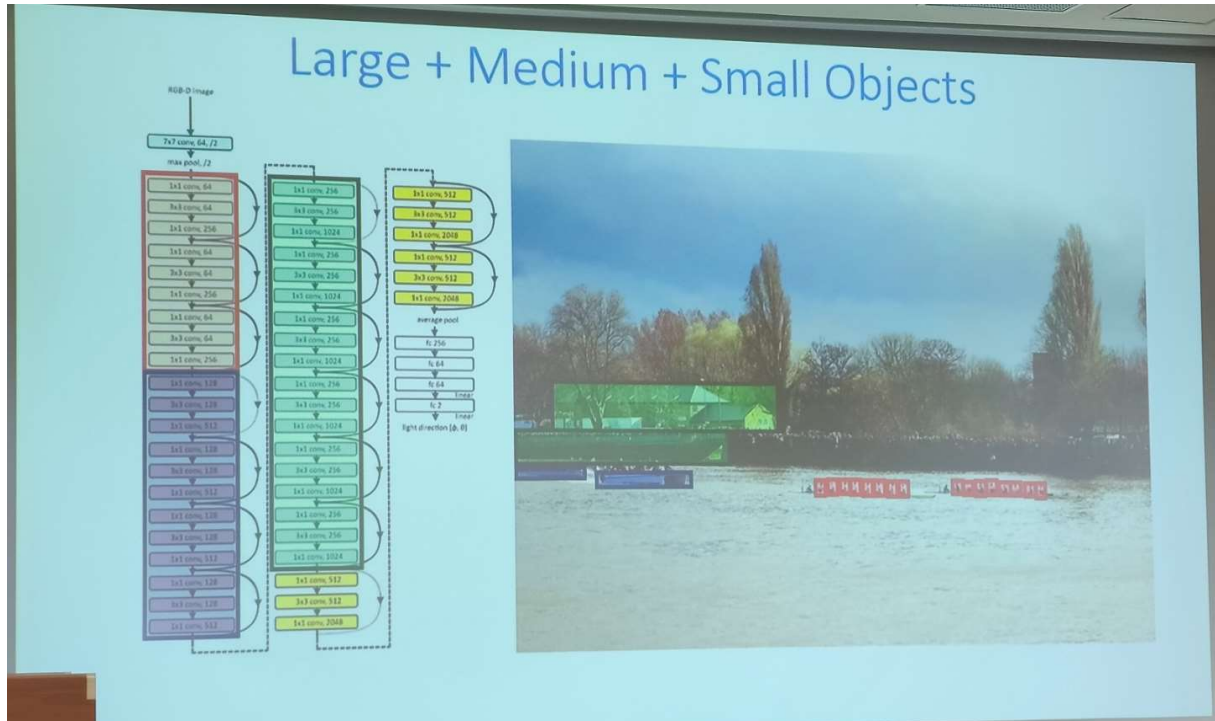
由於硬體的限制，設計模型時須有下列三大考量：

1. 模型必須輕量化，且效能必須快又準。
2. 必須能在各平台部屬 e.g., Cloud、Local、Edge (Cell Phone)。
3. 先 survey 現有的方法。

根據以上考量，先挑選出了 ResNet-50 隨機選出 7 層進行遮蔽處理，使其減少運算的參數量，並且增加模型結構的多樣性，經過這些改動其架構達到了不錯的效果 Map@0.5 為 49.5%，並且 FPS 為 103，是當時最快的方法。



在實際應用中，圖片中可能存在大、中、小物件，分別需要不同的架構進行偵測，因此模型中有個別對應大、中、小物件偵測的架構。



最後向我們展示了 YOLO 在各領域中的應用：

交通：全景十字路口車流偵測、公車死角安裝鏡頭提醒公車司機可能有物件進入到死角。

生物醫學：血球偵測、噬菌體培養皿偵測、細胞有絲分裂檢測、埋葬蟲軌跡偵測。

病蟲害防治：使用 YOLOv7 判斷樹葉上是否有害蟲，並使用雷射進行殺蟲，一秒鐘可以擊殺 7 隻。

今天很感謝中央研究院資訊科學研究所的廖弘源博士跟我們分享了 YOLOv4、v7、v9 的架構及其應用，讓我學習到了很多專業知識，也學習到了做研究的方法，生活的方式(過簡單的生活，有顆單純的心，思考著複雜的問題)，真的受益良多。

參考文獻：

[1] ImageNet. (2023, January 11). Retrieved from 維基百科, 自由的百科全書: <https://zh.wikipedia.org/w/index.php?title=ImageNet&oldid=75472839>

[2] AlexNet. (2025, February 4). Retrieved from 維基百科, 自由的百科全書: <https://zh.wikipedia.org/w/index.php?title=AlexNet&oldid=85953217>

[3] Wikipedia contributors. (2025, January 19). Region Based Convolutional Neural Networks. In *Wikipedia, The Free Encyclopedia*. Retrieved 08:54, April 1, 2025, from https://en.wikipedia.org/w/index.php?title=Region_Based_Convolutional_Neural_Networks&oldid=1270337136

[4] 廖弘源博士 [From YOLOv4,YOLOv7,to YOLOv9]，演講投影片，2025 年 4 月 1 日。