日期：2025/04/01

標題：From YOLOv4, YOLOv7 to YOLOv9

講者：中央研究院 資訊科學研究所 廖弘源 博士

關鍵字：YOLOv4、YOLOv7、YOLOv9、Computer Vision、Object Detection、ImageNet、R-CNN、Edge Detection

台灣中央研究院的廖弘源博士本次演講聚焦於YOLO系列即時物件偵測模型的演進，特別是YOLOv7和YOLOv9的創新與應用。演講者說明在2010年之前未知影像辨識的技術瓶頸到ImageNet的公開在到實際案例：在Nvidia Jetson TX2上計算並擷取交通參數並上傳到雲端從而開始接觸YOLO。

首先從YOLO的起源開始自2016年由Joseph Redmon等人提出成為即時物件偵測的領先框架，演講者有特別提出原作者怕該技術被拿去使用在軍事上所以放棄發展。當時是YOLOv4當紅的時候，前面有提到因為需要在計算量有限的電腦上運行所以講者等人重新設計一套物件偵測系統，並提到設計上會遇到的問題：network architecture、feature integration method、detection method、loss function、label assignment method、training method，並使用Stage level design設計每stage負責不同大小的物件，CSPNet(Cross-Stage Partial Network)讓參數量、連結量、計算量都減少相反讓組合數多樣性增加這更適合在邊緣設備上使用。也介紹Tiny YOLOv4是由VoVNet變化而來使用的是CSPVoVNet其框架考慮參數量、運算量、計算密度，還分析了gradient path，使不同層的權重學到更多樣性的特徵，使推論速度更快、準確度更高。

在設計YOLOv7系統時需要面對的基本問題，Stage層級->network層級，設計ELAN(European Liberal Arts Network)且要適用於邊緣運算以及運算力有限的筆記型電腦，雲端則需要另外設計E-ELAN處理model scaling造成的問題。為提升inference speed而引進implicit knowledge的概念與方法。

其中展示實際應用例如：俯視角路口交通路況辨識及分析、車站混和人流追蹤、國道交通路況辨識、無人機追蹤、在醫學上使辨識紅血球以及噬菌體培養皿監測、X光片骨折檢測等。

YOLOv9基於YOLOv7和Dynamic YOLOv7，引入了可程式化梯度資訊（PGI），這是一種輔助監督框架，旨在解決深層網絡中的資訊瓶頸和可逆函數問題。PGI 在推論時僅使用主分支，避免額外成本，並保持目標任務的關鍵特性，克服了可逆架構、遮罩建模和深層監督的限制。YOLOv9結合了通用高效層聚合網絡（GELAN），這是一種輕量級架構，使用CSPNet塊和計劃性 RepConv替代ELAN。

參考資料：

Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." arXiv preprint arXiv:2004.10934 (2020).

Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023.

Wang, Chien-Yao, I-Hau Yeh, and Hong-Yuan Mark Liao. "Yolov9: Learning what you want to learn using programmable gradient information." European conference on computer vision. Cham: Springer Nature Switzerland, 2024.