

比較平行處理技術與前處理方法對 YOLOv4-tiny

物件偵測模型於性能與精度的影響

Comparing the Impact of Parallelization Techniques and Pre-processing Methods on the Performance and Accuracy of YOLOv4-tiny Object Detection Model

指導教授：何建忠教授

專題成員：莫寶琳

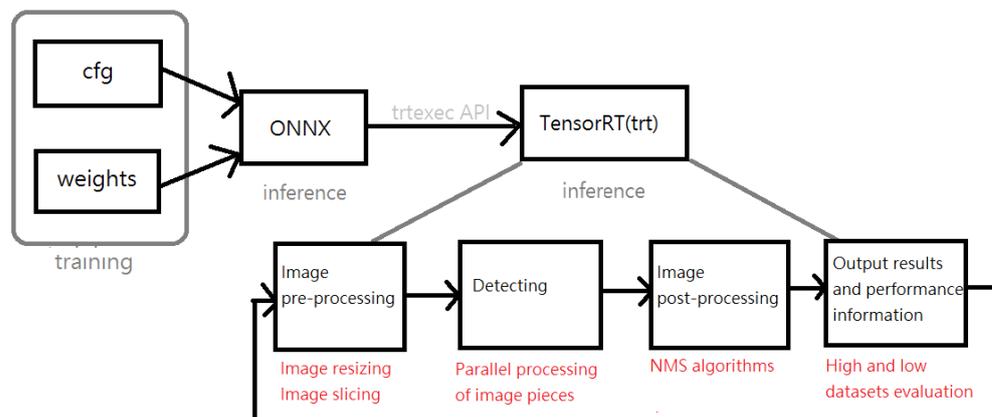
開發工具：Python 3.10.6、Cuda 11.7、cuDNN 8700、TensorRT 8.5.2.2、OpenCV 4.7.0 with GPU、NumPy 1.23.0

測試環境：Ubuntu 22.04、Jetson Nano 4GB

一、簡介：

現今的物件偵測技術發展迅速，隨著這個趨勢，有越來越多的嵌入式應用場景會需要在運算資源以及記憶體資源受限的裝置上運行物件偵測模型。在不同嵌入式系統使用場景，適合它們的執行速度與精確度也會不同，因此，探討不同的 performance 與 accuracy 間權衡的可能性是個有發展性的研究議題。在我的專題中，我使用 TensorRT 將 YOLOV4-Tiny 轉成更輕量的模型、探討物件偵測模型於 Inference parallelization、Image processing、NMS algorithm 的優化方法，以及實作基於高解析度與低解析度影像資料集的驗證環境，未來若延伸此題目，可以繼續使用實驗環境開發。

以下為系統架構圖：

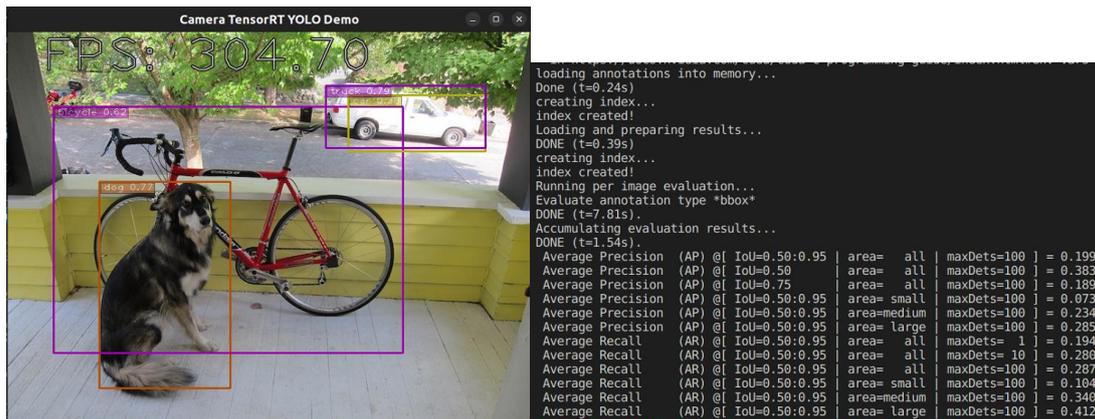


二、測試結果：

圖一圖二分別為 TJU-DHD(high resolution dataset)於 YoloV4-Tiny inference 的 performance 與 accuracy 結果。



圖三圖四分別為 COCO(low resolution dataset) 於 YoloV4-Tiny inference 的 performance 與 accuracy 結果。



Methods	High resolution dataset Performance(FPS)	Low resolution dataset Performance(FPS)
Original version	54	300
Parallel NMS	51	210
Image slicing in 4 pieces	48	180
Image resizing with GPU	45	275
INTER_NEAREST Image resizing	56	310

表格比較不同 Inference 方式在處理高解析度與低解析度影像的速度差異，FPS 值越高代表處理速度越快。