

# Model Description

- 使用YOLOv8為主要訓練的模型
- 實作data augmentation提高模型預測準確率
- 使用YAML定義網路架構，調整超參數train from scratch
- 不使用YOLO的COCO訓練集，不載入pre-trained weight
- 不斷嘗試找出最適合的Hyper parameter

# Implementation

- 考量到VRAM限制12G與其他參數的調整，在嘗試過各種參數量的版本後，最終選擇yolov8m進行training
- Hyper Parameter設定與Data Augmentation: (紀錄分數最高的設定)

150 epoch

imgsize=1920 -> 因為圖片是空拍機去拍攝，很多object非常小，為了讓模型做更好的辨別所以將模型size提高  
batch size=2 -> batch size調低一點，讓模型有更好的generalize能力

mixup=0.1

mosaic=1 -> 常駐開啟mosaic

close\_mosaic=10 -> 最後10 epoch 關閉mosaic

lrf=0.01 cos\_lr=True -> 使用dynamic learning rate調整，使用cosin lr strategy

weight\_decay=0.0006 -> 權重遞減調多0.0001

copy\_paste=0.15 -> 做到類似Oversampling的效果

Patience=15 -> 紿15 epoch機會 若結果不斷震盪無提升則停止training

hsv\_h: 0.015

hsv\_s: 0.7

hsv\_v: 0.4

degrees: 5

將training set中切出10%作為validation set

# Result Analysis



- 將調高之後，對於一些遠距離空拍畫面能有更好的辨識
- 在上面這兩張圖可以看到，Car明顯能更精準的辨識，除了受到Head數量夠多之外，因為car本身體積也大，預測的結果非常好，而person & motorcycle則辨識困難，除了有許多漏框之外，confidence也偏低
- 雖然Hov的量是car的1/20倍，但也因為體積較大的緣故，辨識上都非常準確

# Result Analysis



- 如上圖所示，可以看到左邊第一張圖，person的confidence普遍都不高甚至有很多連0.3都不到，而且畫面的右手邊也都沒有成功辨識出來
- 除了人物過小之外，我也發現**拍攝角度也是會大幅影響辨識結果的變因**，有幾張圖的角度是協拍，預測最好的Car在bounding box and Confidence都表現不好
- 可以看到右邊的圖，有趣的是，雖然Car的資料量非常多，本身體積也很大，但是在這張圖我不管怎麼調，**他都只能框一半的車身，而且confidence都不高**

# Short conclusion

## 實作過程遇到的困難：

1. Data Set除了Long Tailed之外，因為都是空拍機的畫面，尤其是person & motorcycle的confidence都不高，在透過視覺化觀察bounding box也很容易看到漏標或是標不準確，person尤其明顯除了Long Tailed外，可能還要研究Tiny Object Detection

## 2. 在Data Augmentation中遇到一些困難

- 為了解決以上問題，在資料增強嘗試使用Oversampling & 局部放大，像是”**copy\_paste**”並不是做“類別物件的**sampling**”，所以如果要做到oversampling要自行手刻並且要按照yaml格式去實作，過程有嘗試想完成但因為格式一直對不上所以就調整**copy\_paste**參數讓他做到類似的效果
- 局部放大也是，原本想將**training data**做四個角的個別放大讓他更好學細小物件，格式部分處理不好，所以只好使用**imgsize**做到類似的效果，值得一提的是，大部分**training data**都是1920x1080p 如果**imgsize**調得太大，除了VRAM不夠的情況之外，本來的圖片反而會被拉糊，在訓練會起反效果所以最後將**imgsize**調整成1920x1920 在**person&motorcycle**有更高的mAP