

明志科技大學考試命題用紙

109 學年度第 2 學期 期末考試 碩士生 高等類神經網路 課程試題 共 1 頁第 1 頁

注意: ☒ 准使用計算器。 ☒ 准參考筆記或書本。 ☐ 另附答案紙每人 0 張 考試時間: 80 分鐘

命題教師__陳思翰__(簽章)_110_年_5_月_9_日 班級: 碩管一甲 座號: M09218001 姓名__周彥廷__
班級: 碩管一甲 座號: M09218010 姓名__葉庭佑__

請以物件偵測的概念分析工安相關的影像集。

1.請介紹你所下載的影像集。

- (1)資料集名稱與來源為何
- (2)描述影像集的基礎訊息(如各偵測物件的樣貌、辛苦標註的過程、樣本數、各類物件個數)

(1)資料集名稱: Automatic Hardhat Wearing Detection(Helmet_Dataset)

kaggle 連結: <https://www.kaggle.com/abhishek4273/helmet-dataset>

(2)資料集含有 3174 張照片,隨機選出 400 張作為整體資料並進行標註,最後訓練前將影像與標註檔整理如圖 1 所示,影像特徵為工人於作業環境影像,本研究探討工人於作業時是否有穿戴基本安全配備工安議題為主軸,偵測對象主要為 1.安全帽(hardhat),2.安全背心(vest),3.口罩(mask)與 4.人(person)。

本研究透過 YOLOv4 深度學習並建立偵測上述 4 類之演算法(詳細實驗流程於第 3 部分介紹)。

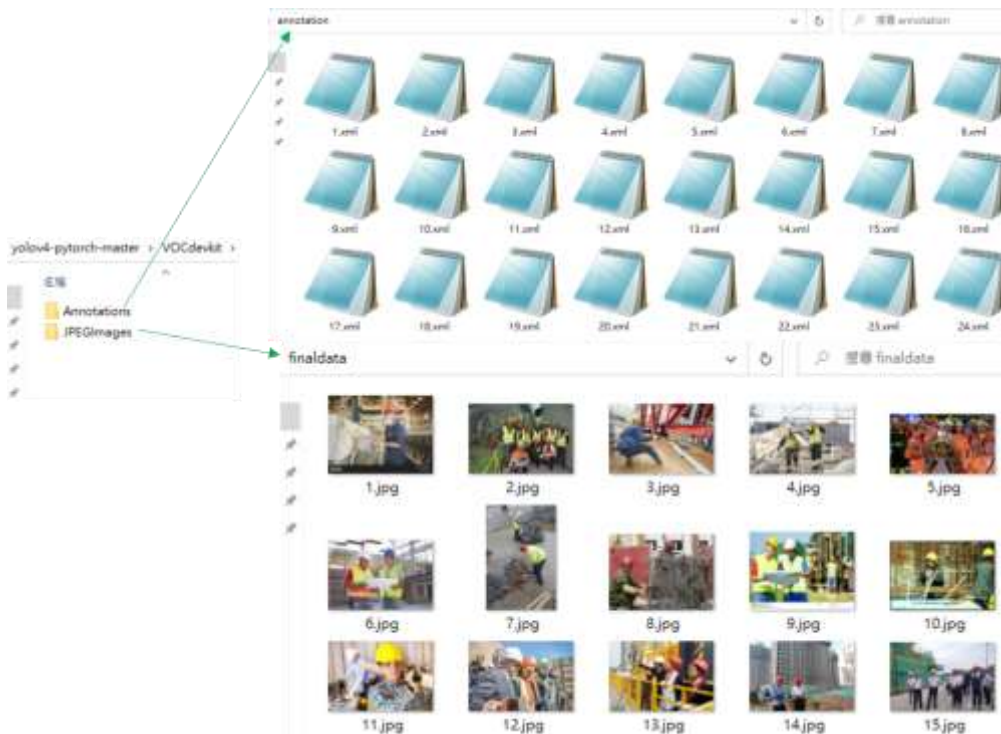


圖 1、影像與對應標註檔整理放置於 YOLOv4 資料夾。

■ 標註過程:

兩人分別標註 200 張影像,如圖 2 所示可以看到偵測物件樣貌,標註工作需要花費許多時間,本資料集主要特徵類別為安全帽與人,有影像會同時出現許多安全帽與人情情形如圖 3 所示,這時候要逐一標註並注意類別是否正確,並於開始訓練前再次確認 4 種標註類別情況是否正確。



圖 2、使用 labeling 分別標註 1.安全帽，2.安全背心，3.口罩，4.人示意圖。

■ 調整標註範圍：

原本標註方式在標註人時會較貼近人的輪廓，這樣標註密集人群時，容易框到同是人的背景，造成人與背景無法有效區分，後續再修正人類別的標註方式如圖 4 所示，標註框儘可能地讓人與背景同時存在，並視情況捨棄人群疊加影像，目的是希望 YOLOv4 能學習區分人與背景，同理亦應用於安全帽和安全背心場景。

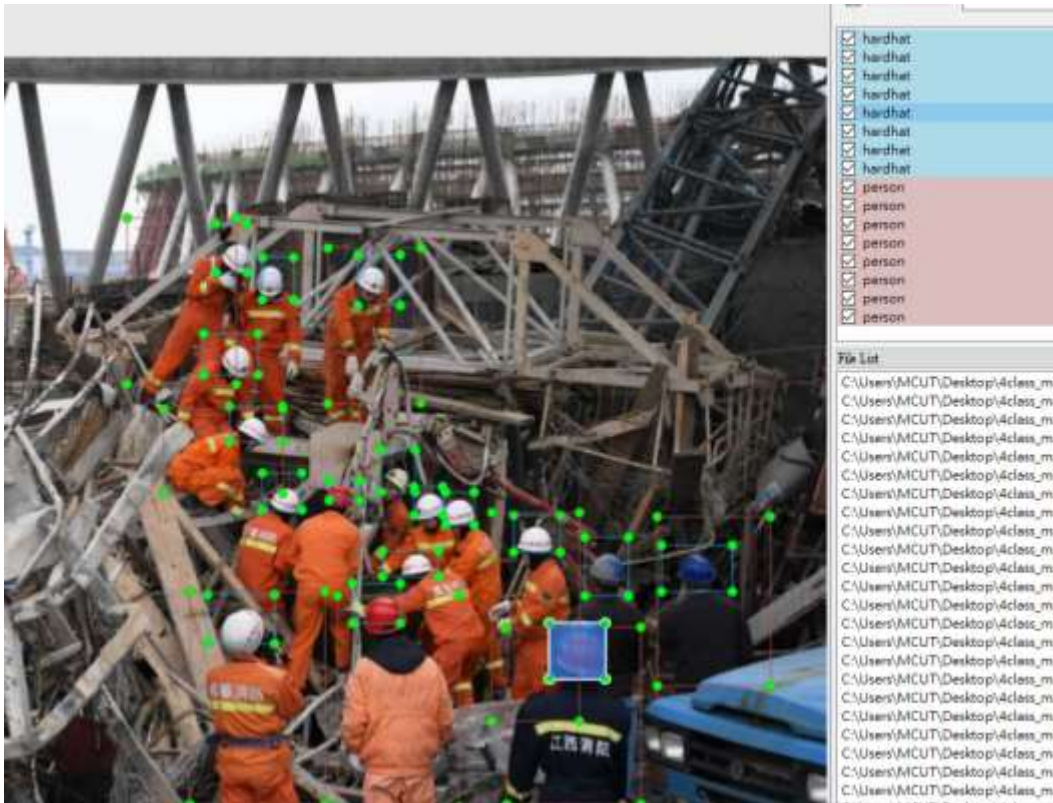


圖 3、同時出現多個安全帽與人影像示意圖。



圖 4、標註改善前(紅框)/改善後(綠框)示意圖。

利用程式碼讀取加總各類別物件個數以 ground-truth 如圖 5 所示，發現整個資料集 mask 樣本僅有 9 個，vest 有 323 個，而 person 與 hardhat 樣本數量接近。

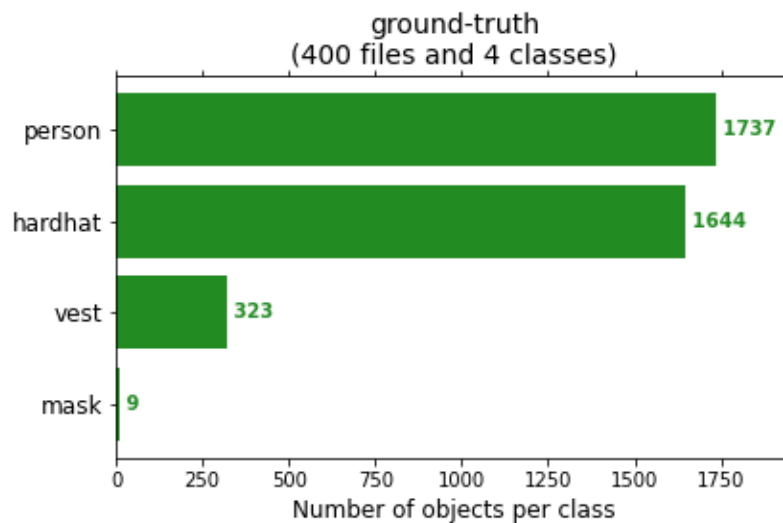


圖 5 顯示 labeling 中所框 4 類別的數量。

2. 建模前請執行前處理。

- (1)請做 50%-10%-40%的資料分割
- (2)是否需做影像前處理
- (3)是否需做影像擴增

(1)資料集分割

根據題目指示切分資料集為訓練資料(含 valid data 60%)與測試資料(40%)，並在 voc2yolo4.py 進行 trainval_percent = 0.6，儲存後執行會進行資料切分。

test.txt	2021/5/23 下午 05:41	文字文件	3 KB
train.txt	2021/5/23 下午 05:41	文字文件	4 KB
trainval.txt	2021/5/23 下午 05:41	文字文件	4 KB

(2)沒有作影像前處理。

(3)本研究使用色彩空間轉換(Hue、Saturation、Value，HSV)和高斯模糊(Gaussian Blur)進行影像擴增。

訓練資料集由原本 240 張影像擴增為 720 張影像，將原圖與處理後影像比較如圖 6 所示。

這樣可以達到影像擴增 2 倍數量與不用重新標註以節省時間成本。



圖 6、顯示原圖，轉換色彩空間與高斯模糊差異。

3.建模與解釋。

■ 實驗設計

實驗流程如圖 7 所示，mAP 為初次測試訓練成果(尚未進行後續優化步驟)，由 4 個類別(hardhat、vest、mask 以及 person)縮減至訓練 2 類(hardhat 與 person)。

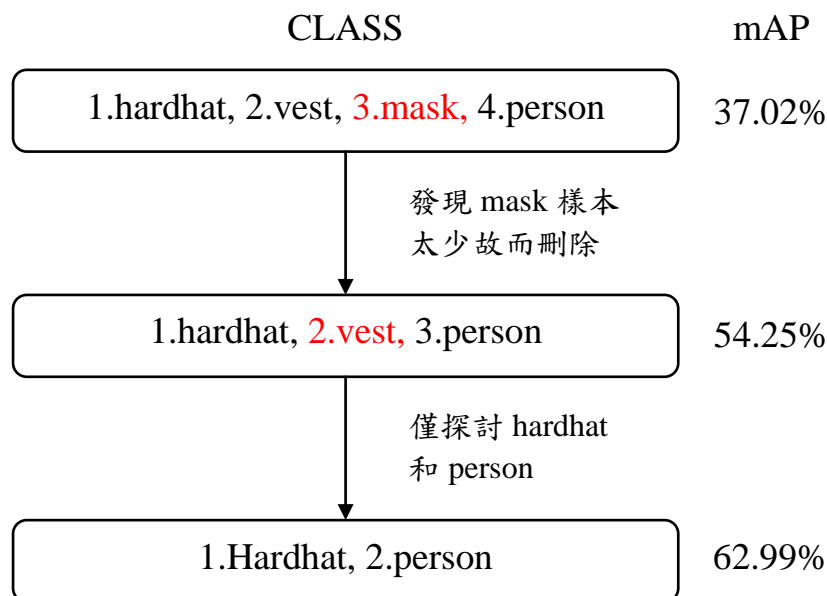


圖 7、實驗設計流程與初步 mAP 結果。

(1)選擇 1 種物件偵測模型以訓練影像集進行建模

如實驗架構所示：

Step1.本研究一開始對 4 個類別進行訓練，結果發現 mask 類別樣本太少且 mAP 為 37.02 如圖 8 所示，因此將 mask 類別刪除改為 3 類。

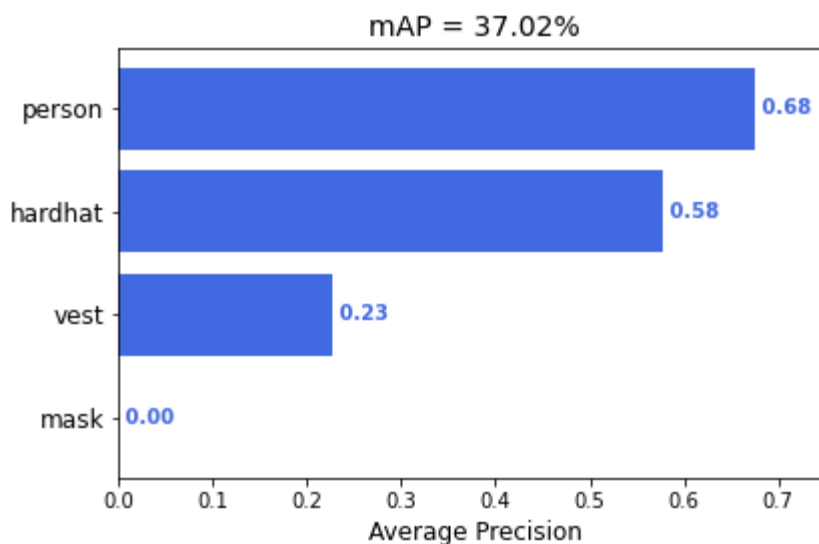


圖 8、測試資料在(1.hardhat，2.vest，3.mask 與 4.person)表現。

Step2.訓練 3 類別，mAP 有顯著提升至 54.25%，後續經過超參數微調與增加訓練步數(epoch)後，mAP 達到 64.31%如圖 9 所示。

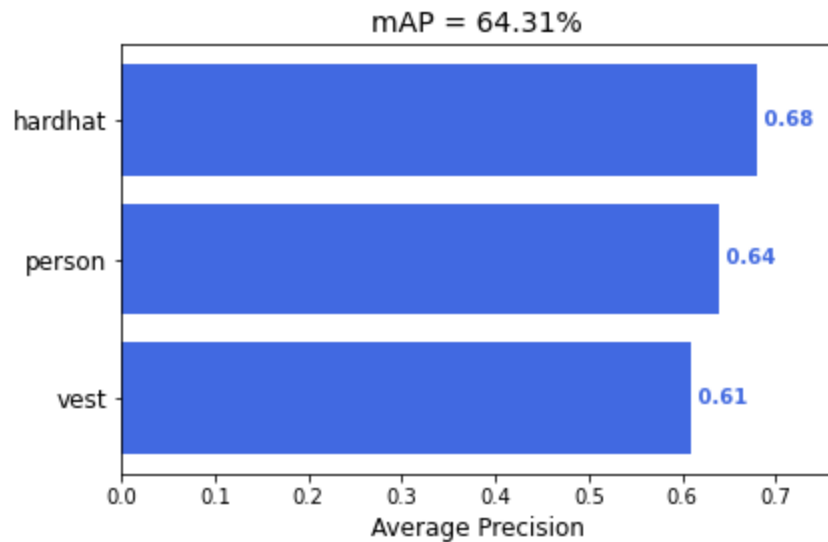


圖 9、測試資料在(1.hardhat，2.vest 與 3.person)經過優化表現。

Step3.最後僅探討 hardhat 和 person 兩類並進行超參數微調與改善策略。

(2)模型輸出的 3 維張量(Tensor)尺寸為何

YOLOv4 沿用 YOLOv3 (head)概念:3 個尺度 feature map 分別對每一個 grid 預測 3 個 bounding box (對應 3 個 Anchor)，每個 bounding box 預測出 5 個值(x, y, w, h 以及 confidence)，輸入影像尺寸為 608×608×3，類別數(C=2)，head 輸出張量為 19x19x3x(5+2)，38x38x3x(5+2)和 76x76x3x(5+2)。

(3)何謂 mean average precision (mAP)？請解釋該指標如何運算

AP(average precision)會考量精度(precision)與召回率(recall)兩種指標，對每一個位置精度做計算然後取平均，mAP 為各類別 AP 值加總除以類別數，數值越大代表模型性能越好。

混淆矩陣(Confusion Matrix)可以得出 TP、FP 與 FN，計算精度與召回率。

		True Condition	
		Positive	Negative
Predicted outcome	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$\text{mAP} = (71.48 + 64.51) / 2 = 68.00\%$$

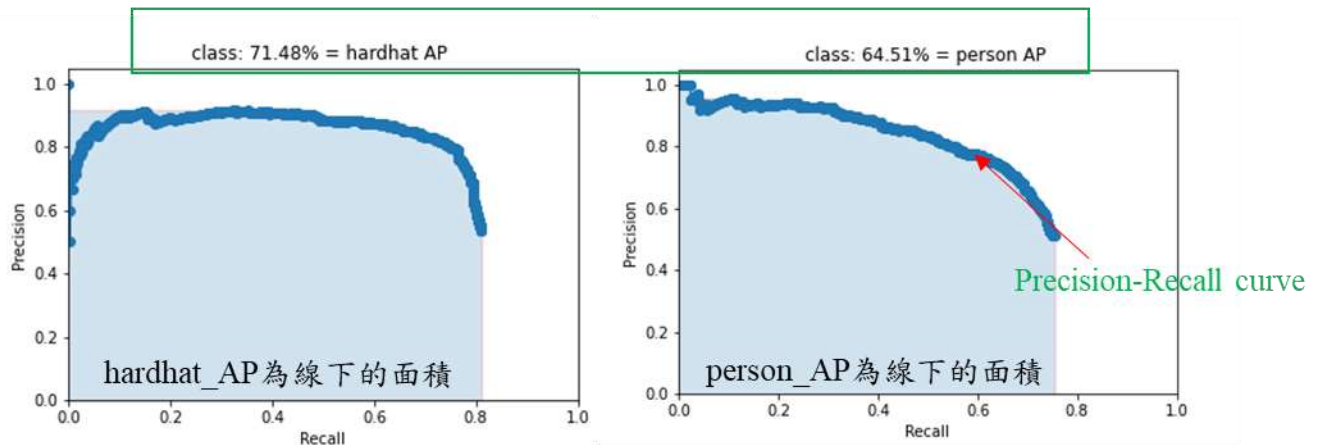


圖 9、計算 mAP 與顯示 AP 為線下的面積。

IOU(Intersection over Union)計算方式如下：

$$\text{IOU} = \frac{\text{面積交集}}{\text{面積聯集}}$$

(4)請以驗證 mAP 或其他指標做些許超參數微調實驗(請參考表一與表二)

表一、4 類、3 類與 2 類初步訓練流程與超參數調整。

實驗項次	類別數	Mosaic	Cosine	Smooth	epoch	k-means_anchors	mAP(%)
1	4	✓			100	✓	37.02
2	3	✓	✓		100	✓	54.25
3	3		✓	0.005	200	✓	64.31
4	3		✓	0.005	300	✓	63.95
5	2		✓	0.005	300	✓	62.99

表二、使用遷移學習與影像擴增方式進行優化。

實驗項次	類別數	Mosaic	Cosine	Smooth	epoch	k-means_anchors	mAP(%)
6	2	✓	✓	0.005	300	✓	63.83
7	2	✓	✓	0.005	300		66.87
8	2				300		68.00
9	2	✓	✓	0.005	500		65.04
10*	2				200		67.09
11*	2	✓	✓	0.005	500		89.85

10*使用遷移學習方式訓練

11*使用影像擴增方式訓練

指標計算參數(iou 與 confidence 設定 0.5)

■ 結果與討論：

表一進行實驗 batch_size 皆設定為 2，學習率設定(1e-4)，表一顯示訓練類別為 4 時，mAP 效果為 37.02，刪除 mask 類效能提升至 52.41%，進行馬賽克影像增強(Mosaic)與餘弦退火(Cosine)策略進行改善，最終 3 類別最高 mAP 為 64.31%，後續選定 hardhat 和 person 訓練並嘗試加深 epoch 次數和影像擴增方式使模型效能提升效能。

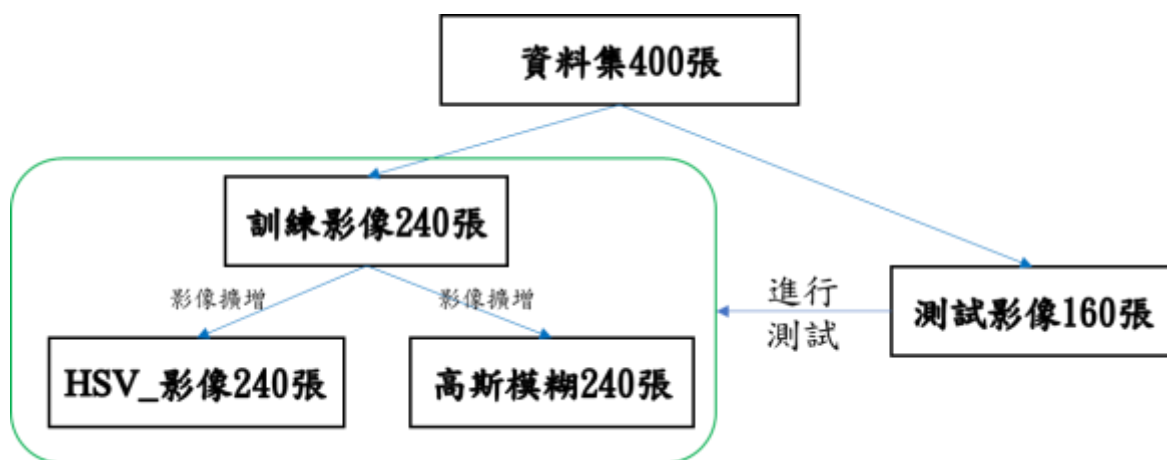
表二進行實驗使用 epoch 為 500 時，凍結層前 250 步 batch_size 設定為 4，學習率設定(1e-4)，解凍訓練後 batch_size 設定 2 與學習率設定(1e-4)，由表二顯示採取各種優化策略得到最高 mAP 有 68.00%且發現使用預設 anchor 較 k-means_anchor 的效果較佳。

遷移學習：

目前 2 類別最高 mAP 為 68.00%，藉由已訓練 300 步最低 loss 權重檔並透過後續訓練 200 步改善效能，結果顯示後續再接 200 個 epoch 之 mAP 並無提升。

影像擴增：

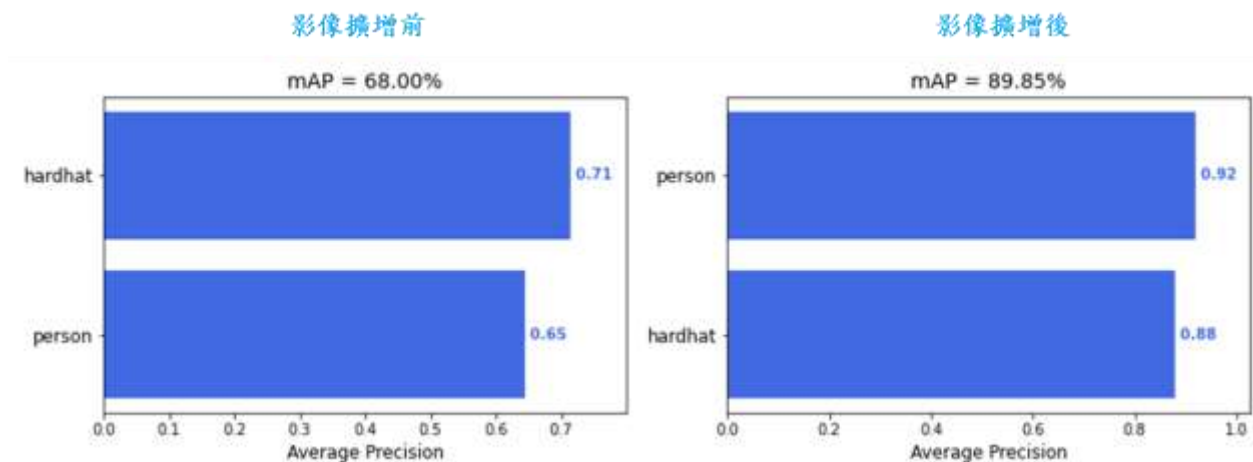
影像擴增使用 1.將原圖色彩空間轉換 HSV 與 2.原圖轉換高斯模糊兩種方式進行如下：



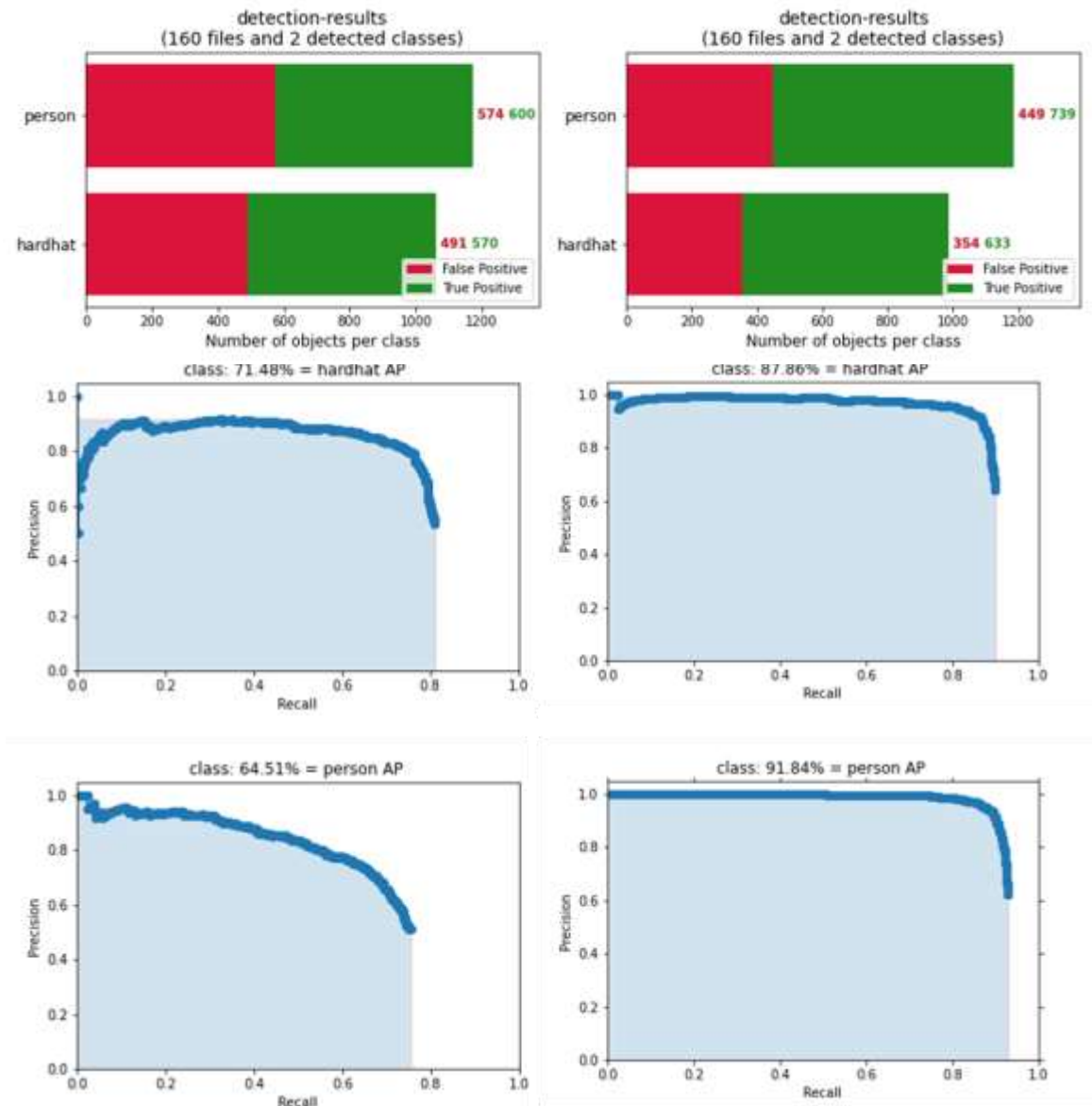
使用影像擴增方式增加資料量，並設定 epoch 為 500，mAP 由 68.00%提升至 89.85%，顯著改善演算法 hardhat 與 person 偵測效能。

(5)以測試影像集計算測試 mAP，針對 160 張影像進行驗證結果如下。

- 比較 hardhat 和 person 的 mAP。



- 顯示偵測結果 TP 與 FP 的數量



(6)改善成果比較：

由(5)可以發現在影像沉擴增後，hardhat 和 person 各自 AP 皆有所提升，hardhat 由 71.48% 提升至 87.86%；person 由 64.51% 至 91.84%，經過平均後 mAP 為 89.85%，在 detection-result 顯示綠色面積 True Positive 有所改善，以下為成果呈現：

1.偵測密集或小物件人影





2. 正確辨識安全帽並辨識小物件人影

案例 1



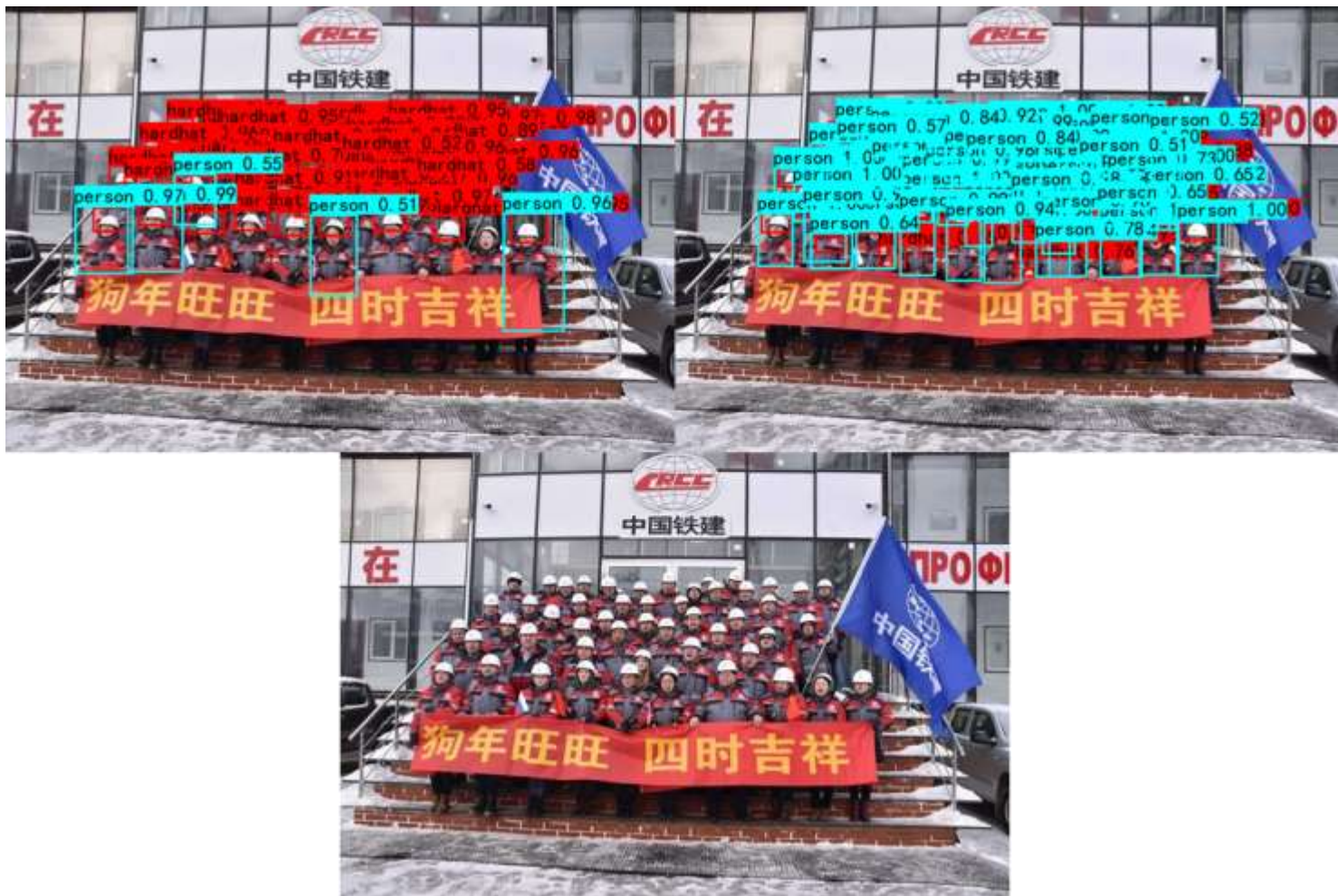
案例 2



案例 3



3.團體照(安全帽與人群)



4.辨識不同人姿勢影像與小物體



5.即時偵測(Real-time Detection)



結語：本研究透過 YOLOv4 深度學習方式建立辨識人和工業安全帽影像演算法，並透過超參數調整與改善策略提升 mAP，希望未來能投入場域進行試驗，偵測人員有無穿戴安全護具以促進場域環境安全與節省人力成本。