

基於機器學習優化遊行人數之小物件偵測與計算

Headcount calculation for parade Based on Machine Learning

組別：B129 指導教授：黃朝宗 組員：蔡湄萋、吳宗哲、陳冠瑋

報告摘要

在民主化的社會中，群眾遊行為人民表達意見與訴求的方法之一，但各方統計方式與立場的不同，往往會產生截然不同的數字，令人無所適從。因此，本專題將使用物件偵測(Object Detection)的技術並加以優化，嘗試估算出遊行影像中的準確人數。

物件偵測是運用機器學習的技術，然而偵測微小物件對此技術而言是一大挑戰。我們分別以兩個對小物件偵測效果較佳的 YOLOv4 模型和 Faster R-CNN 模型進行實驗，針對此研究設計了三種資料集以進行模型訓練，並比較這三種資料集對 YOLOv4 模型和 Faster R-CNN 模型的訓練結果之差異。但實際測試遊行影像後，僅能偵測到人頭較清晰的部分，因此我們進一步提出分割圖片的方法，使其更加適合應用於人頭大小不均的影像上。接著，分析背景複雜度對模型偵測的影響，以選出在多數情況下皆能有較佳表現的模型。最後，將最佳的物件偵測模型之偵測結果經過擬合後，得出估算遊行人數的公式及合理的誤差值，並將其結合攝影機，實現遊行群眾的即時偵測與人數估算。

報告內容

目前常見的遊行人數計算方式主要有兩種，其一是直接派駐人力在遊行現場駐點統計，然而這種方法相當耗費人力，也可能無法顧及所有遊行現場的狀況，再加上計算上的人為誤差，導致此方式浪費人力又容易失準。另一種則是以人群密度與面積計算的方式得出遊行人數，然而遊行路線上每一處的人口密度不盡相同，在統計上也會有一定的落差。

近期機器學習已成為一項熱門的研究，且也有一定程度的發展，因此本專題將以機器學習為基礎，並加以優化改良，嘗試做出一個偵測模型以改善傳統遊行人數計數的方法，不僅不會耗費人力，也不必擔心人群密度不均，只要拍下遊行的照片就能計算出遊行人數，解決遊行人數計數不準確的問題。

一、系統設計

我們首先依據權重製作三種資料集，並用這三種資料集分別訓練 YOLOv4 與 Faster R-CNN 兩種模型，因此共會獲得六種模型結果，再將訓練完的六種模型進行分割實驗，最後選出一個最佳的模型應用於即時影像偵測。

1. 資料集處理

我們主要使用 CrowdHuman Dataset，並取出標記為 head box 的屬性作為訓練的資料集，但由於此資料集並不是全部都是密集人群的圖片，因此我們製作了三種資料集以符合我們的偵測目標並比較其差異。第一個資料集選用 CrowdHuman Dataset 所有的資料，並使每張圖片權重都相同，我們將其取名為 All Dataset；第二個資料集選用 CrowdHuman Dataset 中人數大於 40 人的圖片，我們將其取名為 Pick Dataset；第三個資料集選用 CrowdHuman Dataset 中所有的圖片，但使出現於 Pick Dataset 中的圖片權重為其他圖片的五倍，意即人數較多的圖片權重較高，我們將其取名為 Weighted Dataset。

以 7：2：1 的比例將 Pick Dataset 分為 Training Data、Validation Data、Test Data；對於 All Dataset 及 Weighted Dataset 則以相同比例 7：2 分為 Training Data 及 Validation Data，並使三種 Dataset 的 Test Data 完全相同，以便使用平均準確率(AP)比較其差異。

我們使用 YOLOv4 與 Faster R-CNN 兩種模型分別對於這三種資料集做訓練，因此總共會有六種訓練好的模型用來進行圖片分割實驗。以 AP(IOU=0.5)及 AP(IOU=0.5:0.95)作為比較依據，YOLOv4 訓練結果中，All Dataset 是三者中數據表現最佳的；Faster R-CNN 模型訓練結果則以 Weighted Dataset 相對表現好一些。相同資料集的情況下，兩種模型之間的比較則是 Faster R-CNN 的表現較佳。

2. 分割方式

由於三種資料集所訓練的模型分割結果非常相似，因此我們在報告中僅以 All Dataset 訓練的模型進行比較。

圖 1(左)為未經分割處理的兩種原始模型直接偵測的結果，圖片左上角標籤為偵測框之數目，可以看出此模型僅能偵測出圖片前半部人頭較大較清晰的部分，因此我們嘗試將圖片分割為十六等份，分別偵測後再將每一等份合併，以提高模型偵測的準確率，兩種模型的結果如圖 1(中)，由實驗可知此方式僅能針對圖片中央人頭為特定大小的部分有較高的準確率，然而常常會因拍攝角度不同而導致一張圖片中人頭大小有所差異，導致模型偵測的效果不佳。

為了解決此問題，我們設計了最佳的分割方式以優化模型準確率，圖 1(右)為兩種模型應用此分割方式的結果，可以看出改良後的分割方式不僅可以偵測圖片後半部人頭較小的部分，對於前半部人頭較大的部分也能有較高的準確率。



圖 1 YOLOv4(上)與 Faster R-CNN(下)不同分割方式前後比較

改良後的分割方式先將圖片分割為九等份，再依據每一等份 bounding box 的數量與所占面積比例決定是否再將其分割為四份。圖 2 為本次改良後的分割方式之流程圖。

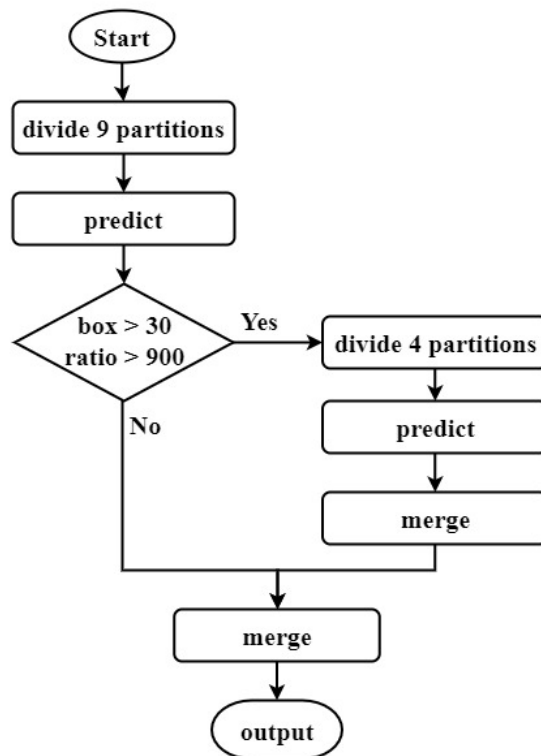


圖 2 改良後分割方式流程圖

二、研究成果

1. 實際測試成果

我們另外挑選出八張遊行圖片，分別以六種模型偵測每張圖片的人數，手動計算出三種資料集所訓練的模型各自的平均 Precision 及平均 Recall 數值，以比較六種模型實際測試的成果。

與 Faster R-CNN 在數據表現上優於 YOLOv4 相異，實際以遊行圖片測試，反而是 YOLOv4 的整體表現都優於 Faster R-CNN，其中又以 All Dataset 和 Weighted Dataset 的模型表現最佳，且 Faster R-CNN 模型所需的偵測時間較長，較不符合即時影像偵測的目標，因此我們最後選擇以 YOLOv4 模型進行進一步的探討。

模型的表現可能也會受背景複雜度影響，因此我們使用 YOLOv4 中的 All Dataset 和 Weighted Dataset 模型，分析圖片差異所影響偵測的原因，並從中選出一個最佳的模型進行公式計算及應用於即時影像偵測。我們使用圖 3 均勻背景圖片(左)與複雜背景圖片(右)分析圖片背景差異與偵測之相關性。均勻背景為背景雜物較少、人頭大小差異較小之圖片；反之，複雜背景為背景雜物較多、人頭大小差異較大之圖片。

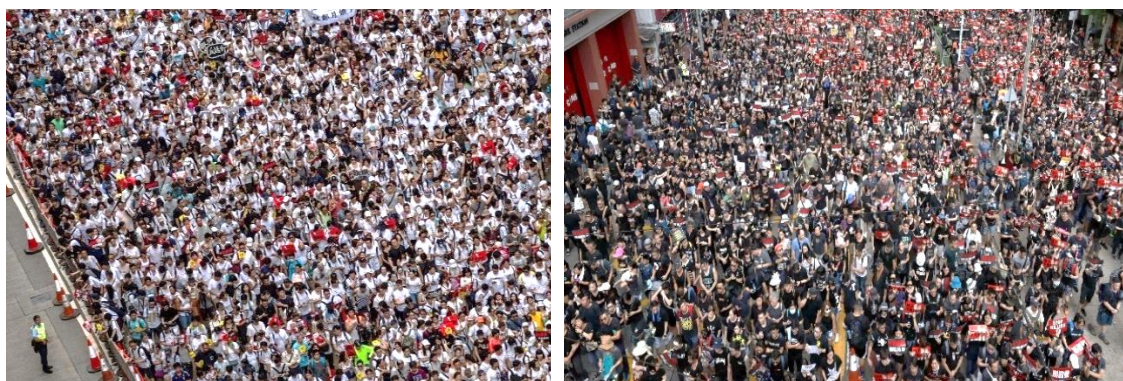


圖3 實驗用之均勻背景(左)與複雜背景(右)圖片

從實驗結果得知，準確率(Precision)都接近百分之百，但均勻背景的召回率(Recall)都高於複雜背景。在均勻背景的情況下，All Dataset 模型能比 Weighted Dataset 模型高出近 10%的召回率；以複雜背景圖片實驗時，All Dataset 模型的召回率也能稍微高於 Weighted Dataset 模型。綜合以上，我們將選用 All Dataset 的模型計算公式，並結合攝影機進行即時影像偵測。

2. 擬合公式與即時影像應用

由於上述實驗結果之召回率尚未達到我們預設的目標，若直接將此模型應用於遊行影像，其偵測結果將無法準確估計實際人數，因此我們想要嘗試透過擬合公式找出偵測人數與實際人數的關係，以降低模型的誤差。

以 All Dataset 所訓練的 YOLOv4 模型將八張測試圖片的偵測結果進行擬合，擬合結果如圖 4，其中 x 軸為模型偵測之人數、y 軸為實際人數。

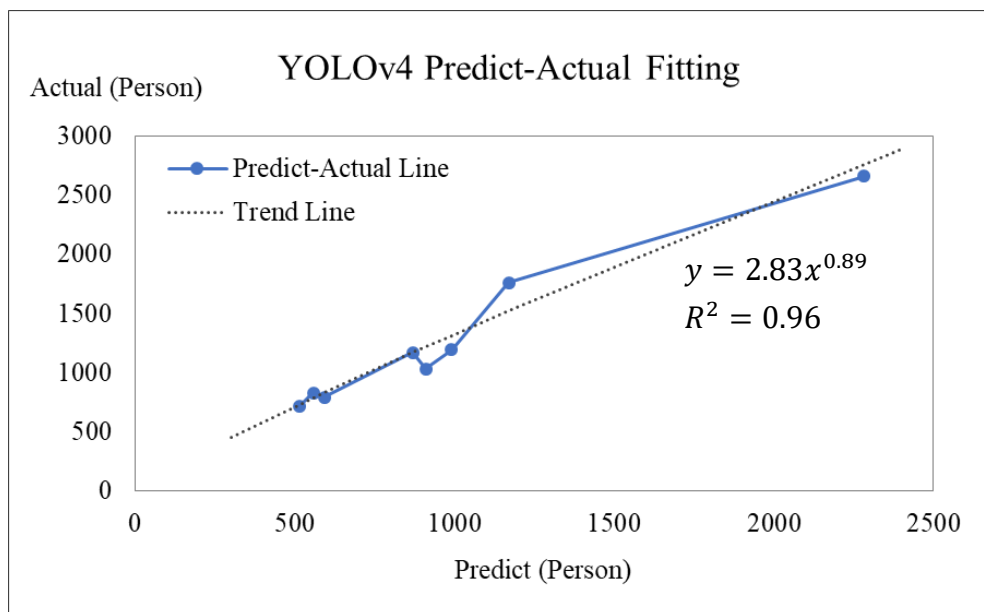


圖 4 YOLOv4 模型之擬合公式圖

由此圖我們得出公式為 $y = [2.83x^{0.89}]$ ，以及此公式的決定係數，亦即相關係數的平方為 0.96，並計算出的人數誤差值為 $\pm 5.1\%$ 。

為了定義此模型的適用範圍，我們將此模型應用於草地音樂祭與教室進行即時影像偵測，人數偵測結果如圖 5，括號內為未經計算的偵測人數。圖 5(左)為草地音樂祭的即時影像偵測，實際人數經由手動計算後得到 549 人，計算誤差值為 2.55%；圖 5(右)為教室的即時影像偵測，因為人數過少且人頭太大，不僅偵測上因為分割而增加錯誤率，預測人數也因為擬合公式被放大將近兩倍，可以得知我們設計的模型僅能運用於實際人數大於 500 人的影像，對於人數較少的影像則無法準確估算。

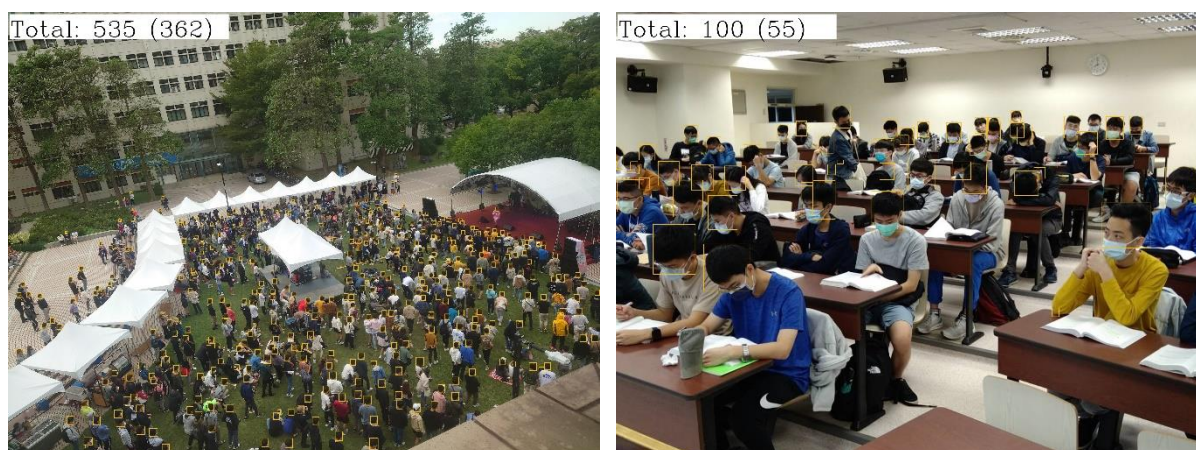


圖 5 即時影像偵測實際測試圖

心得感想

經過兩個學期的實作專題訓練，最大的收穫就是自主學習能力的提升，以往的課程大部分都是教授直接給需要學習的內容，屬於比較被動式的學習，但在實作專題中大部分都是靠自己完成，包含要尋找資源、訂定目標等，也發覺網路上的資源有很多都不一定都是正確的，需要自己去實驗加以驗證，進而篩選出正確的資訊。

在實作專題初期，為了要確立研究方向而閱讀了大量的文獻資料，去比較不同的模型在不同的領域有什麼優缺點，選出適合本次專題目標的模型；中期開始訓練模型，在選擇與處理 dataset 的期間，因為每個 dataset 都有專屬不同的標記檔名，會遇到檔名格式與模型需求不符的問題，需要去尋找可以正確轉檔的工具，過程中常常會徒勞無功，不是每一次的嘗試都會有進展，但這些錯誤的經驗會成為我們的養分，在最後問題迎刃而解的同時，使我們有所成長。

不同於一般課程都有的標準答案，實作專題的結果並不是固定的標準答案，要有怎樣的成果取決於我們想要達到什麼目標，我們不停地討論與實驗，盡我們所能的努力，將所有問題一一排除，從一開始苦惱該如何在圖片上顯示偵測框並計算總數，一步一步的修正與優化讓偵測框越來越多，錯誤率也越來越小，即使最後呈現的結果不如預期中的完美，但看到每一張圖片上面滿滿的偵測框，也會感到有所成就，這或許正是研究最重要的價值。