­­­­­­­逢 甲 大 學

資訊工程學系

專題研究報告

使用FairMOT計算屋內人數

Using FairMOT calculating population indoor

指導教授：張哲誠 教授

學生：D0782956 平祖安

D0746384 李筱文

D0746353 岳 靈

D0746413 李琛瑤

D0745782 杜怡嫺

中華民國一百一十年五月

**摘　　要**

自從電腦在1950年代被發明後，科學家就一直在思考，如何讓電腦變得和人類一樣聰明？如何能代替人們進行所有的工作？時至今日，有些工作可以交給人工智慧可以代勞，然而，將人物特徵檢測出來，並維持同一個編碼，且依靠追蹤技術，確保遺失資料的補償方式，要做到即時性的處理，依然是一個大問題。

這篇論文主要研究即時性問題，提出一個系統，可以連結影像串流，和Line Notify API，並傳送即時情況給場域使用者。

這篇論文使用基於Jointly Detector and Embedding model (JDE)架構所延伸的Fairness of Detection and Re-Identification in Multiple Object Tracking (FairMOT)，提出一套監控系統，不僅用即時性影像取代原本用影片做為輸入，也可直接觀察資料結果而不用在經過轉為影片處理，最大重點是他還能做到即時性。

關鍵詞：多目標物件追蹤、特徵編碼、即時性處理

**Abstract**

While computer invented, scientists wonder how to make computer intelligent as humanity and how to replace manual works with it. Until up to now, we have installed artificial intelligent to replace manual labor. However, scientists are troubled by real-time for detecting human feature and Re-ID and confirming compensation for lost data.

This thesis mainly studies real-time and proposed a system connected with stream from camera and send message concluding currently population in the field we maintained by Line Notify.

This thesis referenced Fairness of Detection and Re-Identification in Multiple Object Tracking (FairMOT) based Jointly Detector and Embedding model (JDE) architecture and proposed monitored system is able to detect immediately data instead of movie input that the source code provided proffered and also directly observed data instead of squandering performance for transforming into movies. And the most important , it accomplished real-time.

Keywords: anchor-free, FairMOT, JDE, Re-ID

**目　　錄**

[第一章 緒論 5](#_Toc71373701)

[**第一節 研究動機** 5](#_Toc71373702)

[**第二節 研究目的** 5](#_Toc71373703)

[**第三節 名詞解釋** 6](#_Toc71373704)

[第二章 文獻探討 7](#_Toc71373705)

[**第一節 Jointly learns the Detector and Embedding model** 7](#_Toc71373706)

[**第二節 FairMOT** 7](#_Toc71373707)

[第三章 研究設計與實施 9](#_Toc71373708)

[**第一節 研究方法** 9](#_Toc71373709)

[**第二節 將鏡頭作為輸入** 9](#_Toc71373710)

[**第三節 Line notify** 11](#_Toc71373711)

[第四章 實證結果與分析 12](#_Toc71373712)

[第五章 結論 13](#_Toc71373713)

[參考文獻 14](#_Toc71373714)

**圖目錄**

[圖 1 camera.py內容 9](#_Toc71371792)

[圖 2 loadvideo修改內容 10](#_Toc71371793)

[圖 3 line notify內容 11](#_Toc71371794)

# 緒論

**第一節 研究動機**

監控系統已廣泛的應用，一般需要在特定之區域範圍做監控錄影，當有犯罪或異常事件時，即可調閱影檔做為破案線索，找出可疑之人事物。有許多大場合，外面都會有技術人員進行數量確認，而當人流大量時，對於人員是十分難跟上，有時會產生極大的誤差。兩者問題都是花費大量人力，但不一定會有期待的結果，因此我們提出一個監控系統，可以在任何時候告訴我們場域內的人數。

**第二節 研究目的**

根據研究動機，本論文研究可分為：

1. 探討當有多物件時，人物追蹤的方式，以及計算人流的方法
2. 為了能隨時掌握場域的情況，需確保做到何時都能了解現況的即時性
3. 當有人的資料遺失時，能重新抓回與否，以及能否再判定為同一個人

**第三節 名詞解釋**

1. JDE
   1. Anchor based

規劃多種比例的框，並以框進行損失函數，以節省運算成本的檢測能力，大幅度地帶領著開發與研究的方向。

* 1. Region Proposal Network (RPN)

是一種 Convolution Network ，Input 是之前 CNN 輸出的 feature map，feature map 上取 sliding window，每個 sliding window 的中心點稱之為 anchor point，輸出是一個 bounding box 以及該 bounding box 包含一個物體的機率。

1. FairMOT
   1. CornerNet

CornerNet為一種Anchor-free的目標檢測演算法。

* 1. Anchor free

與Anchor based不同，是以點去尋找框，省去設置不同比例的Anchor及一些參數的麻煩。

# 文獻探討

**第一節 Jointly learns the Detector and Embedding model**

傳統的多目標追蹤採用的是Separate Detection and Embedding (簡稱SDE)，先進行目標檢測，再進行特徵提取與物體關聯，兩個部分是完全分開的。而Towards Real-Time Multi-Object Tracking這篇論文提出了另一種概念，Jointly learns the Detector and Embedding model(簡稱JDE)，其做法將目標檢測與特徵提取在同一個網路中訓練，原因是這種做法可以讓兩個分支共用一部份的特徵，提升速度。在論文中，實驗結果展示了JDE確實大幅提升了運行速度，略微的準確度下降(Wang Z., Zheng L., Liu Y., Li Y., Wang S., 2020)。

**第二節 FairMOT**

JDE相比SDE速度確實提升了不少，不過相對的犧牲了準確率，FairMOT (Zhang, Y., Wang, C., Wang, X., Zeng, W., & Liu, W. ,2020)針對這點提出幾個原因，一是檢測採用的Anchor-based方法，二是在re-ID除了需要高維度的特徵，也需要低維度的特徵，三是較低維度的特徵更適合JDE這種融合檢測與特徵提取的演算法。

文中提到JDE方法大多採用Anchor-based，而從Anchor中心提取特徵，這會有幾種狀況，一個Anchor中可能包含其他干擾或背景，影響準確率，或是相鄰Anchor，只要IOU夠高可能辨識同一身分，造成training的模糊性。並且Anchor-based方法的feature map通常是8/16/32倍的下採樣產生，這樣的分辨率會降低re-ID的準確率。所以FairMOT在檢測部分採用CenterNet(Zhou, X., Wang, D., & Krähenbühl, P. ,2019)。

CenterNet(Zhou et al. ,2019)從CornerNet( Law, H., & Deng, J. ,2018)演化而來，CornerNet( Law, H., & Deng, J. ,2018)為目標檢測演算法，相比起其他Anchor-based的目標檢測，其特點為以物體左上角的點與右下角的點來定位物體，在backbone部分受到人體姿態分析啟發，使用hourglass network(Newell, A., Yang, K., & Deng, J. 2016)，在姿態分析方面，因辨識不同的身體部位，適合的特徵維度不同，因此hourglass network的特點在於融合低維度與高維度的特徵，架構如沙漏般對稱，前半為原圖經過卷積，後半則將feature map進行採樣放大尺寸，與前半部分不同層級的feature map結合。

經過hourglass network，網路產生兩個分支，分別負責預測物體左上角與右下角的點，CornerNet提出了一種新的概念，corner pooling，用於求出角點，以左上角的點為例，點本身不一定會包含物體，而點所在的行列中會有物體邊緣，為了求出點，將某點的行中最大值與列中最大值相加，這樣所找出的最大值有可能是corner所在位置。

最後輸出三個分支，點的heatmap，embedding，offset，再透過兩點的embedding找出相配的關鍵點。

CenterNet與CornerNet類似，差別在CenterNet認為可以直接預測物體中心，因此它的結構在hourglass network後只有中心點的預測，最後輸出三個分支，中心點的heatmap，offset，物體的高與寬。

回到FairMOT，採用了CenterNet，還有一個好處，此類的anchor-free演算法會需要更高的分辨率，這使得re-ID準確率提升。在基礎網路部分，使用了encoder-decoder類型的網路，DLA-34(Yu, F., Wang, D., Shelhamer, E., & Darrell, T.,2018)，而檢測後的追蹤部分則與JDE同樣，使用匈牙利演算法和卡爾曼濾波進行track。

# 研究設計與實施

**第一節 研究方法**

我們將FairMOT作為主要的追蹤方法，其作者提供的開源程式碼，輸入為影片，輸出則是將處理過的frame生成影片。我們的目的為實時監控，因此需將輸入的部分改成本機鏡頭，並且在輸出時開啟窗口顯示目前偵測到的人，以及連接line notify，固定間隔傳送訊息，告知目前人數。

**第二節 將鏡頭作為輸入**

我們參考一篇實時追蹤的文章，在專案的src新增camera.py，內容與demo.py類似的檔案，只是將影片路徑來源改成了鏡頭。

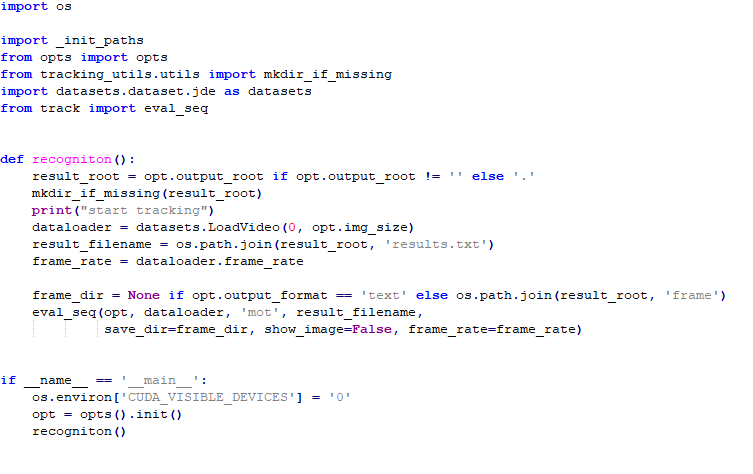
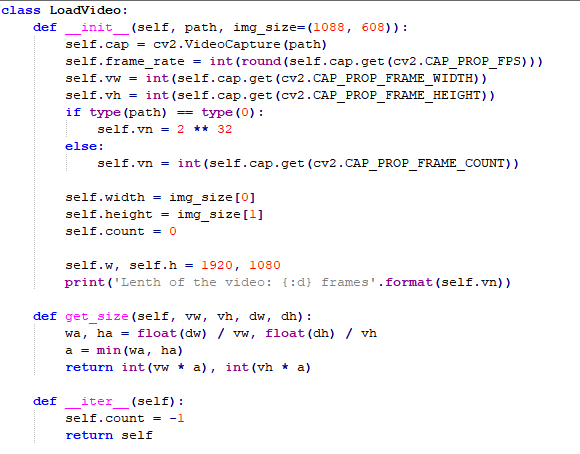


圖 1 camera.py內容

另外，在JDE裡的loadvideo，由於輸入不是影片，因此需將幀數修改掉。FairMOT的程式碼中有一個show image的函式，將參數從false改成true，代表顯示即時窗口。



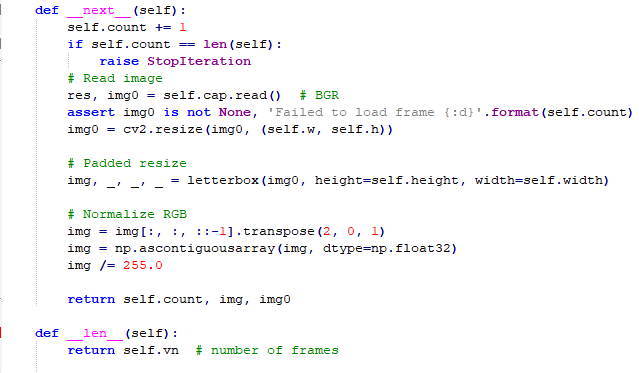
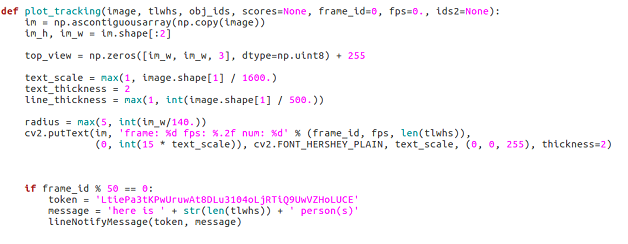


圖 2 loadvideo修改內容

**第三節 Line notify**

為了能方便達到實時監控，我們設定每50個frame，系統就會傳送訊息到對應的line帳號，告知目前人數，具體的程式碼在處理frame的檔案中實現。



Line

Line token

圖 3 line notify內容

# 實證結果與分析

請參考實測影片:

<https://drive.google.com/file/d/1qB7ynLphk098LhQN7SwZfaNJUQ6E2GEU/view?usp=sharing>

影片解釋 :

實測影片中我們可以關注三件事 ，分別為「即時性」、「姿勢的檢測」、「FPS」。以即時性來說，我們成功做到了每50frame傳送一則訊息，裡面告訴我們場域內有幾人，且不論是進場或是出去都要好好地將人框出來，對於即時性監控系統定義而言是我們所預期的結果，以檢測姿勢而論，我們影片中多為站姿，會有一些伸懶腰、走動和彎腰的行為，多數時候都有成功檢測，如果遺失資料實，下個frame也有做到重新抓取。而FPS的部分，因為我們硬體設備薄弱 (1060)，未達FairMOT的建議標準 (2060)，因此多數時候FPS都是低於1，以即時性而言，有進步的空間。

整體報告內容

<https://drive.google.com/file/d/1IoEfdtmeY5kgbrpDt4uS0kENtk9Thedd/view?usp=sharing>

# 結論

FairMOT具有實作出即時監控系統的能力，而資料部分可以更加齊全。因為硬體部分的不足，等待贊助，而我們提出一個新的方式，其介於JDE和FairMOT之間，因為人體站姿和坐姿的比例以圖片而論，所差無幾，而Anchor based 是因為要自行假設比例而吃經驗的問題 ，單論人體是可以避免的，因此我們認為可以嘗試用Anchor based的方式去實測哪一個較有效率，以及哪一個精確度可以更高。

# 參考文獻

周威（2020）。【MOT】对JDE深度解析。上網日期：2020年9月25日，檢自<https://zhuanlan.zhihu.com/p/243290960>。

周威（2020）。【MOT】对FairMOT深度解析。上網日期：2020年9月25日，檢自<https://zhuanlan.zhihu.com/p/259356109>。

周先森愛吃素（2020）。FairMOT实时多目标跟踪。上網日期：2020年10月24日，檢自<https://blog.csdn.net/zhouchen1998/article/details/109253638>。

Law, H., & Deng, J. (2018). Cornernet: Detecting objects as paired keypoints. In Proceedings of the European conference on computer vision (ECCV) (pp. 734-750).

Newell, A., Yang, K., & Deng, J. (2016, October). Stacked hourglass networks for human pose estimation. In European conference on computer vision (pp. 483-499). Springer, Cham.

Wang Z., Zheng L., Liu Y., Li Y., Wang S. (2020) Towards Real-Time Multi-Object Tracking. In: Vedaldi A., Bischof H., Brox T., Frahm JM. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science, vol 12356. Springer, Cham. <https://doi.org/10.1007/978-3-030-58621-8_7>

Yu, F., Wang, D., Shelhamer, E., & Darrell, T. (2018). Deep layer aggregation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2403-2412).

Zhang, Y., Wang, C., Wang, X., Zeng, W., & Liu, W. (2020). FairMOT: On the Fairness of Detection and Re-Identification in Multiple Object Tracking. arXiv preprint arXiv:2004.01888.

Zhou, X., Wang, D., & Krähenbühl, P. (2019). Objects as points. arXiv preprint arXiv:1904.07850.