

Object detection in insole recognition using faster R-CNN and Hough transform

108034573 蔡和諺

國立清華大學

摘要

物件偵測(object detection)是電腦視覺的應用之一，其目標在於標註多個物體所在的位置及大小，在這次的報告中，我們即將應用這項技術來辨識鞋墊工廠中射出成型的鞋墊位置，希望能進而利用機械手臂夾取，達成自動化生產的目的。此外，由於物件偵測無法標出被偵測物的方向，因此我們利用霍夫轉換找出鞋墊上的直線特徵，接著透過法向量求取鞋墊擺放的大致方向。

一、介紹

近年來，機器學習的方法快速的發展，在影像辨識方面也因此有了重大的突破，越來越多的演算法被提出來解決各式各樣的影像處理問題。首先，Convolutional neural networks (CNNs)的出現取代了更早之前用手動提取特徵的方法，進而導致在影像辨識方面能有更加客觀且快速的結果，之後又有針對物件偵測(物體辨識且定位)的神經網路產生，分別是 RCNN、Fast RCNN、Faster RCNN、YOLO...等，有了這些強大的模型，物件偵測的應用能更快速的推廣到各個領域。

隨著工業的快速發展，公司間的競爭也越來越激烈，因此節省成本就成了公司提升競爭力的方式之一。在生產業中，勞力始終佔據成本相當大的一部分，但科技的進步使工廠能自動化的生產，結合現今影像辨識的技術及越來越優化的機械手臂，工廠中的人力功能已有許多可被機器所取代。這次的報告以一間鞋墊工廠為主題，結合了深度學習及電腦視覺的方法，在偵測出鞋子的位置及方向後使用機械手臂夾取，讓生產流程能更加自動化。

二、方法

在這次的報告中，我們使用 Faster R-CNN 的方法來進行物件偵測，在資料的處理部分，透過公布在網路上的 `labellmg.py` 檔案為圖像進行標籤。Faster R-CNN 包含 4 個關鍵模組，分別是特徵提取網路、生成 ROI、ROI 分類、ROI 迴歸。(1) 特徵提取網路：這次使用 google 開發的 inception-v2 為 CNN 架構，並預先在 Oxford-IIIT Pets 資料集上訓練完成，用來提取圖像的特徵圖(feature map)；(2) 生成 ROI：在獲得的特徵圖的每一個點上做多個候選 ROI，然後利用分類器將這些 ROI 區分為背景和前景，同時利用迴歸器對這些 ROI 的位置進行初步的調整；(3) ROI 分類：在 RPN 階段，用來區分前景（於真實目標重疊並且其重疊區域大於 0.5）和背景（不與任何目標重疊或者其重疊區域小於 0.1），在 Fast-RCNN 階段，用於區分不同種類的目標；(4) ROI 迴歸：在 RPN 階段，進行初步調整，在 Fast-RCNN 階段進行精確調整。RPN 是 Faster R-CNN 中相當重要的模塊，它是一個直接生成候選區域的神經網路，經典的檢測方法生成檢測框都

非常耗時，如 OpenCV adaboost 使用滑動窗口+圖像金字塔生成檢測框；或如 R-CNN 使用 SS(Selective Search)方法生成檢測框。而 Faster RCNN 則拋棄了傳統的滑動窗口和 SS 方法，直接使用 RPN 生成檢測框，這也是 Faster R-CNN 的巨大優勢，能極大提升檢測框的生成速度。

當 Fast R-CNN 模型能成功偵測出鞋墊上我們想要的位置時(12 點及 3 線)，接著使用霍夫轉換(Hough transform)對所偵測出的區域圖像進行直線偵測，並透過數學運算計算出鞋墊大致的方向。詳細作法首先是找出三條直線的其中一條，並計算其中點及法向量，然後從 bounding box 的四個點取對角線的兩個點找平均數可以得到 bounding box 的中心點，從中心點前後延伸 0.1 倍法向量可得兩點，最後，分別計算與直線中點的距離，距離較短者為鞋墊尾，距離較長者則為鞋墊頭，如此一來可得到鞋墊大概的方向。

三、案例研究

本次報告是針對鞋墊工廠的鞋墊進行偵測。首先對鞋墊進行標籤，標籤的位置在鞋墊上的 12 個氣孔及上方三條線，只有完整的具有上述特徵才會被標註在圖像上，這次一共對 500 張鞋墊影像進行標註。接著將標註好的資料餵進 Faster R-CNN 的模型中，當 loss 降到 0.5 以下時即可透過 checkpoint 檔案確認偵測的結果。在測試結果階段，我們分別使用了繪圖軟體繪製的圖片及真實的鞋墊圖片來做測試，在電腦繪製的圖片上 bounding box 皆能順利的被框出來而且是在正確的位置(皆有超過 90%的信心程度)，但在霍夫轉換的過程中，線的位置並未完全如我們所預期的出現在三條直線上，而是有可能出現在鞋墊的邊緣，這樣的偵測結果導致法向量的計算出現錯誤，在畫出鞋墊頭尾兩個點時方向也因此產生錯誤。在真實的鞋墊圖像上，由於多了厚度的特徵，因此 bounding box 的誤差比電腦繪製的圖片大，在計算 bounding box 中心點時也較容易失準，但也因為如此在邊緣的直線反而較不明顯，在經過霍夫轉換後直線的位置大致上是正確的，在有成功匯出結果的鞋墊中表現優於電腦軟體繪圖資料的結果。

四、結論

這次報告的結果在 Faster R-CNN 部分相當的成功，由於剛開始標註圖片時，只標註了具有完整特徵的區域，因此在測試結果時未有完整特徵的鞋墊雖有可能被偵測，但其信心分數會較低，可以用此來篩選最有可能是鞋墊的位置；但由於真實情況相較訓練的資料及更為複雜，因此在真實的圖片上經常會出現無法框出 bounding box 的問題，在尋找直線特徵上也較易受到鄰近的鞋墊干擾。整體來說，在物件偵測這方面已經達到良好的結果了，但在找出物件方向上，這次提出的方法仍然有許多進步的空間。最新的技術已經可以將標註圖片的 bounding box 加上旋轉角度的參數，接著在訓練時將角度的變化加入損失函數的計算，如此以來也能生成旋轉過的 bounding box 並回傳角度參數，相信

這項技術在現實生活中會更加實用。

五、參考資料

- 1 Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91-99).
- 2 <https://reurl.cc/D1kXG5> Object Detection API 物件辨識分類器實作-採用 TensorFlow(CPU) 在 Windows10 上
- 3 <https://kknews.cc/zh-tw/code/k2yqmvb.html>
- 4 <https://medium.com/@ywchiu84021121/object-detection-s3-faster-rcnn-%E7%B0%A1%E4%BB%8B-5f37b13ccdd2>