

國立雲林科技大學資訊管理系

機器學習-作業三

Department of Information Management

National Yunlin University of Science & Technology

Assignment

貨櫃資料集分析

Container Data Set analysis

楊欣蓓、黃裕鳴、游榮翔

指導老師：許中川 博士

Advisor: Chung-Chian Hsu , Ph.D.

中華民國113年5月

May 2024

摘要

本研究旨在利用先進的物件偵測模型和文字辨識技術，自動且準確的辨識貨櫃號碼，提升全球物流運輸效率。隨著國際貿易規模的擴大，人工辨識貨櫃號碼的方式已經難以應對日益增多的貨櫃數量，因此自動化辨識系統成為急需解決的問題。本研究利用不同種類貨櫃的影像資料集，運用深度學習技術，開發出能夠在複雜背景下精確定位和辨識貨櫃號碼的模型。實驗方法包括兩個階段:物件偵測模型訓練和影像文字辨識整合。物件偵測模型訓練使用大規模貨櫃照片資料集進行，並通過多項性能指標(mAP、Recall、Precision 和 F1 分數)評估模型。影像文字辨識整合階段則將物件偵測模型與 OCR 文字辨識模組結合，建構出完整的辨識流程。系統使用貨櫃影片資料集驗證其在實際運行中的效能，並通過多數決決策法提高影片整體辨識的準確率本研究採用 Python3.8 及相關深度學習框架(YOLO、OpenCV)進行開發。資料及包含 2125 筆訓練集和 536 筆驗證集影像。本研究結果顯示，經過 50 個 epoch 訓練,YOLO 模型在識別貨櫃物件上表現卓越,mAP、Precision、Recall 和 F1 分數等指標均達極高水準。在 OCR 環節,採用影像預處理的 EasyOCR 工具較 pytesseract 有更高 0.69 的準確率。針對貨櫃影片,提高 YOLO 模型的置信度閾值可有效排除干擾目標。最終,將優化後的 YOLO 與 EasyOCR 相結合,在貨櫃影片號碼識別任務上達到 0.9 的正確率,展現出該綜合方案在實際場景中的卓越性能,為港口物流智能化發展提供有力支持。

關鍵字：物件偵測模型、光學文字辨識、深度學習、電腦視覺

一、緒論

1.1 研究動機

本研究欲利用物件偵測模型和文字辨識來辨識貨櫃號碼。貨櫃運輸是全球物流體系的重要樞紐，準確辨識貨櫃號碼可以提升運輸效率。此外，隨著國際貿易規模不斷擴大，每天進出港口的貨櫃數量極為龐大，以人工辨識方式不僅效率低，更容易產生錯誤，因此自動化貨櫃號碼辨識為重要的發展項目。

近年來，電腦視覺及機器學習技術日新月異，為開發自動化貨櫃號碼辨識系統帶來新契機。然而，實務場景中貨櫃外觀條件多變、光線及背景干擾等因素，使得準確辨識貨櫃號碼仍面臨重重挑戰，為此本研究欲研究貨櫃號碼辨識方法以期能夠提供理想的解決方案。

1.2 研究目的

本研究旨在研究使用物件偵測模型和文字辨識來自動準確辨識出貨櫃號碼。為達成此目標，研究工作首先將建立一個大規模的貨櫃號碼辨識訓練資料集，蒐集實際港口場景中不同種類貨櫃的樣本影像，作為研究的基礎。

接著，本研究將深入探索利用深度學習等先進機器學習技術，研究出強大的貨櫃號碼辨識模型。本研究該模型不僅需精確定位並分割出影像中的貨櫃號碼區域，更需能從複雜背景、光照變化、不同角度和汙損情況下，使物件偵測模型和文字辨識來有效辨識出號碼字元，並正確解析出完整的貨櫃號碼序列。

最後，本研究將對所開發的貨櫃號碼辨識演算法進行全面評估，驗證其在各種極端環境下的辨識準確率及效能表現，確保系統的穩定性和可靠性。總體而言，本研究譽為港口貨櫃物流營運開發出高效可靠的自動化解決方案，提升整體作業效率，降低人力成本，並有助於促進全球供應鏈的智能化升級。

二、實驗方法

2.1 實作說明

本研究旨在開發穩健準確的貨櫃號碼辨識系統。實驗分為物件偵測模型訓練和影像文字辨識整合兩階段。物件偵測模型訓練階段使用大規模貨櫃照片訓練集、驗證集和測試集。使用模型 YOLO 辨識圖片並詳細記錄。使用測試集評估各模型的 mAP、Recall、Precision 和 F1 分數，綜合考量選出最佳模型。

影像文字辨識整合階段將物件偵測模型與 OCR 文字辨識模組相結合，構建完整辨識流程。物件偵測定位分割出貨櫃號碼區域，文字辨識輸出對應文字序列。

使用貨櫃影片資料集驗證系統實際運行效能。對每一幀自動檢測是否存在號碼，若存在輸出辨識結果並疊加原影像。統計每一幀辨識準確率，使用多數決策獲得影片最終號碼輸出，與實際號碼比對計算整體準確率。

通過嚴格實驗評估，期望開發出在港口實務場景中具備卓越性能的智能貨櫃號碼辨識解決方案，促進全球供應鏈物流效率提升。

2.2 操作說明

本研究採用 Python 3.8 作為編程語言，並使用 Visual Studio Code 作為開發環境。本研究選擇 YOLOv7、ORC 文字辨識作為深度學習框架，用於構建、訓練和推理神經網路模型。

三、實驗設計

3.1 資料集

3.1.1 貨櫃資料集概述

名稱：貨櫃資料集

貨櫃圖片資料集的總筆數：2125筆(訓練集)+536筆(驗證集)=2661筆

貨櫃位置資料集的總筆數：2125筆(訓練集)+536筆(驗證集)=2661筆

上述圖片資料集解析度皆為1920*1080



圖 1

部分貨櫃圖片資料集展示

```
<annotation>
  <folder>訓練集</folder>
  <filename>image_0001.jpg</filename>
  <path>E:\貨櫃訓練資料\yolov4(1)_資料\訓練集\image_0001.jpg</path>
  <source>
    <database>Unknown</database>
  </source>
  <size>
    <width>1920</width>
    <height>1080</height>
    <depth>3</depth>
  </size>
  <segmented>0</segmented>
  <object>
    <name>container</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>814</xmin>
      <ymin>33</ymin>
      <xmax>998</xmax>
      <ymax>122</ymax>
    </bndbox>
  </object>
</annotation>
```

圖 2

部分貨櫃位置資料集展示

名稱：影片資料集

貨櫃影片資料集總筆數=10筆

上述影片解析度皆為1920*1080，影片長度大約落在6秒至51秒不等。

3.1.2 標註位置資料集

原始貨櫃位置資料集為 VOC Pascal xml 格式，為了符合模型需求，使用公式將座標作轉換成 YOLO 的 txt 格式，使得位置資料得以輸入模型訓練中。

表 1

部分貨櫃位置資料集資料預處理前後內容

	座標			
資料轉換前	814<xmin>	33<ymin>	998<xmax>	122<ymax>
資料轉換後	0.4719	0.0718	0.0958	0.0824

$$x = \left(\frac{X_{min} + X_{max}}{2} - 1 \right) \cdot \frac{1}{W} \quad (1)$$

$$y = \left(\frac{Y_{min} + Y_{max}}{2} - 1 \right) \cdot \frac{1}{H} \quad (2)$$

$$w = (X_{max} - X_{min}) \cdot \frac{1}{W} \quad (3)$$

$$h = (Y_{max} - Y_{min}) \cdot \frac{1}{H} \quad (4)$$

3.2 資料前處理

3.2.1 貨櫃資料集

- 資料前處理
 - 於文字辨識前，將切割好的貨櫃圖片使用增加對比度，使明暗更加分明，此外也做銳利化處理，使交界處更加明顯。



圖 3

左圖為預處理前的貨櫃圖片，右圖為預處理後的貨櫃圖片

3.3 實驗設計

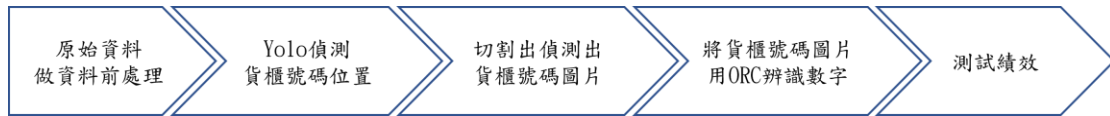


圖 4

貨櫃資料集實驗設計流程圖

3.2.2 貨櫃資料集

本研究欲利用物件偵測模型訓練和影像文字辨識技術以研究貨櫃資料集。首先進行物件偵測模型的訓練，使用大規模貨櫃照片數據集包含訓練集、驗證集和測試集，採用 YOLO 物件偵測算法進行訓練，辨識出貨櫃號碼區域。在測試集上評估各種模型的 mAP、Recall、Precision 和 F1分數等指標，綜合考量選出性能最佳的模型。將物件偵測模型與光學字元辨識(OCR)模組相結合，構建完整的號碼辨識系統，接著物件偵測定位分割出貨櫃號碼區域後，OCR 模組輸出對應文字序列，即為最終辨識結果。通過上述嚴格的實驗設計與評估，本實驗期望透過貨櫃號碼辨識研究以提升供應鏈物流效率。

3.4 實驗結果

本實驗主要用 YOLO 偵測模型和 ORC 文字辨識來分析貨櫃資料集，辨識對象分別為貨櫃測試資料集、貨櫃影片資料集，對上述二者進行偵測，使用 mAP、Recall、Precision 和 F1 分數等指標進行績效分析後，切割出偵測圖片做 ORC 文字辨識，顯示出圖片文字內容，做準確率績效測試。

3.4.2 貨櫃資料集分析

本實驗使用 YOLO 偵測模型對貨櫃資料集進行訓練，參數為 batch size=16、epoch=50 進行訓練，下方圖4為 mAP、Recall、Precision 和 F1 分數等指標表現，可觀察出模型於訓練時，績效不僅逐漸收斂也表現優良。

Precision 衡量模型對偵測到的物件的準確性，Recall 則評估模型對真實物件的覆蓋率。隨著訓練的進行，這兩項指標不斷提升並最終達到相當高的水準，意味著模型能夠準確地識別並包含大部分真實貨櫃物件。此外 F1 分數最終達到極高水準，反映出訓練出的 YOLO 模型在辨識貨櫃物件上具備出色的整體表現。mAP 是計算物件偵測模型在多個物件類別上的加權平均精準度，廣泛用於評估模型的整體精確度。實驗結果顯示，經過50個訓練週期後，mAP 值已收斂至相當高的水平，證明了模型對於辨識貨櫃物件這一單一類別也有極為出色的準確性。

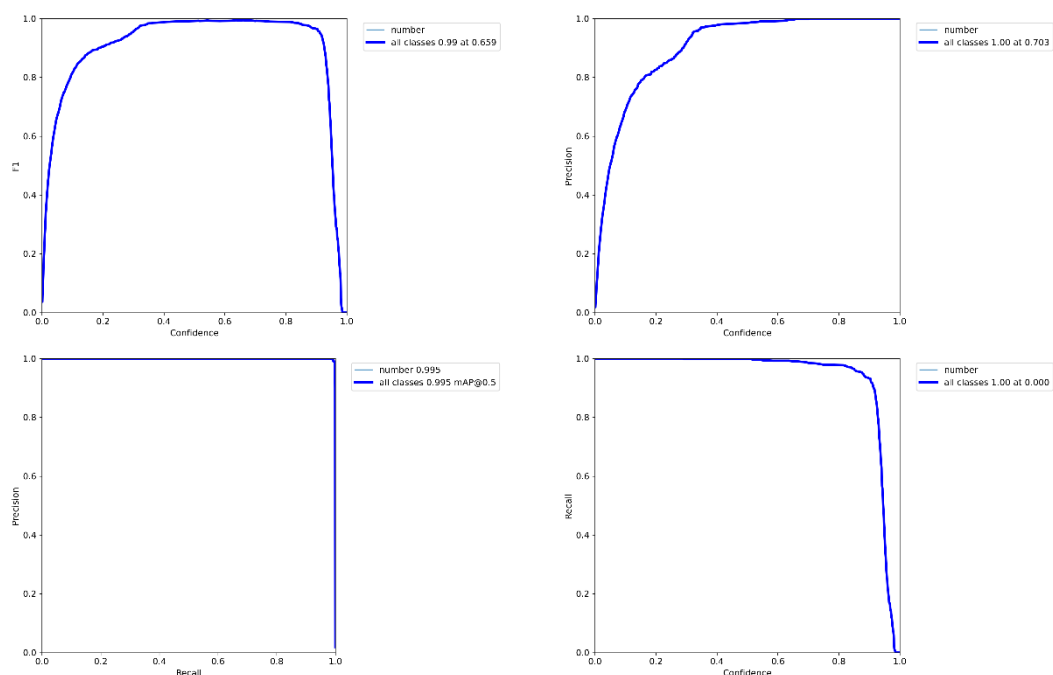


圖 5

貨櫃資料集的 F1 分數(左上)、Precision(右上)、mAP(左下)、Recall(右下)

接著使用訓練好的 YOLO 模型去偵測測試集的貨櫃圖片，並將偵測到的位置進行切割，並將切割好的圖片輸入 ORC 文字辨識中進行分析，並將分析後的資料進行正確率績效測試。

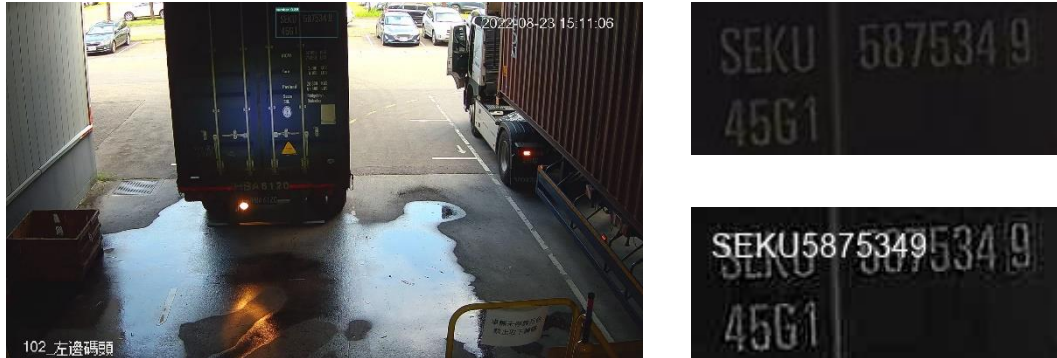


圖 6

左圖為偵測中的貨櫃資料集，右上圖為切割後的貨櫃資料集，右下圖為 ORC 文字辨識後的展示

本實驗使用兩種 ORC 文字辨識進行績效比較，分別為 EasyOCR、pytesseract，將所有測試集進行實驗作分析後，EasyOCR 的準確度=0.69、pytesseract 的準確度=0.57。EasyOCR 的優異表現於其在輸入圖片時先對其進行了預處理，包括對比度增強和銳利化等步驟。這些影像增強技術的應用使得原始圖片中模糊或細節不清的貨櫃號碼部分得以改善，從而提高了後續 OCR 識別的精準度。該做法印證了適當的影像預處理對於 OCR 文字辨識的重要性，因此也選用 EasyOCR 對貨櫃影片資料集進行分析。

於貨櫃影片偵測部分，首先 YOLO 偵測模型使用預設 $\text{conf}=0.5$ 進行實驗。然而實驗結果顯示，在此設定下偵測到的目標物件除了所需的貨櫃號碼之外，還包含了其他一些干擾目標，如下圖所示。這些額外的干擾目標不僅會增加後續處理的工作量，更可能導致最終的號碼識別精確度下降。為了解決這個問題，本實驗嘗試提高 conf 。經過多次實驗，當 $\text{conf}=0.9$ 時，YOLO 偵測模型能更加精確地識別和定位影片中的貨櫃號碼目標，同時有效過濾掉其他不相關的物體。這確保了獲得的影像區域僅包含貨櫃號碼，不會被其他干擾目標所影響，為後續的 OCR 文字識別階段提供了高度純淨且準確的輸入數據。

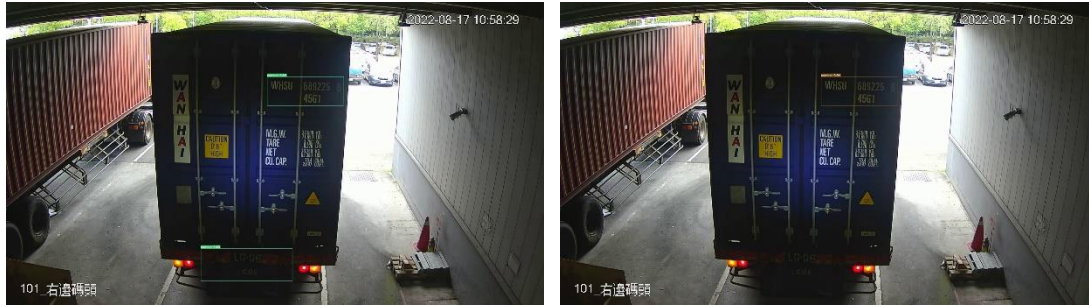


圖 7

對貨櫃影片進行偵測，左圖為 $conf=0.5$ ，右圖 $conf=0.9$

於影片偵測的過程中將每幀有偵測到目標物的位置一併儲存下來，並將其檔名為當時的幀數，接著讀取偵測後的位置檔對應到原始貨櫃影片之幀數進行圖片切割，過程以影片的每一幀對應該影片位置幀數檔名下去做切割，為 EasyOCR 的文字辨識做好準備。接著，將這些切割後的圖片依序輸入到 EasyOCR 模組中。EasyOCR 會對每一幅圖像中的文字區域進行掃描和解析，輸出對應的號碼字符序列，最後統計了各個號碼字符序列出現的次數。由於每一個影片中的貨櫃號碼是固定不變的，因此出現次數最多的序列就是該影片真實的號碼。最終輸出結果與影片的 ground truth 標註資料進行對比，計算出號碼辨識的正確率=0.9。此實驗結果驗證了該號碼辨識解決方案在貨櫃影片分析任務上的實用價值和可靠性，為其在港口物流實際應用場景奠定了堅實基礎。

```
影片資料集/video_0001.avi:SEKU5875349:51
影片資料集/video_0002.avi:TSSU5099404:139
影片資料集/video_0003.avi:WHSU5927854:57
影片資料集/video_0004.avi:WHSU6728694:89
影片資料集/video_0005.avi:WHSU5368194:95
影片資料集/video_0006.avi:WHSU2403178:14
影片資料集/video_0007.avi:WHSU6167124:59
影片資料集/video_0008.avi:WHSU6892251:61
影片資料集/video_0009.avi:WHSU5223794:34
影片資料集/video_0010.avi:TLLU4080734:157
```

圖 8

各個貨櫃影片資料集偵測出貨櫃號碼出現最多次展示

四、結論

本研究使用 YOLO 偵測模型和 ORC 文字辨識來分析貨櫃資料集。首先，針對貨櫃圖像資料集，經過50個訓練 epoch 後，我們訓練出的 YOLO 目標檢測模型表現出色。各項評估指標如 mAP、Precision、Recall 和 F1分數等均達到了極高水準，證明了模型在辨識貨櫃物件這一單一類別上具有卓越的準確性和完整性。其次，在 OCR 文字辨識環節，比較了 EasyOCR 和 pytesseract 兩種工具。實驗結果顯示，採用了影像預處理步驟的 EasyOCR 在準確度上領先 pytesseract，達到0.69的高水準。這突顯了適當的影像增強對 OCR 識別的重要性。接著，在貨櫃影片資料集的分析中，先通過調整 YOLO 模型的 conf，成功排除了大量干擾目標，僅保留貨櫃號碼的有效區域。然後將 YOLO 與 EasyOCR 模組相結合，建立完整的號碼辨識流程。最終在這一綜合解決方案下，貨櫃影片號碼的辨識正確率高達0.9，驗證了該方法在實際應用場景中的價值。

本實驗充分展現了基於 YOLO 目標檢測和 OCR 文字辨識的貨櫃分析解決方案的優異性能。通過模型調參數和優化，構建出貨櫃號碼自動識別系統，為港口物流等領域的智能化發展提供了有力支撐。

參考文獻

Dario Radečić (2021) 。如何使用 TensorFlow 優化學習率——比你想像的要容易。

[How to Optimize Learning Rate with TensorFlow — It's Easier Than You Think | by Dario Radečić | Towards Data Science](#)

Wang, C., Bochkovskiy, A., & Liao, H.M. (2022). YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7464-7475.