智慧化企業整合

Final Project

License plate recognition via ESRGAN and Faster R-CNN

109034554 蔡丞洲

摘要

近年來隨著目標偵測(Object detection)的技術進步，國道的電子收費系統(ETC)也加入了自動偵測車輛收取通行費，根據 2019 年統計平均一個月的交易量為 5 千萬做又。從系統偵測結果發現，正確分辨車牌的機率達 99.91%，代表一個月仍有四萬五千次的錯誤扣款或是無法偵測車牌，因此想藉由此專題來提升物件偵測模型的準確度，並使用生成對抗網路(Generation Adversarical Network)來生成解析度較佳的圖像來幫助模型學習。

近年隨著深度學習(Deep Learning)持續發展，生成對抗網路(Generation Adversarical Network)被提出，並在影像生成相關任務上有出色的表現。因此本專題想藉由 GAN 的技術來提升物件偵測的解析度，使物體更容易被檢測器偵測。

1、 緒論

1.1 問題定義 5W1H

WHY:目標檢測技術仍有改善空間
WHAT: 提升目標檢測準確率。
WHERE:高速公路的 ETC 裝置位置。
WHEN:當用戶經過 ETC 時。
WHO: 用於所有經過高速公路的駕駛。
HOW:利用 GAN 生成高解析度的車牌影像，並以 Faster R-CNN 來訓練車牌資料庫，並使檢測準確度提高。

1.2 研究動機與目的

由於目前 ETC 使用了目標檢測方式來偵測車牌，並進行收取通行費，雖然分辨車牌及扣款的正確率達 99.91%，但經 ETC 車輛眾多，代表一個月仍有 4 萬多筆交易處理錯誤，因此想透過此專題來進行目標檢測準確度的改善。

透過 5W1H 分析法，來幫助我們對問題分析並思考解決方法，為了使現有的目標檢測技術準確率更高，因此此專案目的為藉由加入生成對抗網路，來測試是否對檢測結果有所提升。

2、 文獻探討

2.1 Faster R-CNN 介紹

2.2.1 Faster R-CNN 介紹

目前的目標檢測網路依靠 region proposal 演算法來假設目標的位置，諸如 SPPnet 和 Fast R-CNN 所取得的進步已經減少了這些檢測網路的執行時間，但也揭露了 region proposal 的計算是一個瓶頸。

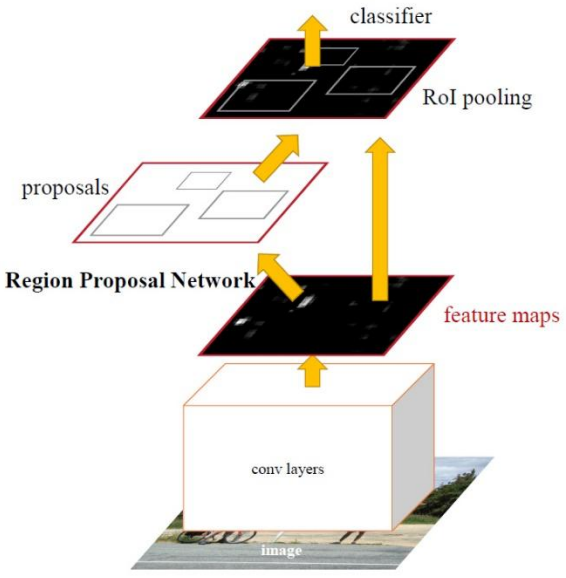
Faster R-CNN 的作者發現用於基於 region 的 detector（如 Fast R-CNN）的卷積 feature map 也可以用於生成 region proposals。於是在這些卷積特徵的頂部，通過新增一些額外的卷積層來構建一個 Region Proposal Network（RPN），對每一位置同時輸出 region bounds 以及 objectness score。因此 RPN 是一個全卷積網路（full convolutional network，FCN），可以進行端到端訓練，生成高質量的 region proposals，然後送入 Fast R-CNN 進行檢測，進而降低檢測網路的執行時間。

2.2.2 Faster R-CNN 方法

Faster R-CNN 的方法為通過共享卷積特徵進一步將 RPN 和 Fast R-CNN 整合成一個網路，並提出了一種訓練機制：在保持 proposal 固定的情況下，交替微調 region proposal 和 object detection。

對於 RPN 網絡，先採用一個 CNN 模型接收整張圖片並提取特徵圖。然後在這個特徵圖上採用一個 N×N 的滑動窗口，對於每個滑窗位置都映射一個低維度的特徵（如下圖）。然後這個特徵分別送入兩個全連接層，一個用於分類預測，另外一個用於回歸。對於每個窗口位置一般設置 k 個不同大小或比例的先驗框（anchors,default bounding boxes），這意味著每個位置預測 k 個候選區域（region proposals）。

對於分類層，其輸出大小是 2k，表示各個候選區域包含物體或者是背景的概率值，而回歸層輸出 4k 個坐標值，表示各個候選區域的位置（相對各個先驗框）。對於每個滑窗位置，這兩個全連接層是共享的。因此，RPN 可以採用卷積層來實現：首先是一個 n×n 卷積得到低維特徵，然後是兩個 1×1 的捲積，分別用於分類與回歸。



2.2.3 Faster R-CNN 步驟

以下為 Faster R-CNN 的運行步驟:

1. 使用 ImageNet 上預訓練的模型初始化特徵提取網絡並訓練 RPN 網絡。
2. 使用在 ImageNet 上預訓練的模型初始化 Fast-RCNN 特徵特徵提取網絡，使用步驟一中訓練好的 RPN 網絡產生的候選框作為輸入，訓練一個 Fast-RCNN 網絡，至此，兩個網絡每一層的參數完全不共享。
3. 使用步驟二的 Fast-RCNN 網絡參數初始化一個新的 RPN 網絡，但是把 RPN，Fast-RCNN 共享的特徵提取網絡參數的學習率設為 0，即使學習 RPN 網絡所特有的參數，固定特徵提取網絡。到此步，兩個網絡已經共享了所有的公共的捲積層。
4. 仍然固定共享的那些網絡層，把 Fast-RCNN 特有的網絡層也加入進來，繼續訓練，微調 Fast-RCNN 特有的網絡層，到此為止，RPN 與 Fast-RCNN 網絡完全共享參數，使用 Fast-RCNN 即可同時完成候選框提取和目標檢測功能。

2.2 ESRGAN

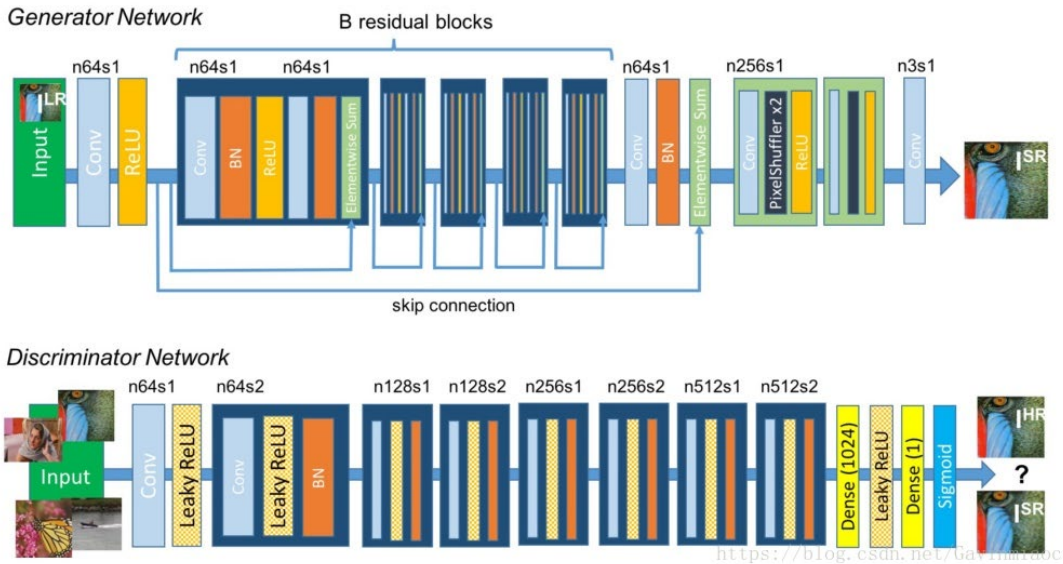
2.2.1 Super-Resolution

超解析度技術（Super-Resolution）是指從觀測到的低解析度影象重建出相應的高解析度影象，在監控裝置、衛星影象和醫學影像等領域都有重要的應用價值。SR 可分為兩類:從多張低解析度影象重建出高解析度影象和從單張低解析度影象重建出高解析度影象。基於深度學習的 SR，主要是基於單張低解析度的重建方法，即 Single Image Super-Resolution (SISR)。

SISR 是一個逆問題，對於一個低解析度影象，可能存在許多不同的高解析度影象與之對應，因此通常在求解高解析度影象時會加一個先驗資訊進行規範化約束。在傳統的方法中，這個先驗資訊可以通過若干成對出現的低-高解析度影象的例項中學到。而基於深度學習的 SR 通過神經網路直接學習解析度影象到高解析度影象的端到端的對映函式。

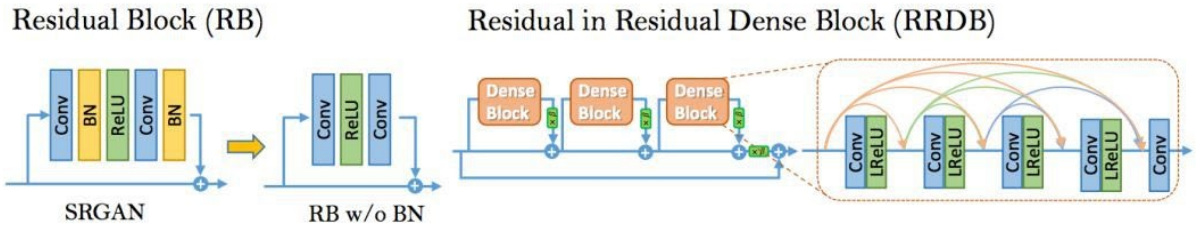
2.2.2 SRGAN

SRGAN 將生成式對抗網路（GAN)用於 SR 問題。其出發點是傳統的方法一般處理的是較小的放大倍數，當影象的放大倍數在 4 以上時，很容易使得到的結果顯得過於平滑，而缺少一些細節上的真實感。因此 SRGAN 使用 GAN 來生成影象中的細節。其中 SRGAN 利用感知損失(perceptual loss)和對抗損失(adversarial loss)來提升恢復出的圖片的真實感。感知損失是利用卷積神經網路提取出的特徵，通過比較生成圖片經過卷積神經網路後的特徵和目標圖片經過卷積神經網路後的特徵的差別，使生成圖片和目標圖片在語義和風格上更相似。下圖為 SRGAN 架構。



2.2.3 ESRGAN

SRGAN 模型極大地提升了超解析度結果的視覺質量，但是 SRGAN 模型得到的圖像和 GT 圖像仍有很大的差距。因此提出了 ESRGAN，為了提升 SRGAN 重建的圖像質量，將生成器網絡的基本單元從基本的殘差單元變為 Residual-in-Residual Dense Block (RRDB)，如下圖，並將 GAN 網路改進為 Relativistic average GAN (RaGAN)，最後在改進感知域損失函式，使用激活前的 VGG 特征，這個改進會提供更尖銳的邊緣和更符合視覺的結果，經過證明後，ESRGAN 對提升輸出圖像的視覺效果都有作用。



3、 研究方法

3.1 資料來源

訓練樣本來自 kaggle 網站提供的資料庫，在 Car License Plate Detection 中，並有 853 張影像，影像中有不同角度、大小的車牌，而在檔案中除了圖檔以外，還附上 xml 檔並記錄 bounding box 座標及長寬，因此不用再做 label 的動作，如圖 3-2。

訓練樣本分為兩個資料集，第一個為來自 kaggle 網站提供的資料庫 Labeled licence plated dataset，並選擇 200 張影像作為測試集，由於經過初步結果，第一個測試集準確度達到 90%以上，因此建立第二個測試集來加強驗證，

第二個測試集為取自於 Google 所提供的車牌影像，挑選較難分辨車牌之圖象，並選擇 200 張做為第二個測試集。

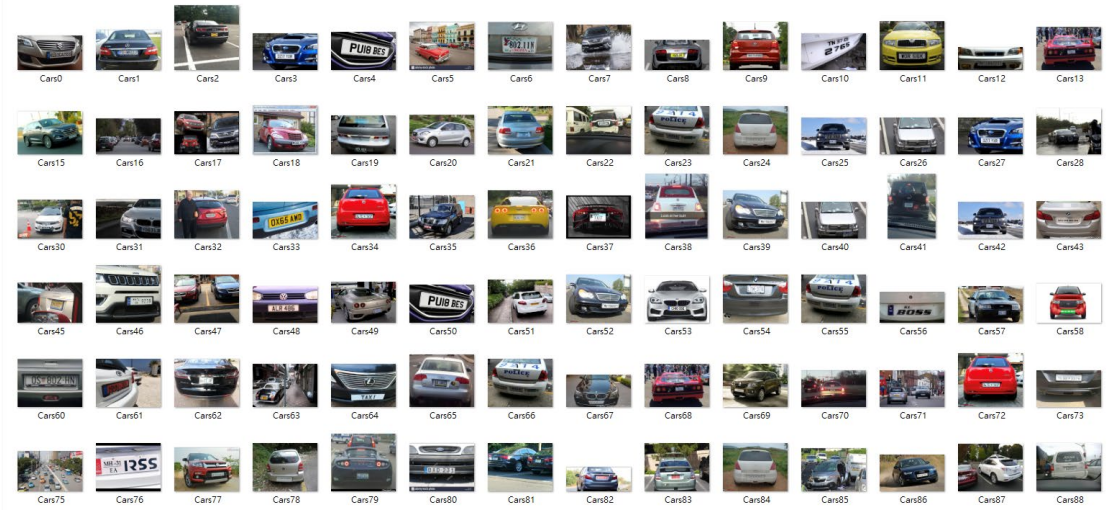


圖 3-1 訓練集影像

圖 3-2 測試集 1 影像

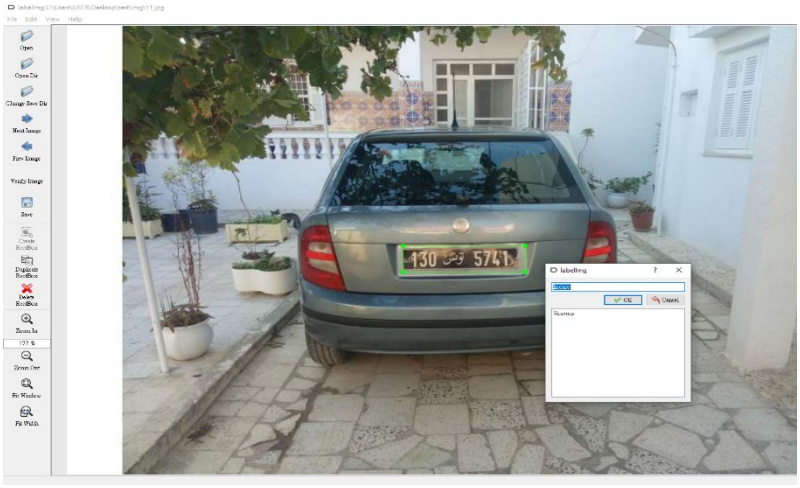
圖 3-3 測試集 2 影像

```
<annotation>
    <folder>images</folder>
    <filename>Cars4.png</filename>
    <size>
        <width>590</width>
        <height>350</height>
        <depth>3</depth>
    </size>
    <segmented>0</segmented>
    <object>
        <name>licence</name>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
        <occluded>0</occluded>
        <difficult>0</difficult>
        <bndbox>
            <xmin>156</xmin>
            <ymin>82</ymin>
            <xmax>503</xmax>
            <ymax>253</ymax>
        </bndbox>
    </object>
</annotation>
```

圖 3-4 影像標記檔

3.1.1 生成標記檔

由於測試集 2 沒有標記檔，因此以 LabelIMG 軟體來標記 bounding box 的座標，並生成標記檔。



3.2 研究架構

本研究架構分為兩個部分，第一部分為將原始圖片轉為超解析度圖片，第二部分為，分別將原始圖片及超解析度圖片利用 Faster R-CNN 訓練及分析出準確度，再比較超解析度圖片是否對目標檢測方法有準確度的提升，最後再將超解析度圖片的 Faster R-CNN 作超參數優化，來確定最佳的超參數組合。

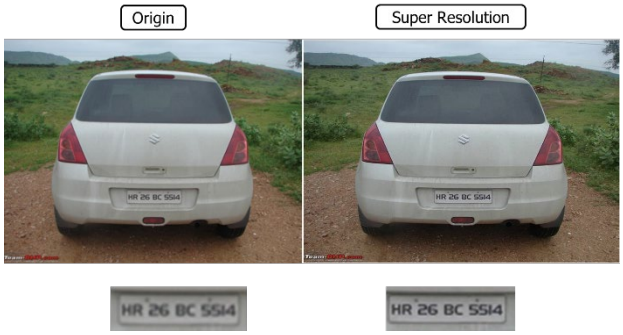
- **Stage 1**

- **Stage 2**

3.3 生成 SR 圖片

將 433 張車牌圖像作為輸入，首先進行預處理，此步驟分為兩部分，第一部分為圖片標準化，圖片類型會有 jpg 和 png，因此透過預處理將圖片磚換成張量，並且統一維度，第二部分為調整圖片尺寸，由於 ESRGAN 產出為原始圖像的 4 倍大，因此要將圖片的長和寬裁為 4 的倍數。接下來使用 tensorflow-hub 所提供的 pre-train-model 進行生成 SR 圖像，最後再將張量轉為圖像。

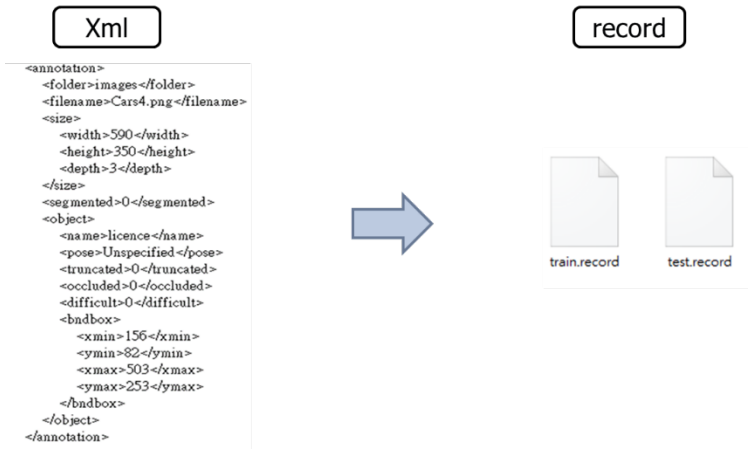
下圖為原圖和生成出的超解析度圖片，經縮放成同樣大小後，比較車牌的清晰度，可看出經 ESRGAN 的超解析度圖片，不管是在紋理和清晰度皆比原始圖片更好。



3.4 Faster R-CNN

3.4.1 前處理

利用 google api 所提供的 Object detection github，進行 Faster R-CNN 分析。首先進行標記檔預處理，將原先的 xml 檔利用 python 轉換為 record 檔，如下圖所示。

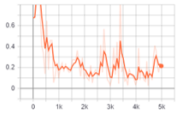


3.4.2 訓練

利用預設的初始超參數對原始圖片及超解析度圖片進行訓練，在特徵提取的模型使用 Inception v2，訓練步數為 5000、學習率為 0.0002。

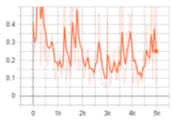
以下為分析結果，並可得到不論是在 testset1 和 testset2 超解析度圖片再準確率都比原始圖片佳，且在 ESRGAN 的 loss 訓練圖可得到學習還沒達到收斂，表示有可能會得到更好的準確度，所以在下個步驟進行超參數優化。

- **Test set 1 : Labeled licence plated dataset**
- 原圖



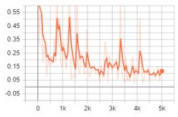
	accuracy
Origin	0.947
ESRGAN	0.961

- ESRGAN



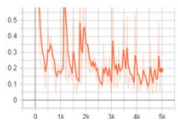
22

- **Test set 2 : Google**
- 原圖



	accuracy
Origin	0.739
ESRGAN	0.778

- ESRGAN



23

4、 超參數優化

4.1 超參數調整

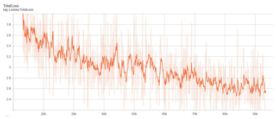
本研究選擇了 Feature extraction model、Learning rate、Steps 和 Iou threshold 作為預調整超參數。

4.1.1 Feature extraction model

以下為調整結果，選擇 Inception v2 作為特徵萃取的模型，SSD MobileNet 在此資料集無法達到收斂，而 ResNet101 v2 則是要經過更長的訓練時間才會得到較好的結果，但衡量訓練時間及準確度，最後仍然選擇了 Inception v2。

- **Feature extraction model調整**

models	accuracy	決策
SSD MobileNet v2	0.658	
Inception v2	**0.968**	✓
ResNet101 v2	0.861	



4.1.2 Steps & Learning rate 調整

由於考慮到 Steps 和 Learning rate 之間可能會互相影響，因此將所有的組合都進行測試，最後選擇了訓練步數 10000、學習率為 0.0005，有最高的準確率達 0.972。

- **Steps & Learning rate調整**

Steps	Learning rate	accuracy	決策
3000	0.0002	0.951	
5000	0.0002	0.961	
	0.0005	0.968	
	0.0008	0.958	
8000	0.0002	0.959	
	0.0005	0.968	
	0.0008	0.952	
10000	0.0002	0.960	
	0.0005	**0.972**	✓
	0.0008	0.953	

4.1.3 Iou threshold 調整

最後進行 Iou threshold 的選擇，經以下測試，選擇了預設值 0.6 作為最佳參數。

- **Iou threshold調整**

	initial
steps	10000
Learning rate	0.0005
Iou threshold	0.6

Iou threshold	accuracy	決策
0.7	0.971	
0.6	**0.972**	✓
0.5	0.957	

5、 研究結果

5.1 結果分析

使用調整過後的超參數進行最後的測試，(Feature extraction model 使用 Inception v2、Learning rate 為 0.0005、Steps 為 10000 和 Iou threshold 為 0.6)，且得到以下結果。在測試集 1 經超參數調整後，可對準確度提升 1.1%；在測試集 2 可提升準確度 4.1%。

- Dataset 1

超參數調整前	accuracy
Origin	0.947
ESRGAN	0.961

1.2%

1.1%

超參數調整後	accuracy
Origin	0.958
ESRGAN	0.972

- Dataset 2

超參數調整前	accuracy
Origin	0.739
ESRGAN	0.778

3.4%

4.1%

超參數調整後	accuracy
Origin	0.764
ESRGAN	0.810

圖 4-2 成果績效

5.2 測試影像

隨機從資料集選出照片進行實際測試，結果如下圖。

圖 4-3 測試結果照片

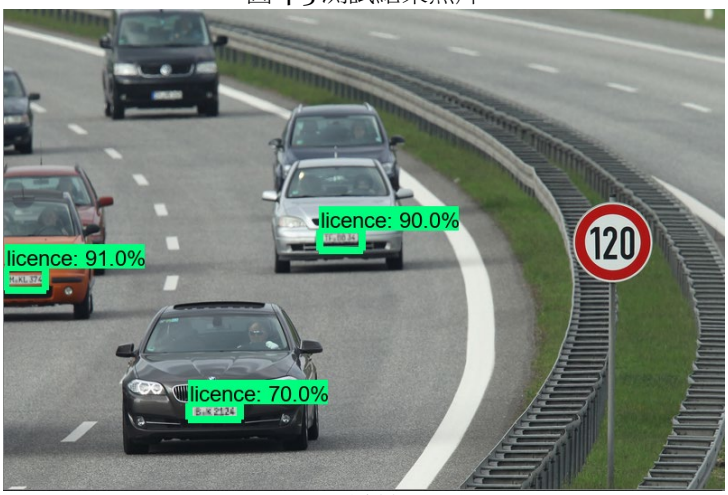


圖 4-4 測試結果照片

6、 結論

經此研究可得到 ESRGAN 所生成的超解析度圖像，不論是在紋理還是色彩皆比原始圖片來的清晰，經結果顯示使用超解析度圖片進行 Faster R-CNN 的檢測，能使準確率得到提升，並且經過超參數優化得到準確率 97.2%。此研究的限制為必須考慮模型的訓練時間，在 Faster R-CNN 的特徵提取模型，只使用常見的模型進行比較，且在生成超解析度圖片沒有進行 ESRGAN 訓練及優化。在未來方向，能比較更多物件偵測模型來驗證超解析度圖片能提高檢測能力，並且利用 OCR 讀取車牌數字，增加研究的適用性。

7、 參考文獻

[1] https://tisvcloud.freeway.gov.tw/
[2] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ...

& Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).
[3] Hicsonmez, S., Samet, N., Akbas, E., & Duygulu, P. (2020). GANILLA: Generative adversarial networks for image to illustration translation. Image and Vision Computing, 95, 103886.
[4] https://zhuanlan.zhihu.com/p/31426458
[5] https://medium.com/@9821343/object-detection-api%E7%89%A9%E4%BB%B6%E8%AD%98%E5%88%A5%E5%88%86%E9%A1%9E%E5%99%A8%E5%AF%A6%E4%BD%9C-%E6%8E%A1%E7%94%A8tensorflow-cpu-%E5%9C%A8windows-10%E4%B8%8A-83d73065f27f