繁體中文場景文字辨識競賽－

初階：場景文字檢測

隊伍: 報名系統一致

成員: 吳亦振

壹、 環境

作業系統 : ubuntu 18.04
語言 : python 3.6.9
套件 :

pandas==1.1.4
numpy==1.17.2
torch==1.3.0
torchvision==0.4.1
tqdm==4.53.0
PIL==6.2.0
cv2==4.5.1

額外資料集 : ReCTS(20,000 筆) (連結)
預訓練模型 : Resnet-152 (torchvision)

貳、 演算方法與模型架構

1. 使用 Mask RCNN 架構，其中 backbone 利用 resnet-152 預訓練模型
2. 利用 Mask RCNN 的語義分割輸出結果推論場景文字邊界框:
   Step1 : 利用 Soft NMS 篩選預測框和預測框，且信心值高於 0.875 者
   Step2 : 語義分割大於 0.3 者設為 1，其餘為 0
   Step3 : 使用 cv2.minAreaRect 求出能包含語義分割為 1 者的最小面積矩陣
   Step4 : 得到最小面積矩陣的 4 個角落座標後，分別以歐式距離計算離 4 個角落座標最近的語義分割為 1 的點為新的角落座標

參、 資料處理

1. 由於使用 Mask RCNN 訓練，因此除了給定邊界框真值，同時也需要給定語義分割真值。其中邊界框真值取 4 個角落座標的 x, y 軸的極大和極小值；語義分割則給定 4 個座標軸內的所有像素點為 1，其餘為 0。

肆、 訓練方式

1.設定 Anchor box 大小及長寬比分別為:

大小 : 32^2, 64^2, 128^2, 256^2, 512^2

長寬比 : 0.3, 0.5, 1.0, 1.5, 2.0, 2.5, 3

2. Optimizer 為 SGD (lr=0.005, momentum=0.9, weight decay=0.0005)

3.資料分為訓練集 (23,400 筆) 與測試集 (600 筆)

4. Batch Size : 6 Epoch : 25

5. Data Augmentation : 水平翻轉

伍、 分析與結論

這次競賽是使用一般的物件偵測模型來實作，而大部分的模型輸出皆為水平的矩形，需額外調整預測框到最小包含文字區域的 4 邊形結果才會準確，因此可以嘗試專門處理場景文字的模型(例如 EAST)為改進方向。

陸、 程式碼

詳見附檔

柒、 使用的外部資源與參考文獻

https://www.twblogs.net/a/5d720e21bd9eee5327ff7374

聯絡資料

- **隊伍**

隊伍名稱	Private leaderboard 成績	Private leaderboard 名次
報名系統一致	0.634161	10

- **隊員(隊長請填第一位)**

姓名(中英皆需填寫)	電話	E-mail
吳亦振 (Yi-Chen Wu)	0972771286	a0972771286@gmail.com