

附件 一、資料集說明文件格式(範本)

「合成場景影像品質變化資料集」(資料集名稱)

資料集說明文件 v0.2.2

建置團隊資訊

子計畫名稱：___可信賴的生活場景文字識別技術

子計畫主持人：___陳祝嵩___

說明文件版本：___v0.2.2___

日期：___111.08.14___

1 資料集基本說明

資料集名稱		合成場景影像品質變化資料集
資料集摘要 (含目的/用途/價值/標註重點等)		• 原清晰場景影像 • 合成雨滴之場景影像 • 合成反射之場景影像 • 合成模糊之場景影像 • 含本地資料，目標使用族群為本地之一般社會大眾，資料之隱私保護將配合總計畫訂定之資料治理規範實行，例如將場景影像中之人物上馬賽克。
資料集聯絡人		(姓名/E-mail/聯絡方式) 李偉華 / weihua9217@gmail.com.tw 廖禾豪 / 110753115@nccu.edu.tw 劉宸羽 / C24152684@gmail.com
資料類型		選擇(結構化數據，語音，文字，影像，其他) 影像
資料	資料集類型	影像
	儲存格式	png
	來源機構	(若有多個提供機構請分列) Youtube
	收集期間	110 年 11 月至 111 年 8 月
	案例數	合成雨滴之場景影像 22103 張 合成反射之場景影像 22103 張 合成模糊之場景影像 22103 張
	筆數/案例	(若每個案例收錄多筆資料請註明)
	總共容量	32.1 G (三種合成影像資料集，各 10.7GB)

	適法性基礎 (本項暫不對外公開)	☐ A 類(新收案) : 000 例 ☐ B 類(曾取得概括同意, 已補告知程序) : 000 例 ☐ C 類(未曾取得同意, 已補告知程序) : 000 例 ☒ D 類(資料已去識別, 無法進行告知) : 000 例 ☐ E 類(資料無個資, 不需進行告知) : 000 例
	其它說明	
標註資料	標註工具	開發之自動合成方法
	標註方式	☐ 自動標註 ☐ 人工標註 ☒ 自動標註+人工校正 ☐ 其他 : _____
	註記方式	☐ 有註記 : _____ ☐ 其他 : _____
	標註儲存格式	
	標註情形	☐ 全部標註完成 : 000 例 ☐ 僅標註部份項目 : 000 例 ☒ 僅原始資料, 未加標註 : 000 例
	其它說明	
與其他資料庫的串接		☐ 保留與本團隊其他資料庫串接可能性 : 未來可透過假 id 串接本團隊其他的資料庫 ☐ 保留與國網中心資料庫串接可能性 : 與上傳國網中心的資料採同樣的假 ID, 以利後續資料更新/維護 (但僅限由本團隊主動進行更新與維護) ☒ 無串接規劃 ☐ 其他 : _____

2 案例相關資訊欄位說明

案例資訊項目		欄位名稱 (呈現於檔案的名稱)	欄位格式 (數值/字串/文字等)	範例/備註說明
個案編號				
基本 資料	出生日期			
	性別			
	其他			
	...			
資料 取得 設備 資訊	取得日期			
	資料取得設備 型號			
	其他			
	...			
每筆 資料 描述 (選 填)	代碼			
	分類			
	其他			
	...			
延伸 資訊	報告			
	其他			
	...			

3 標註說明範例

使用到下述檔案

標註項目	資料檔	標註檔

3.1 資料與標註資料的對應關係

3.2 標註內容

標註欄位說明

欄位名稱	說明	檔案截圖

4 工具

可以使用下列工具從儲存的檔案中讀取資料

資料類別	公開/免費軟體名稱/版本	商用軟體名稱/版本
資料	一般的影像瀏覽器 Python 之 opencv 套件	
標註資料		

5 去識別化

去識別化方式	■ 遵循本專案共同制訂方式
	☐ 本資料集自行定義方式 (勾選本項請逐欄位說明各欄位之處理作法，使用工具，參考資料與連結) 說明： ● 使用 FaceMaskDetection 工具針對場景影像進行人臉偵測，並且將該區域進行模糊化，工具連結： https://github.com/AIZOOTech/FaceMaskDetection ● 使用自行訓練之 yolov5 車牌偵測模型，將影像中的車牌偵測出來，並且進行車牌文字的代換