**機器學習 專案作業二**

本作業主要使用Python語言及Keras/Tensorflow/Pytorch實作CNN圖像分割。

作業內容包括：

* + 安裝與練習使用Anaconda裡的Jupyter Notebook系統。有問題可請教助教。
  + 資料集 (Endotracheal Tube)

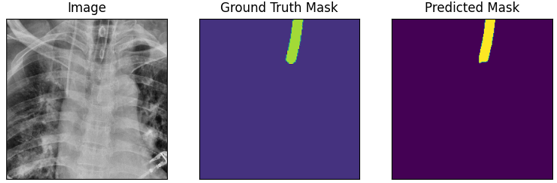
ETT資料集下載網址: <https://drive.google.com/drive/folders/18-nxkaKLNrZZb90fIlYXNg2lzLy-fhmn?usp=sharing>

* + 利用Python及Keras/Tensorflow/Pytorch建構CNN並分別進行以下操作。
    1. 請下載ETT資料集
    2. ETT資料集分成5個Fold，train為訓練集、val為驗證集、test為預測集；trainannot為訓練目標遮罩、valannot為驗證目標遮罩、testannot為測試目標遮罩。
    3. 請使用卷積神經網路(CNN)架構進行預測，並分別呈現績效。(可用Github open source現成的model, 例如，UNet, UNet++, DeepLab, …等，需要進行不同模型之績效比較、需要說明超參數之設定)
    4. 請使用IOU指標評估模型，此外**需自行設計三種自定義指標用於計算ETT端點的預測位置包括：平均誤差公分、誤差在0.5公分內準確率、誤差在1.0公分內準確率用於評估模型**。

|  |
| --- |
|  |

|  |
| --- |
|  |

|  |
| --- |
|  |

* + 1. 為分割氣管內管端點的y座標，為實際氣管內管端點的y座標。*P*為1公分在圖像轉換為pixel數，*K*代表測試資料集的圖片數量，**在此研究情境下，將每72 pixel轉換為1公分**。
    2. 流程實驗需進行以下架構：
    - 說明進行何種資料前置處理？
    - 說明網路的架構設置與介紹：model structure, number and size of feature maps, number and size of kernels, stride, loss function , activation function, optimizer, batch size, epoch, etc。
    - 請試著比較不同的設置會造成什麼影響? (例如，模型串接的方法、Input的設置等，請自由發揮。)
    - 分別列出5個Fold訓練資料及測試資料的預測績效，以及5個Fold的平均績效。
  + 視覺化預測結果，預測範例如下：
  + **實作流程建議**：
    1. 讀取圖像資料(同時載入image與mask)
    2. 將讀取的圖像資料進行前處理(包括圖像大小的調整、裁切：建議大小為256 x 256，或是調整成運算設備能負荷之程度，應注意裁切位置不能包含到mask)
    3. 套用強化方式至圖像資料上(例如：圖片翻轉、對比度強化等，但應注意強化方式的合理性)
    4. 選擇合適的模型進行訓練，包括FCN、U-Net、U-Net++、SegNet、DeepLab等(根據不同的模型，所適合的參數也不同，也可嘗試替換不同的骨幹架構到各種模型中)
    5. 設計合適的損失函數，例如圖像分割任務中常見的二元交叉熵、Dice損失函數等
    6. 將處理完的圖像載入模型中進行訓練(應設定適當的epoch或使用early stopping，以及learning rate)
    7. 建議可將訓練過程中較好的模型權重保存，以利結果的預測(例如，保存驗證集績效較好的權重)
    8. 以視覺化的方式呈現原圖、真實圖、預測圖三種結果
  + 訓練深度學習模型很花時間，建議及早做本次作業，以免屆時時間不夠
  + 遇到問題，可以向助教諮詢
  + 若有需要使用資訊中心AI平台者，可向助教登記後，老師幫忙申請。也可使用Google Colab
  + 請遵循學術論文寫作格式撰寫專案報告，內容至少必須包括下列要項：

**摘要** (學術論文摘要通常不分段，內容簡要地涵蓋動機、目的、方法與實驗結果)

**一、緒論**

**1.1動機** (所選用的資料集之動機)

**1.2目的** (想進行哪種分析，期望了解甚麼？)

**二、方法**

(1. 簡單說明程式架構，2. 執行程式的方法，無法執行程式，視為不正確)

**三、實驗**

**3.1資料集** (簡介資料集：名稱、筆數、欄位、型態…、並以資料表格呈現部份資料內容)

**3.2前置處理** (例如，圖像裁切方式、大小調整、圖像強化方式等)

**3.3實驗設計** (實驗如何進行、若有參數的話，參數如何設定等）

**3.4實驗結果** (呈現結果)

1. **結論**

* 評分標準:

內容(難度、豐富性、方法正確性)：60%，論文格式：20%，寫作技巧(文筆通順)：20%

* 任何複製剪貼網站資料或學長姐以前作業，皆屬抄襲，嚴重者零分。引用網站或他人資料，應註明引用來源。
* 此份報告最多15頁，請加註頁碼。請遵循投稿論文寫作格式，使用一欄式，無需製作論文目錄。可以觀摩 ACM Transactions on Knowledge Discovery from Data期刊論文的寫作技巧與格式。
* 紙本報告於上課前繳交或直接放入許老師的系辦信箱。專案程式直接上傳GitHub，專案名稱：hw2\_組長學號，並新增助教為專案成員，以利下載評分。缺紙本或程式皆視為未完成繳交作業。
* 繳交紙本時，在左上角訂一訂書針即可，無須其他任何裝訂。
* 每遲交一天，扣該次作業成績10分。
* 繳交日期：112/05/11（四）