

## 人工智慧於醫療應用與服務 Homework 2

NM6101072 張育嘉

1. 訓練 AI 模型分辨 NIH Chest X-rays 資料集，並分析訓練結果。

程式環境：Python3.8

函式庫：

Torch==1.10.2

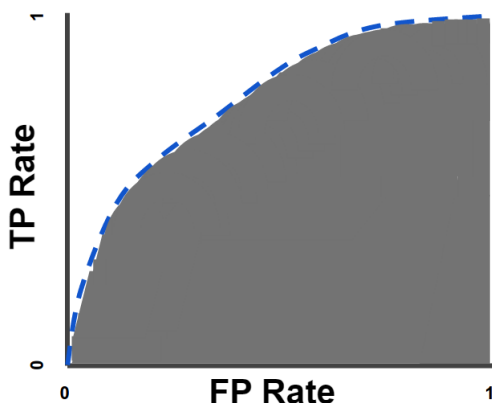
Torchcam==0.3.2

Torchvision==0.11.3

訓練模型程式為 retrain.py、nih\_labels.csv 為小型資料集、nih.py 為預測模型表現的程式。

這邊選擇醫療 AI 模型 DenseNet121 做訓練，因其特點是加強特徵的傳遞還有減輕梯度消失問題，很適合用做醫療影像學習，模型評量指標選用 AUC，用 sklearn.metrics 實現。

ROC 曲線用來衡量檢測有無偽陽性，而 AUC 則是 ROC 曲線下的區域，面積由 0 到 1，越接近 1 代表模型越準確。(TPR 是真陽率，FPR 是偽陽率。)



在 kaggle 上訓練，因此照片路徑

為 ../input/data/images\_001/images/00000001\_000.png。先訓練整個資料集，

查看模型表現，學習率為 0.01，優化器使用 SGD+momentum，損失函數為

binary cross entropy loss

，是參考[2]的做法，若 loss 下降太慢則再調降學習率。表 1 是經過 9 次迭代後，各個類別的總體 AUC 表現：

表 1

label	auc
Atelectasis	0.5
Cardiomegaly	0.5
Consolidation	0.5
Edema	0.5
Effusion	0.5
Emphysema	0.5
Fibrosis	0.5
Hernia	0.5
Infiltration	0.5
Mass	0.5
Nodule	0.5
Pleural_Thickening	0.5
Pneumonia	0.5
Pneumothorax	0.5

結果並不如預期，AUC 只有 0.5，模型基本上是在猜測，沒辦法取代醫生。

圖 1 是迭代 9 次的 loss 值，驗證部分的 loss 值沒有下降，可能是資料庫每個類別數量比例不一致導致模型無法收斂。

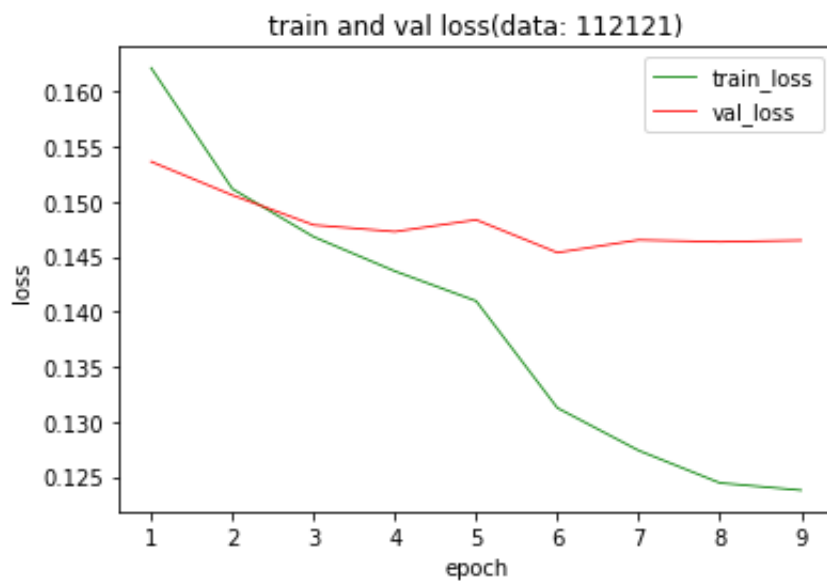
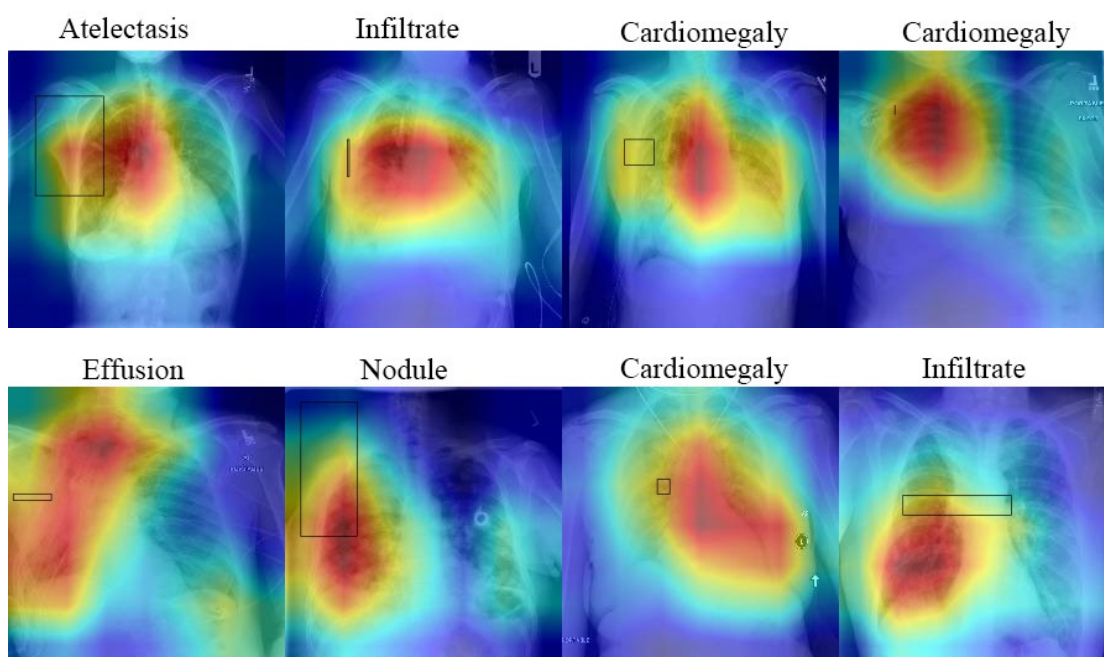


圖 1 整個資料庫十幾萬筆資料迭代 9 次的 loss 值

第二次訓練，採樣了比較小的資料集，只針對 Mass 標籤做訓練看模型表現如何。(資料集 train：15613 筆，Mass 佔 29%，val：2076 筆，Mass 佔 30%，test：2018 筆，Mass 佔 30%。)

經過第二次訓練，總體 AUC 表現還是跟表 1 一樣，模型預測能力只有 0.5，這次使用 Grad-Cam 查看模型關注區域，再跟 bounding box 比對有無吻合。隨機採樣 8 張，結果如下。



第二次訓練只迭代 5 次，損失值結果如下：



## 2. 心得

從 Grad-Cam 可看出，模型關注的地方並沒有跟標籤或 bounding box 一樣，可能是因為醫學病徵都比較微小，而資料在訓練過程中又從 1024\*1024 的尺寸縮小到 224\*224，照片可能因此損失重要醫學特徵，比如某些小於 1 公分的腫瘤或病灶可能因為照片縮小而難以學習到。

## 3. 參考資料

- [1] ROC 曲線和 AUC <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>
- [2] Rajpurkar, P., Irvin, J.A., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D.Y., Bagul, A., Langlotz, C., Shpanskaya, K.S., Lungren, M.P., & Ng, A. (2017). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *ArXiv, abs/1711.05225*.