

HTC:

## Weakly-Supervised Learning for Disease Detection

### 問題描述

在醫學影像分析領域，精密的標記很難獲得。因此，在這個項目中，我們探索一個新的方向來緩解這個問題。用weakly-supervised 影像標記，我們想設計一種新的方法來自動識別潛在的、代表疾病根源的醫學影像異常區域。在這個項目中，我們專注於胸部X光影像。以 bounding box為基礎進行evaluation。

### 模型輸入/輸出

輸入: 胸部X光影像 (1024x1024) 以及對應的疾病標記  
輸出: 疾病根源的 bounding boxes + 疾病label

### 可用資料集

NIH ChestX-ray 14 dataset

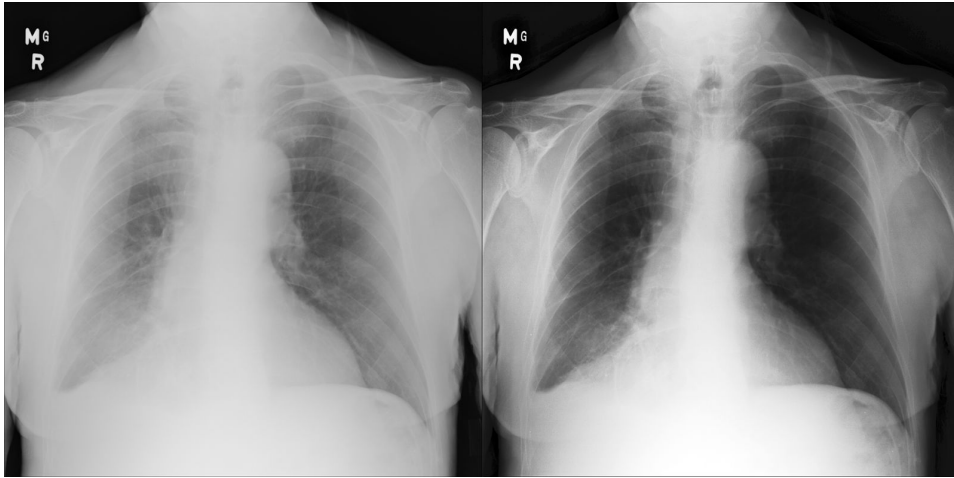
### 資料與前處理

使用NIH ChestX-ray 14 dataset，我們過濾掉not finding和Bbox沒出現的class最後剩下8個類別。圖片我們有使用Histograms Equalization和 CLAHE等方法預處理，最後resize成224x224。

#### Histograms Equalization

因為是黑白的圖片，所以圖片的Histogram分佈相當重要。我們發現在原本的dataset中有許多的圖片並沒有先經過histogram equalization的處理，予以處理之後可以讓我們所有data的黑白照片標準化。

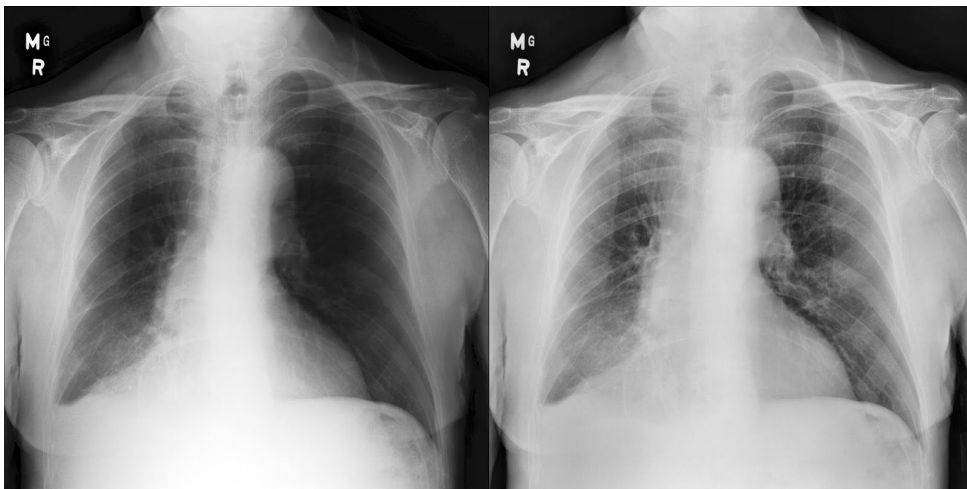
下圖：Histograms Equalization前/後



### CLAHE (Contrast Limited Adaptive Histogram Equalization)

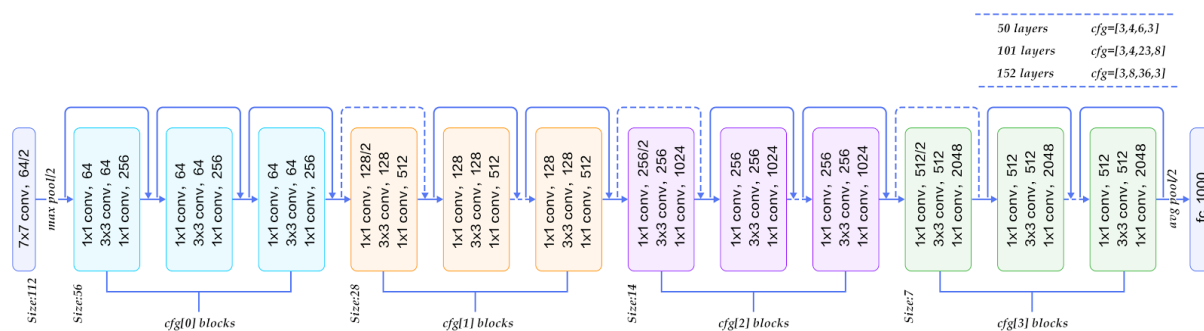
利用Histograms Equalization有一個小缺點，就是他是使用global contract of images當作標準化的標準，但是有時候這反而會讓我們的圖片在重點區域有over-brightness，也就是過度的黑白對比(可以發現上圖中右邊，HE過後的肋骨非常不明顯)，因此我們需要用到CLAHE的技巧。透過將一張圖片切成小張小張的tiles，能夠限制contract的差異大小，避免過度失焦。

下圖左：Histograms Equalization，下圖右：再經過CLAHE



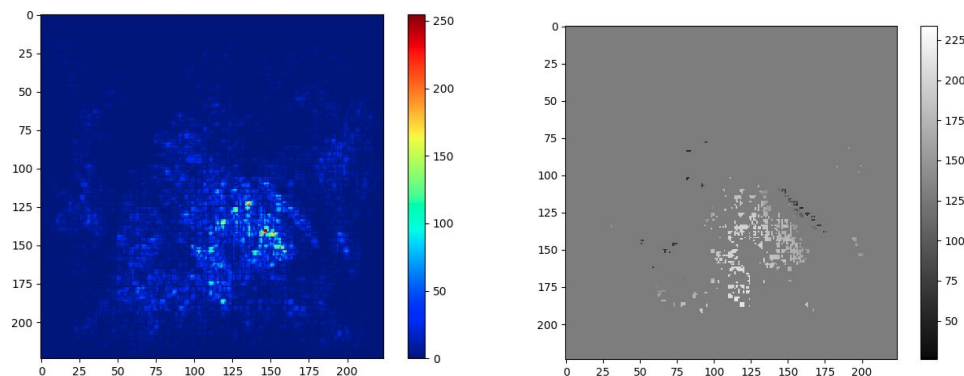
## 訓練方法

我們使用ResNet-50 和 InceptionResNetV2 兩個不同的model去訓練。



(Picture Reference: [http://book.paddlechrome://history/epaddle.org/03.image\\_classification/](http://book.paddlechrome://history/epaddle.org/03.image_classification/))

我們希望先做出一個不錯的classifier，再根據該classifier最後一層layer的saliency map，來推斷network在判斷這張X光是屬於何種疾病時，會注重於影像的何處。最後將所有heatmap的值高於經過fine tune過後的值框出。



此圖為心室肥症X光照的saliency map，可以看出它關注於左胸心臟的部位

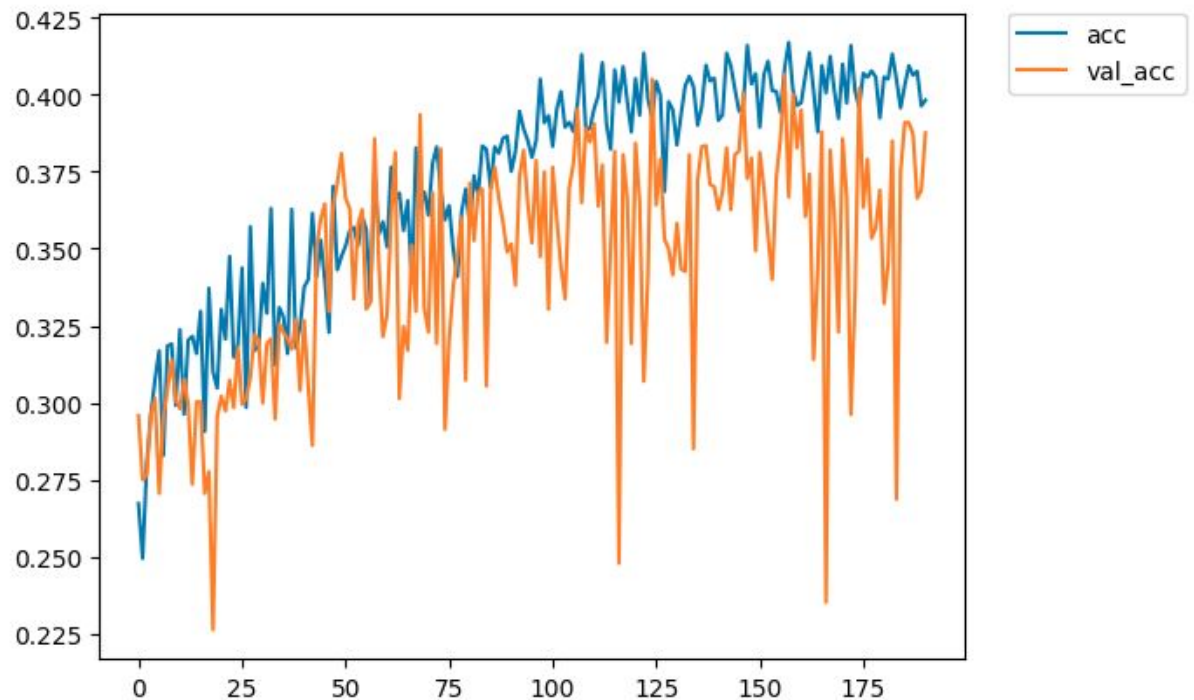
## 模型架構與Classification結果

下列Model的loss function皆選categorical\_crossentropy。

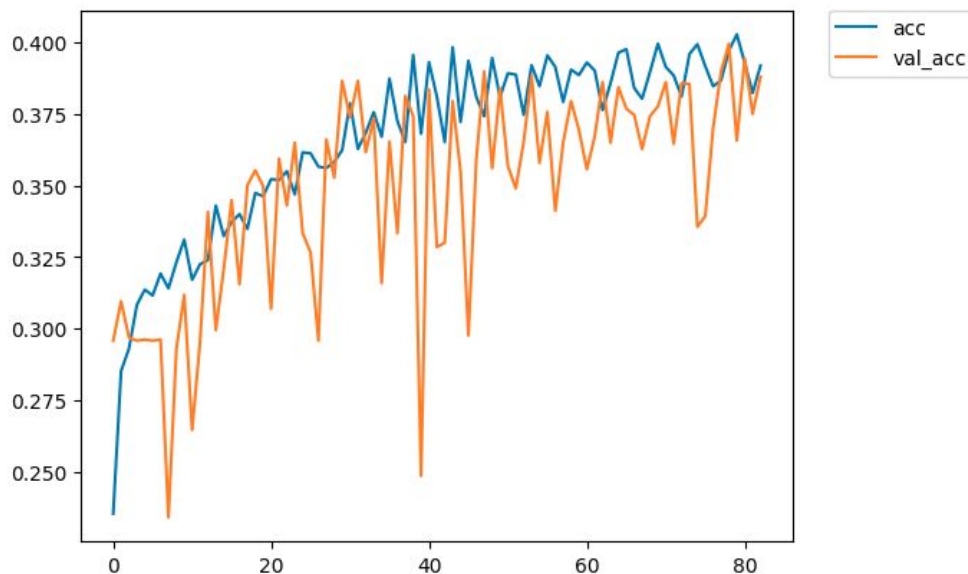
1. Model採用InceptionResNetV2，並使用 [ImageNet](#) 的pre-trained weights。並將network output flatten後接兩層2048(或512)的Dense layer。嘗試了許久，在validation的accuracy始終在30%附近徘徊，故後來就捨棄該模型。

2. Model採用ResNet50，並使用 [ImageNet](#) 的 pre-trained weights，並將其layers freeze住 ( untrainable)。在network output flatten後接兩層2048的Dense layer。

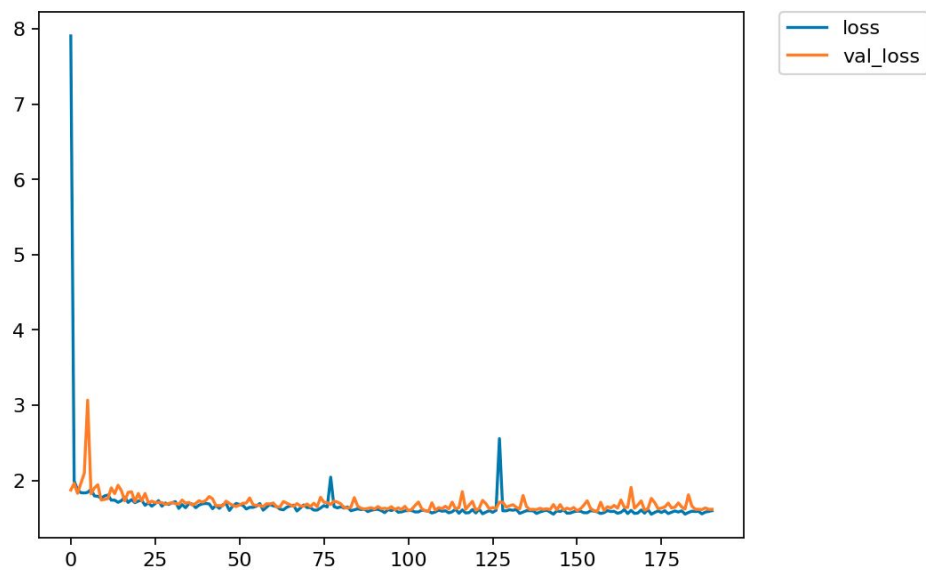
訓練結果 ( 使用440筆 validation data)



訓練結果 ( 使用2700筆 validation data，較穩定)



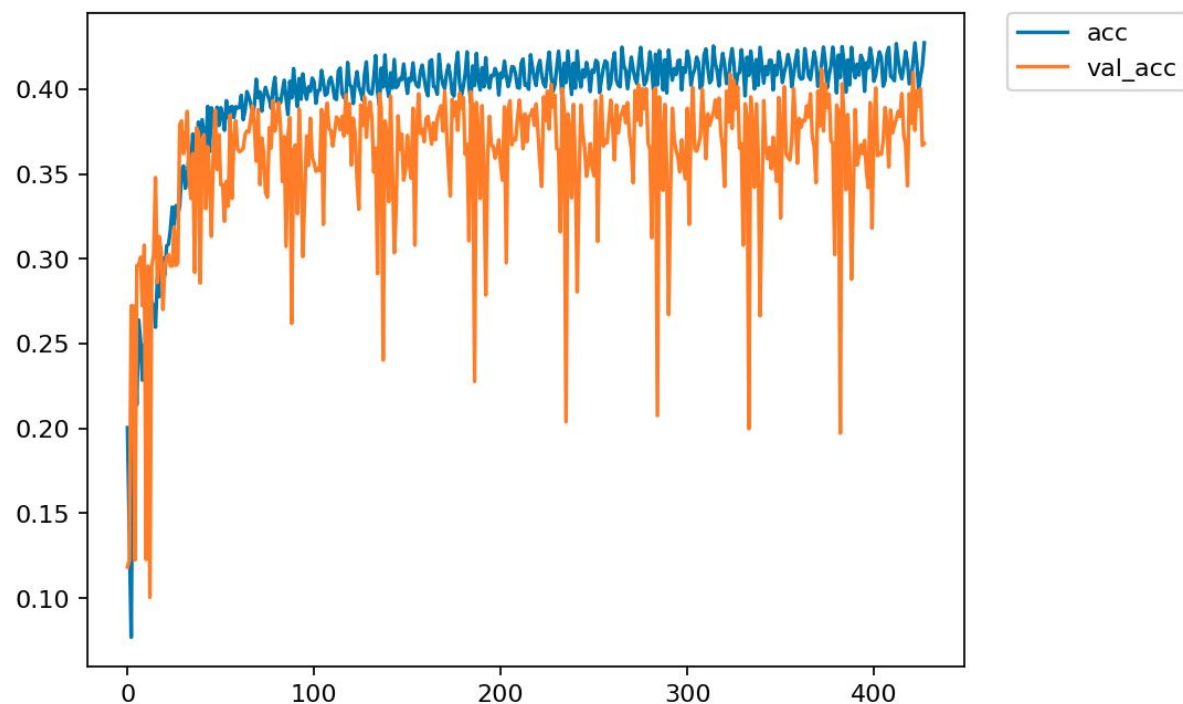
可以看到training accuracy 大概到 40%附近時就到上限了，而validation accuracy亦然。Validation的accuracy始終在35%~40%徘徊，略輸training 一些。



若觀察loss的話，可以看到training 跟validation loss都降很快且達平衡了，model難以在進步。

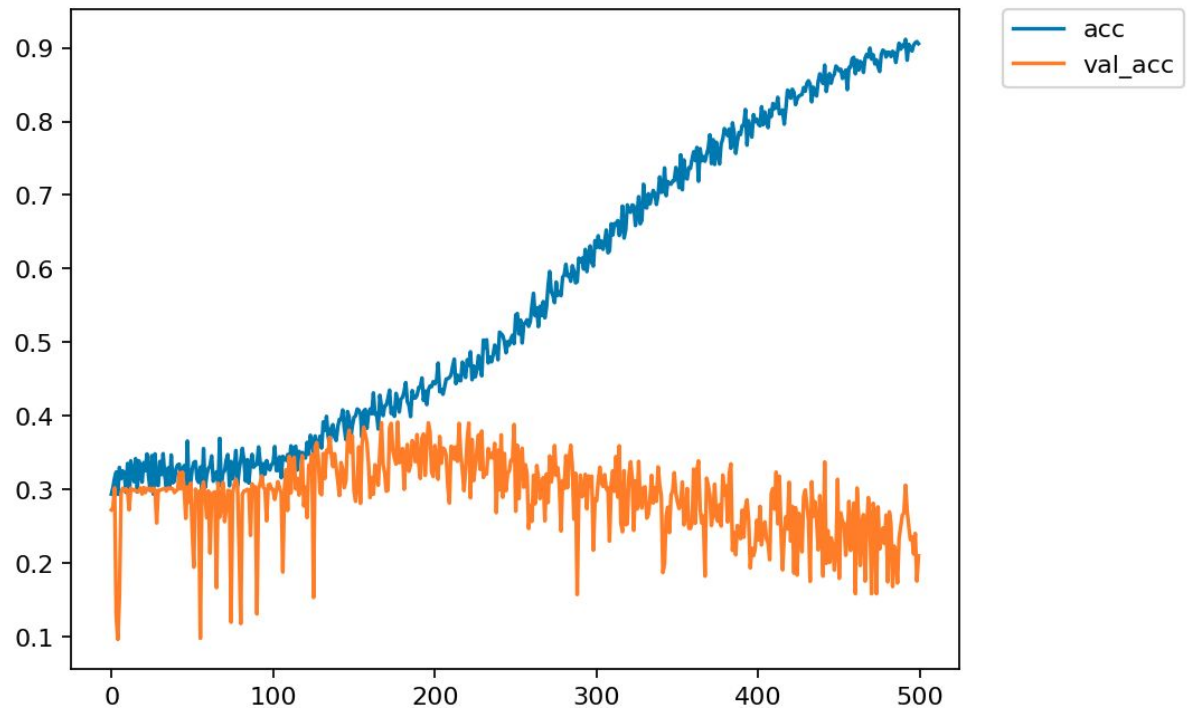
3. Model採用ResNet50，並使用 [ImageNet](#) 的pre-trained weights，並將其layers freeze住 ( untrainable)。在network output flatten後接一層4096的Dense layer

。訓練結果



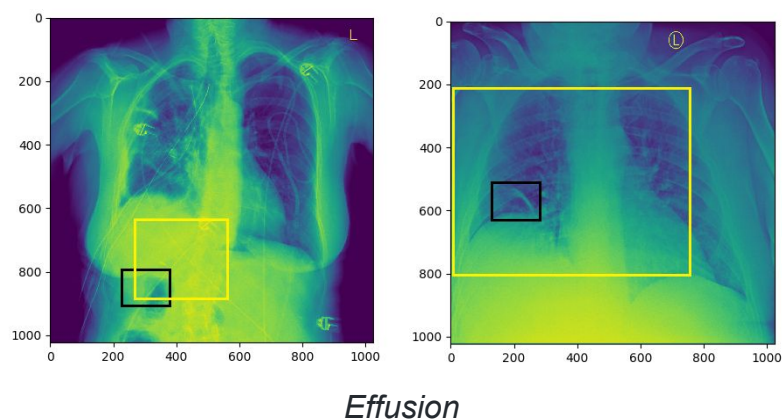
只使用一層 4096的dense比使用三層要更穩定一些，可以看到training的accuracy穩定於40%。此model的validation 的accuracy上限也是約40%。

4. Model採用ResNet50，並不使用任何pre-trained weights，也沒有freeze住任何的layer。在network output flatten後接一層4096的Dense layer。（模型與第3點相同，差再沒有load pre-trained weights）

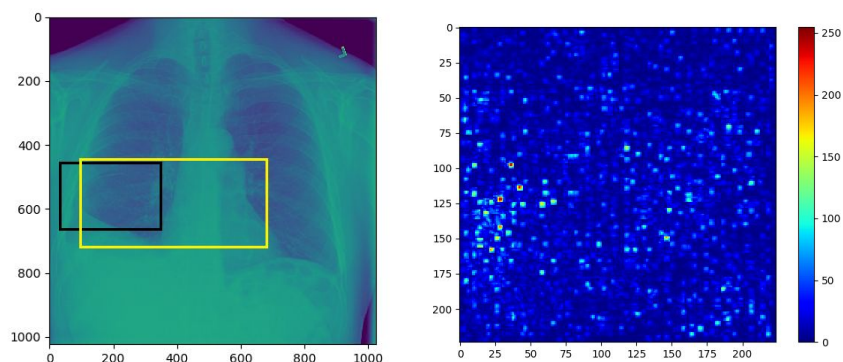


可以看到一樣validation 在約40%後達到上限，之後training繼續下去就overfitting了。故我們推測此模型在本次task的極限在約40%附近。而本組別online judge分數僅0.023，相當不理想，還有可能是boxing的問題會於下說明。

## Boxing結果







Effusion 的 boxing 結果與 saliency map

做法是抓取 saliency map 上的紅色像素值作為框選基準，大於給定的threshold即納入 boxing範圍內。較為顯著的問題是經常到outlier的影響導致框出的結果過大，我們嘗試將像素值在一定範圍內取平均以剔除 outlier，但推測因為範圍不夠大，並沒有顯著進步。

## 結論

故綜合上述實驗結果，目前使用的模型架構，在本次task預測疾病的準確率最佳為約41%，並不算是相當高的數字。有可能是選擇的CNN model不適合。而在一名放射線醫師 Luke Oakden-Rayner ( PhD Candidate / Radiologist)的[blog](#)中有提到

*"I believe the ChestXray14 dataset, as it exists now, is not fit for training medical AI systems to do diagnostic work."*

本次的資料集他認為由於label的正確性和label在影像辨識的代表性等因素，在機器學習的task中可能難以有很好的performance。這也許也是我們可以再探討和深入研究的部份。

## Reference

[Exploring the ChestXray14 dataset: problems]

<https://lukeoakdenrayner.wordpress.com/2017/12/18/the-chestxray14-dataset-problems/>

[ Deep Residual Learning for Image Recognition] <https://arxiv.org/abs/1512.03385>

[ *Classification of Common Thorax Diseases* ]

<https://github.com/srm-soumya/chest-scan-classifier>