

Weakly Supervised Learning for Findings Detection in Medical Images

9595 七講辦公室

r05945039 張宇軒, r05922091 何俊宏, r05945040 許靖, R05922120 蔡佳睿

Introduction:

每年都有非常多的人死於肺部相關的疾病，為了能夠診斷出疾病，會讓患者照 X 光片來做為診斷的依據，胸部 X 光的判讀需要有專門的放射科醫師進行判斷，這些在 X 光片中寶貴的資訊若能透過電腦自動判讀，能夠省下大量的時間及資源，達到電腦輔助診斷的結果。

在我們的研究中，我們使用深度學習的方式閱讀 X 光片，透過神經網路學習到每張 X 光片中關於肺部疾病的特徵，進而達到分類的目的，在深度學習的領域中，分類圖像中的物件種類，偵測圖像中的物件位置已經有相關的研究，對 X 光片做疾病分類若能成功，對 CAD 系統有大量的幫助。

Data description:

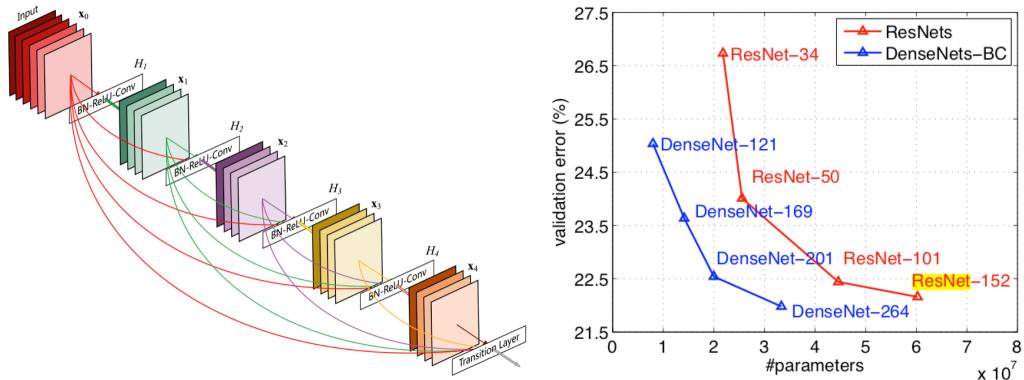
在我們的 dataset 中，共包含 112120 張胸部 X 光片，總共包含 14 種疾病，包含 Cardiomegaly, Consolidation, Edema, Effusion, Emphysema, Fibrosis, Hernia, Infiltration, Mass, Nodule, Pleural Thickening, Pneumonia, Pneumothorax，同一張胸部 X 光片可能有複數種疾病，每張圖片的大小為 1024*1024，其中只有 validation set 與 testing set 各 440 張 X 光片有疾病於 X 光片中的位置資訊 (Bounding Box)，並且只有 8 種疾病有位置資訊。

X 光的照片中，會有從前面或後面拍攝的照片，照片的位置也不一定固定在正中間，故之後使用資料增強時使用左右翻轉，旋轉及偏移等等認為是可行的資料處理。

Methodology:

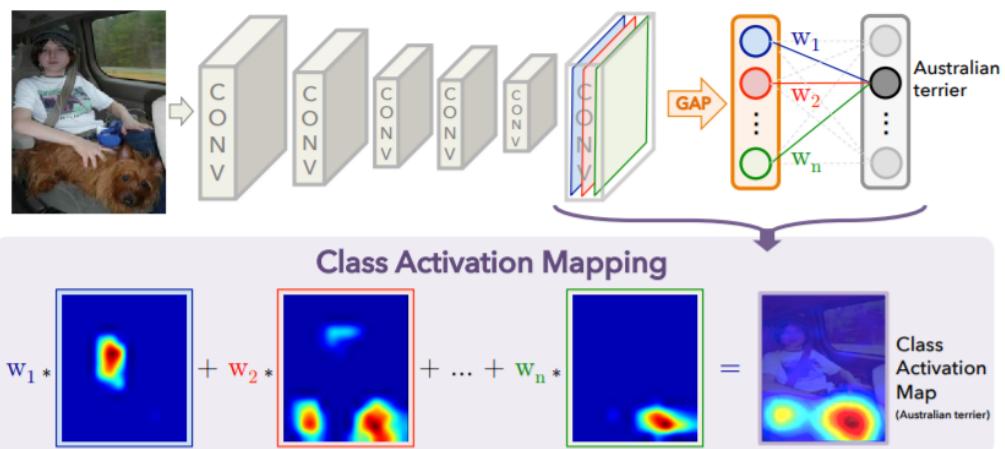
Model:

使用的 model 是 DenseNet (Densely Connected Convolutional Networks)，這個 model 來自 2017 年 CVPR 的 Best Paper Award，paper 裡提到，與 ResNet 相比，DenseNet 可以用大約一半的參數達到一樣的結果 DenseNet 的結果如下左圖所示，將每一層的 dense block 的 layer output 與後面的 layer output 一起運算，五層 layer 就有 10 個連接，DenseNet 中的 Growth Rate 代表每個 Dense block 的 layer output 產生的 feature map 數，在 ImageNet 的 validation dataset 上，DenseNet 得表現比 ResNet 的表現好很多， DenseNet264 用一半的參數量就達到與 ResNet152 一樣的結果。

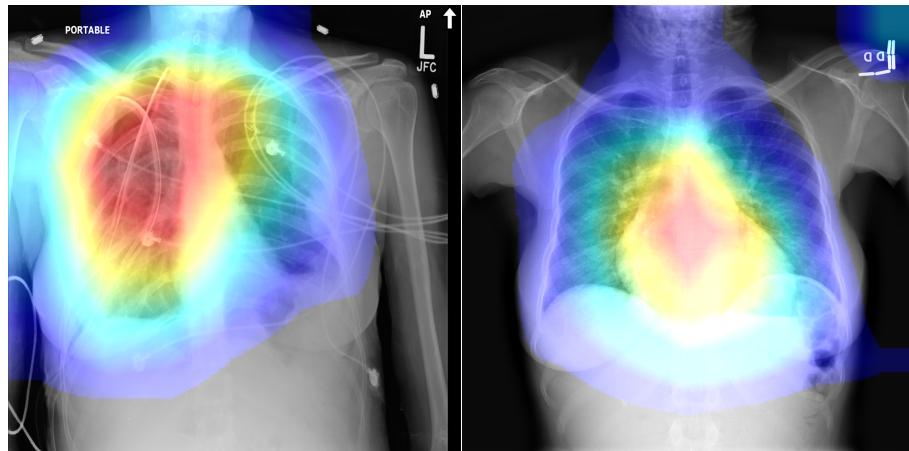


Class Activation Map:

如下圖所示，model 中的最後一層 Convolution layer output 通過 Global average pooling (GAP) 將每個 feature maps units 取平均並再接上一層 Fully connected layer 從 1024 dimension mapping 到 14 類，透過找出對於每一類在 fully connected layer 的 weights，則是 Convolution layer 的 feature maps 對於每個類別的權重，疊加起來就可以知道相對於這個 Class 的 Activation Map。



下方左圖為 predict 為 Pneumothorax 的 class activation map，右圖為 predict 為 Cardiomegaly。



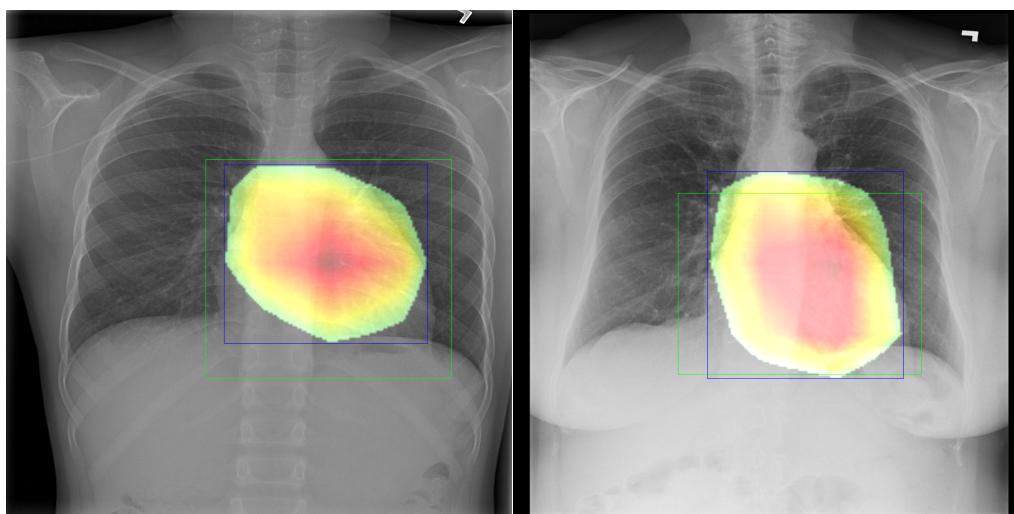
Experiment:

我們將圖片透過 INTER_LANCZOS4 resize 到 224*224 的灰階圖片，送入 DenseNet121 這個模型中，loss 用 sum binary cross-entropy，optimizer 用 Adam，用 loss 及 f-measure 作為標準，最後分類成 14 種疾病，並使用 class activation mapping 產生 model predicted class 的 heatmap，再依據 threshold 值留下 heatmap 中較重要的部分（排除藍色區域）並產生 bounding box 將其框選出來，bounding box 的 evaluation 是使用與 Ground Truth 的 IOU (Intersection of union)。

為了能更準確的預測疾病的種類，進而框出 bounding box，我們將資料中沒有 bounding box 的資料移除，剩下 8 種疾病的 Data 做 training，有效提高於 validation 與 testing set 的 f-measure 與 mean IOU，在訓練時將圖片做旋轉及左右翻轉，作為增強資料。

Result:

在預測疾病的準確度上，於 validation set 的 f-measure 能達到 0.55，而預測疾病位置的部分，平均的 IOU 能達到 0.2，下圖為 model predicted 的疾病於 X 光中的位置，包含較重要區域（部分 heatmap）、框選的 Bounding Box（藍色框），與 Ground Truth 的 Bounding Box（綠色框），雖然可大約預測到正確位置，但是與真正的區域還是有些差距。下圖為預測有 Cardiomegaly 之 X 光，而左圖之 IOU 為 0.72、右圖為 0.67，我們也發現 model 預測為 Cardiomegaly 時產生的 Bounding Box 是 8 種疾病中平均 IOU 最高的。



Conclusion:

使用 DenseNet 時，是否使用 pre-train model 對結果並不影響，不使用 pre-train model 也可以讓 f measure 達到 4.3 以上。

疾病分類的準確率上，使用 DenseNet121，loss 使用 weighted binary cross

`entropy`，同時將最後一層的 `pooling` 換成 `GlobalLSEPooling2D` 會有最好的結果。

用 `heatmap` 框出 `bounding box` 的實驗中，用 `RGB` 圖訓練出的結果，對比灰階圖訓練出的結果，灰階模型產生的 `heatmap` 比 `RGB` 模型產生的更不容易出現一大片紅色，而且也更接近 `Bounding Box`，在平均 `IOU` 上的結果比較好。

在 14 類疾病中，`Infiltration` 和 `Nodule` 的 `heatmap` 圖特別混亂，在疾病分類的階段，`Infiltration` 的準確率就特別的低，可以推測若分類的階段就能有較好的結果，會對後面 `bounding box` 的圈選有正面的影響。

在訓練的過程中，是否使用增強資料對結果有很大的影響，在我們的實驗中，嘗試過對圖片作旋轉及水平翻轉，整個訓練的過程變得緩慢，與原本相比，有使用增強資料後，分類的準確率反而降低，且訓練所需時間大幅增加，從 `heatmap` 的結果來看，有使用增強資料的 `heatmap` 較為集中，框出的 `bounding box` 較好。

模型深度的部分，`DenseNet121` 及 `DenseNet264` 在分類與 `heatmap` 兩個問題上並沒有太大的區別，選用 `DenseNet121` 即可。

Reference:

1. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases
2. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning
3. Learning Deep Features for Discriminative Localization
4. Densely Connected Convolutional Networks