

Bone Segmentation in Scanography

一、動機

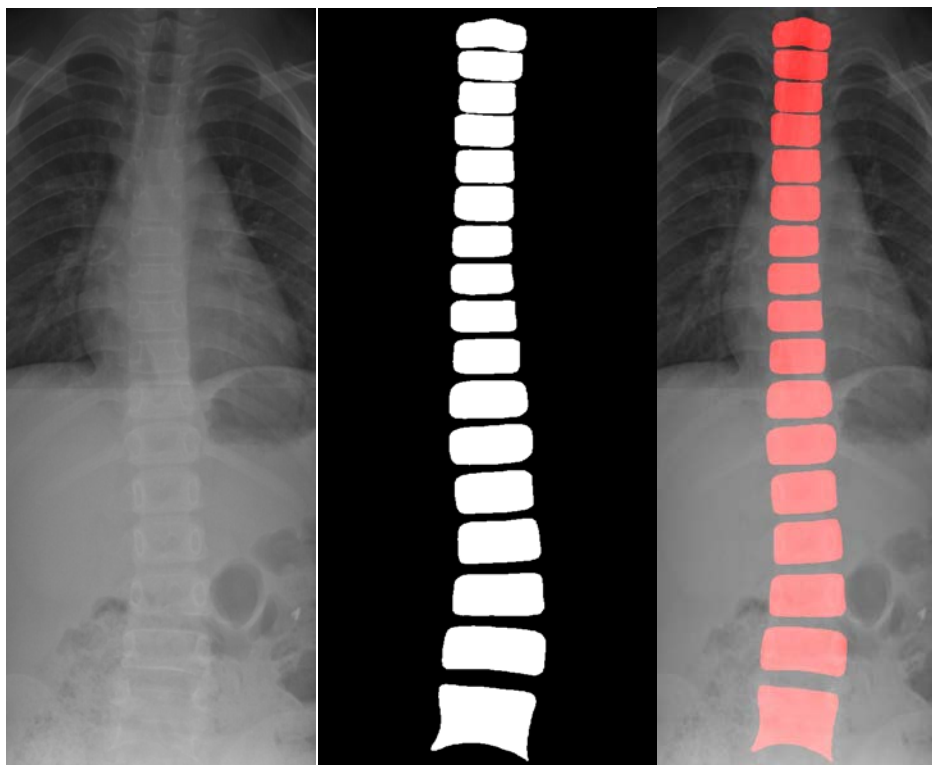
在平均壽命逐漸增加的社會上，肌肉骨骼老化是一個重要的問題，骨骼方面可能產生的問題包括癌症骨轉移，骨質疏鬆甚而壓迫性骨折，又以脊椎為最常見發生區域。Bone scan(圖一)為可近性佳並可評估全身骨頭活性的一項檢查，然影像解析度較低，正確評估病灶發生位置較為困難，常需要借助解析度較佳的其他醫療影像如 X-Ray, CT, MRI 等。我們希望可以在 Bone scan 上做 Segmentation，然而資料標註困難，因此選擇嘗試同樣為 2D 影像的 Scanography 做 Vertebra body 的 Segmentation，一來可以更清楚的對照出脊柱的位置，二來未來也可嘗試做 Transfer learning。



圖一、whole body bone scan。

二、 資料集介紹

取用 60 對 T-L 胸部 X 光圖與其 label 做後續的 Segmentation(圖二)。



圖二、右圖為胸部 X 光圖，中間圖片為 label 圖，左圖為兩圖結合的圖。

三、 方法介紹

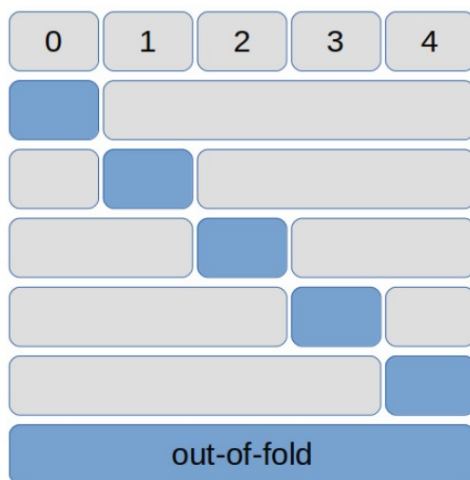
1. K-Fold(圖三):

步驟 1. 將數據集分成 k 個部分。

步驟 2. 選擇一部分作為測試集，其餘數據集仍然是訓練集。

步驟 3. 通過重複執行步驟 2，我們可以獲得 k 個相似但不同的模型。

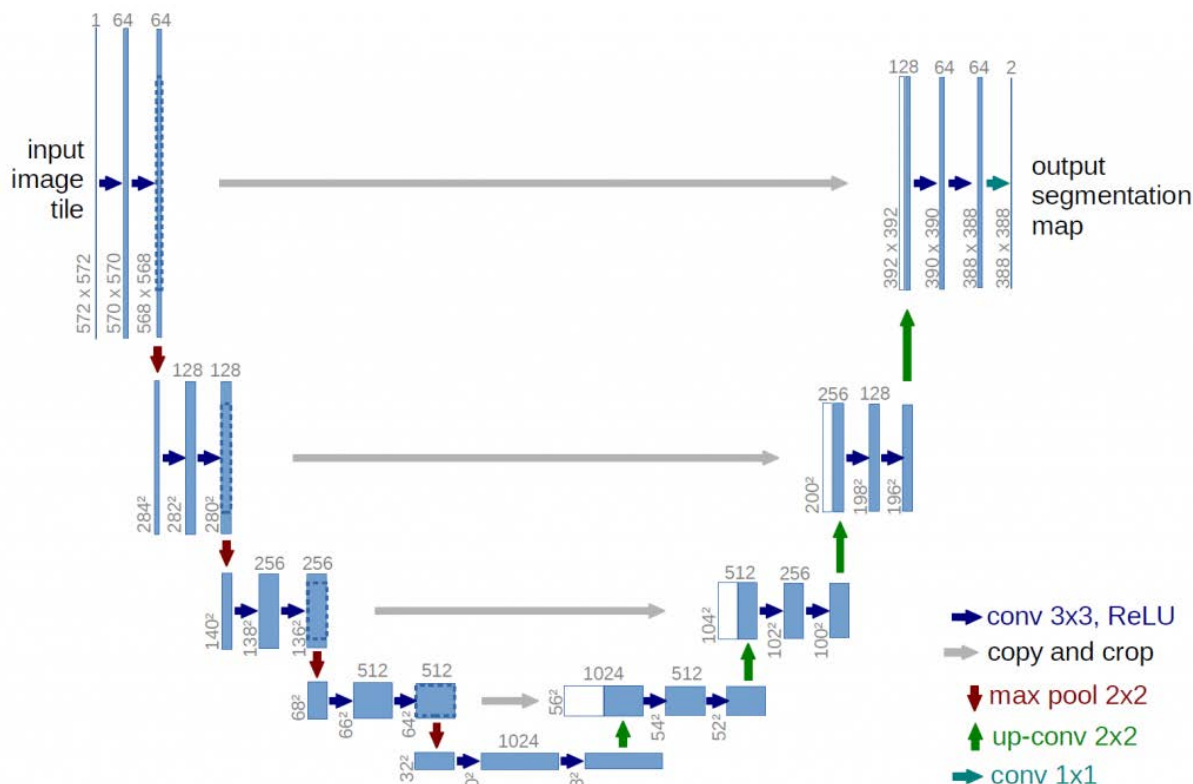
步驟 4. 在 Inference 階段，每個模型都可以計算自己的結果並對最終輸出進行投票。



圖三、此處以 $k=5$ 為例，out-of-fold 將得到 5 個結果，並投票選出最終結果。

2. U-Net ^[1]:

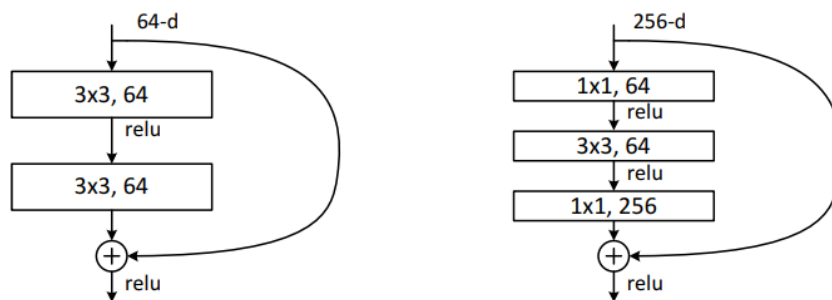
U-Net 是一種在醫學影像分割中最常用的模型，如下圖(圖四)所示，左側可視為一個 Encoder，右側可視為 Decoder。Encoder 中每層包含多個 Convolutional layer，不同層之間會使用 Max pooling 實現向下採樣；Decoder 中每層也同樣包含多個 Convolutional layer，不同層之間則會利用 Upsampling 進行向上採樣，而為了融合特徵信息，會在 Encoder 與 Decoder 同一層間加入 Skip connection。



圖四、U-Net 架構圖。

3. ResNet ^[2,3]:

Computer vision 大部分主流的網路幾乎都使用了 ResNet 的概念，它的出現開啟了深層網路的時代，ResNet 的網路設計就是簡單地增加一條路線做單純的加法（如圖五左），此作法使得深層網路訓練變得容易許多。若網路再更深，可以考慮使用 ResNet bottleneck block(如圖五右)，此方法用到的參數量較少，可以加快訓練速度。



圖五、左圖為 ResNet block，右圖為 ResNet bottleneck block。

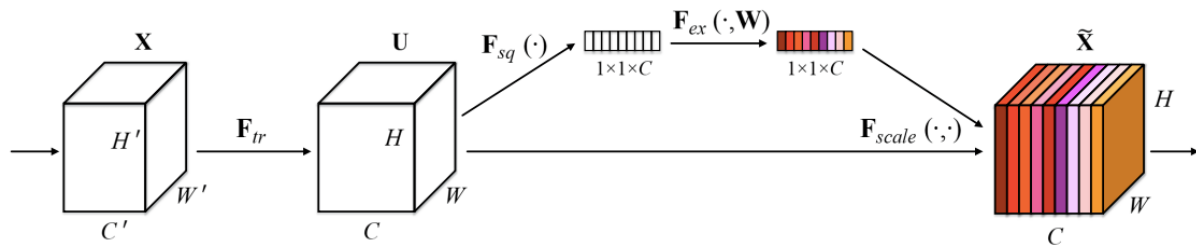
4. SE-ResNet [4,5]:

Squeeze and Excitation(如圖六)主要是去學習 Channel-wise 的 Attention，它會去學習不同 Feature Maps 之間的重要性，作為其權重來對其配重，而這樣的子架構可以再加入到其他模型中，例如：ResNet，因此 衍生出 SE-ResNet。

步驟 1. 透過 F_{sq} 壓縮操作，作者用 Global average pooling 作為 Squeeze 操作。

步驟 2. 透過 $F_{ex}(w)$ 的兩個全連接層和兩個非線性激活函數來學習不同 Channel 之間的 Attention weight。

步驟 3. 最後透過 F_{scale} 操作，輸出 re-wight 後的 Feature Maps。



圖六、Squeeze and Excitation 架構。

5. Attention UNet [6]:

在 UNet 上採樣的過程中，重建的空間信息是不精確的。為了解決這個問題，U-Net 使用 Skip connections 將來自下採樣路徑的空間信息與來自上採樣路徑的空間信息結合起來。然而，這會帶來許多冗餘的底層特徵提取，因此 Attention Unet 在 Skip connections 中加入了 Attention Gate，將有效地抑制不相關區域中的激活，從而減少冗餘特徵的數量。

步驟 1. Attention Gate 接受兩個輸入， g 是從網絡的下一層提取的。考慮到來自於更深的網絡，它具有較小的維度和較好的特徵表示。

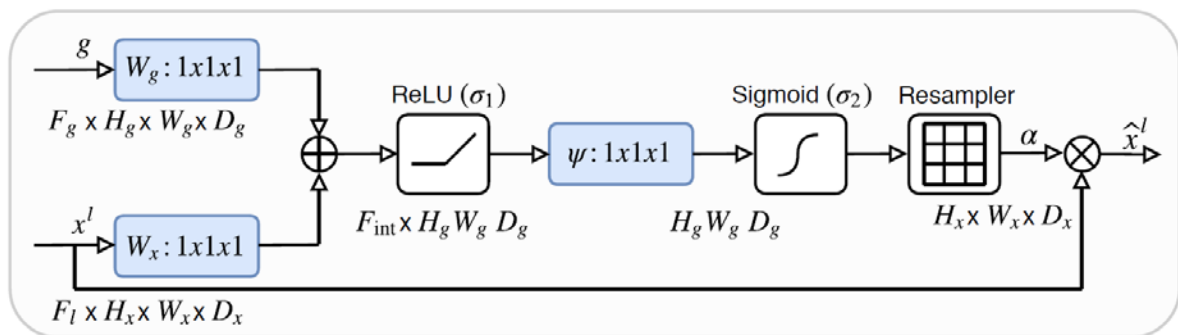
步驟 2. 將兩個輸入 Pixel-wise 相加。此過程導致對齊的權重變大，而未對齊的權重相對變小。

步驟 3. 合成的向量經過 ReLU 激活層和 1×1 卷積。

步驟 4. 這個向量經過一個 Sigmoid 層，該層在範圍 $[0, 1]$ 之間縮放向量，產生 Attention 係數(權重)，其中係數接近 1 表示更相關的特徵。

步驟 5. 使用三線性插值將 Attention 係數向上採樣到 x 向量的原始維度。

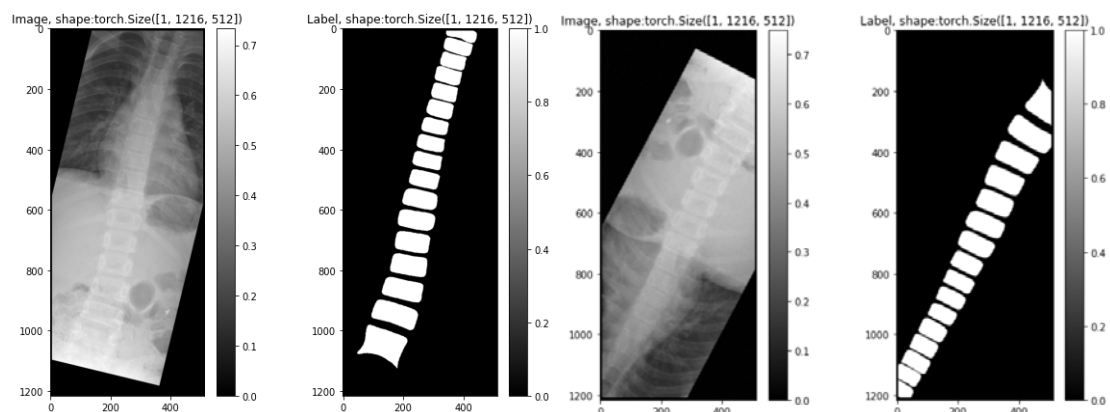
步驟 6. Attention 係數以 Pixel-wise 的方式與原始 x 向量相乘，根據相關性縮放向量，然後在 skip connection 中正常傳遞。



圖七、Attention Gate。

五、 前處理

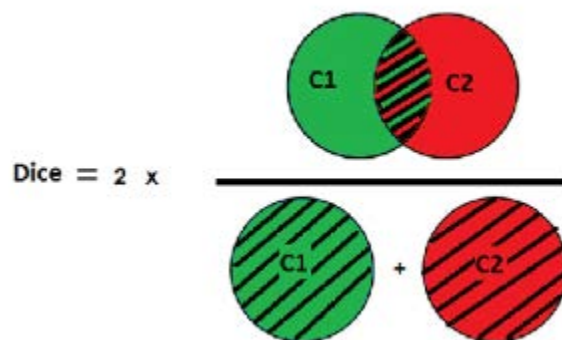
因資料較少的緣故，使用 albumentations 的 Shift、Scale、Rotate 來將圖片做 Data augmentation(如圖八)，並且使用 K-Fold 的方式，讓訓練集與驗證集可以被充分利用與學習。



圖八、為此實驗兩組 data 與 label 圖片的旋轉(隨機旋轉 -30° ~ 30°)範例。

六、 模型評估

本實驗採用 Dice coefficient 方法來評估模型(圖九)，並且將 K-fold 所產生的三組 Dice 做平均。

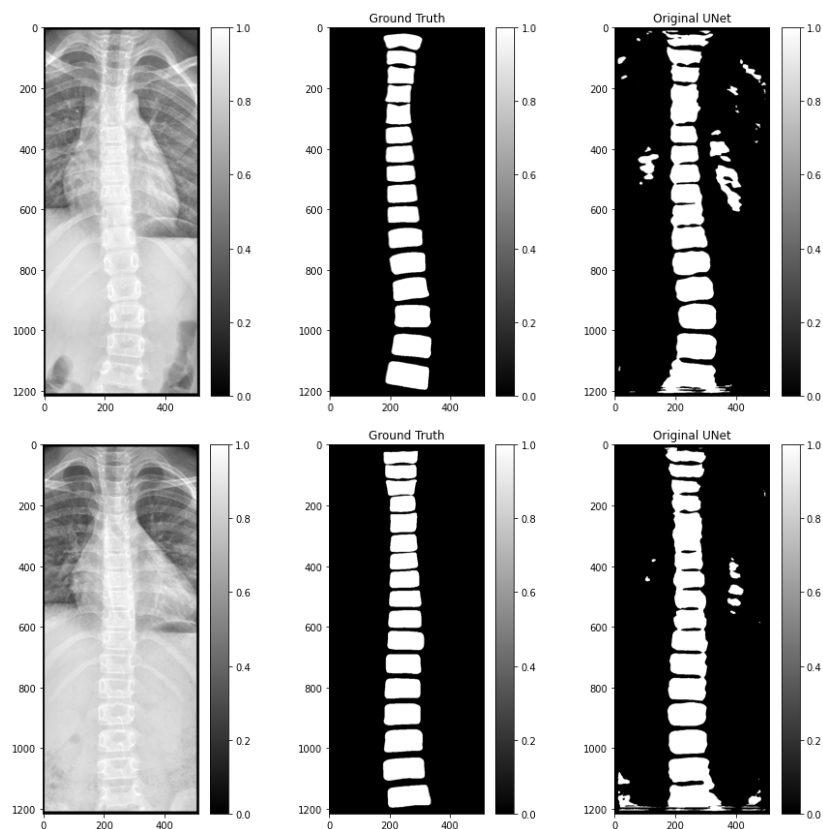


圖九、Dice coefficient，將兩個資料的交集乘以 2，並除以兩個資料的總和。

七、實驗過程與結果

1. UNet

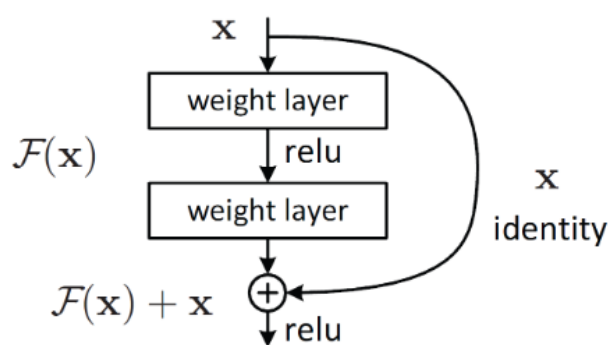
在這個方法中，我們在下採樣與上採樣的路徑中每層都只放兩層的 Convolutional layer，經過評估後 K-Fold Average Dice 為 0.718，由實際分割的結果來看(圖十)，周圍非脊柱的地方雜訊還是偏多



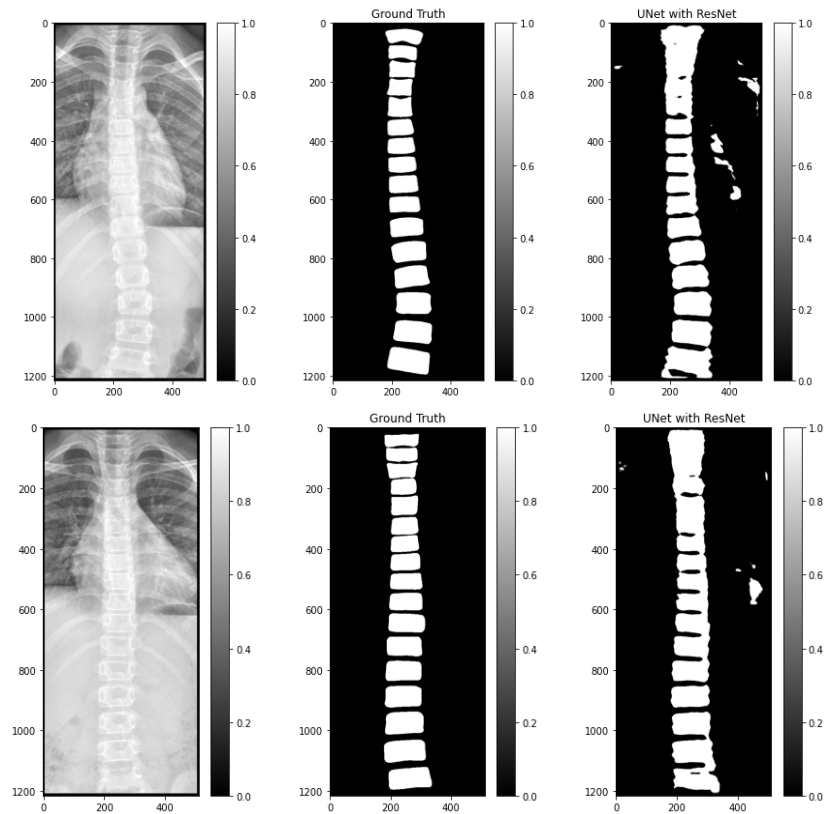
圖十、兩組為實驗 1 的實驗結果，左為胸部 X 光圖，中為脊椎 label 圖，右為預測出的結果圖。

2. UNet with ResNet

將 UNet 向下採樣的每層改為一層的 Convolutional layer 與兩層的 ResNet block (圖十一)，如此一來，每層就總共有五層的 Convolutional layer 去做特徵擷取，此方法經過評估後 K-Fold Average Dice 為 0.806，由實際分割的結果來看(圖十二)，周圍的雜訊減少了很多。



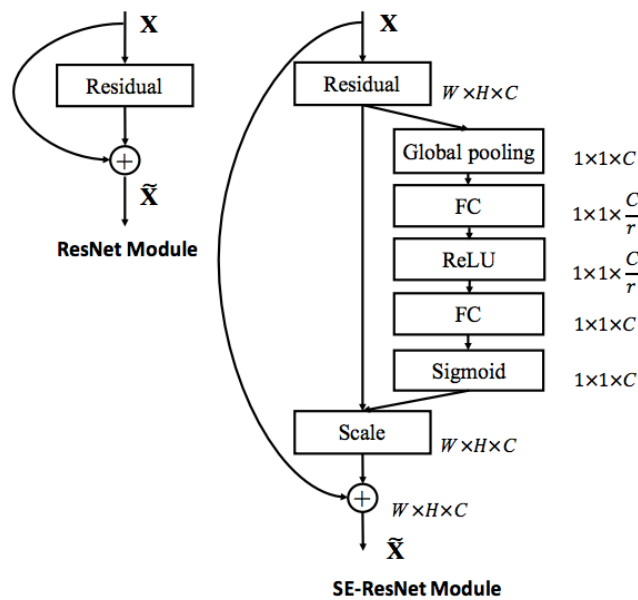
圖十一、為實驗 2 所用之 ResNet 架構。



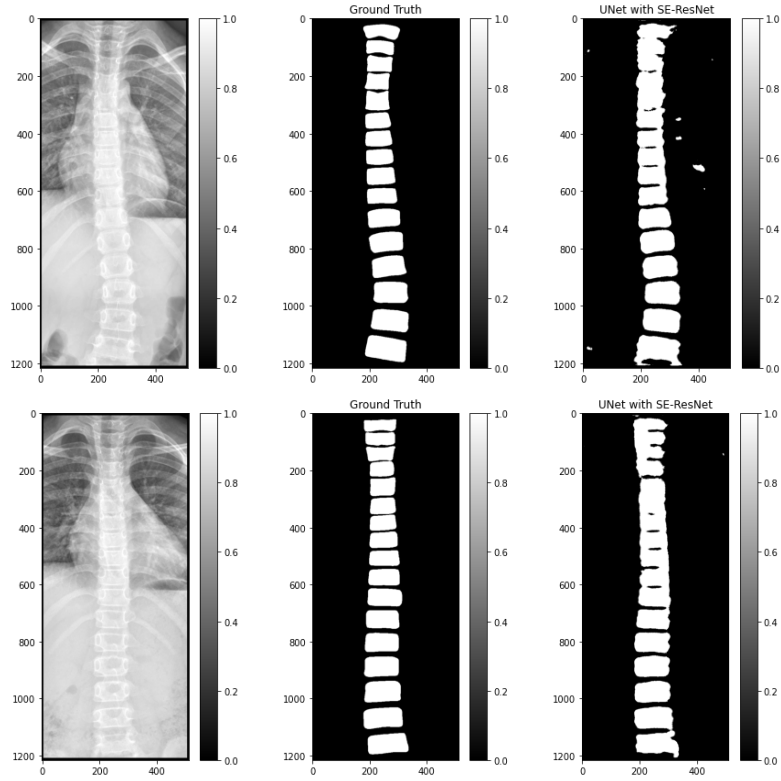
圖十二、兩組為實驗 2 實驗結果，左為胸部 X 光圖，中為脊椎 label 圖，右為預測出的結果圖。

3. UNet with SE-ResNet

將實驗 2 UNet 中的 ResNet Block 改為 SE-ResNet Block(圖十三)，經過評估後 K-Fold Average Dice 為 0.849，由實際分割的結果來看(圖十四)，周圍的雜訊又再變得更少，但是脊椎的部分還是會有連在一起無法清楚切分的問題。



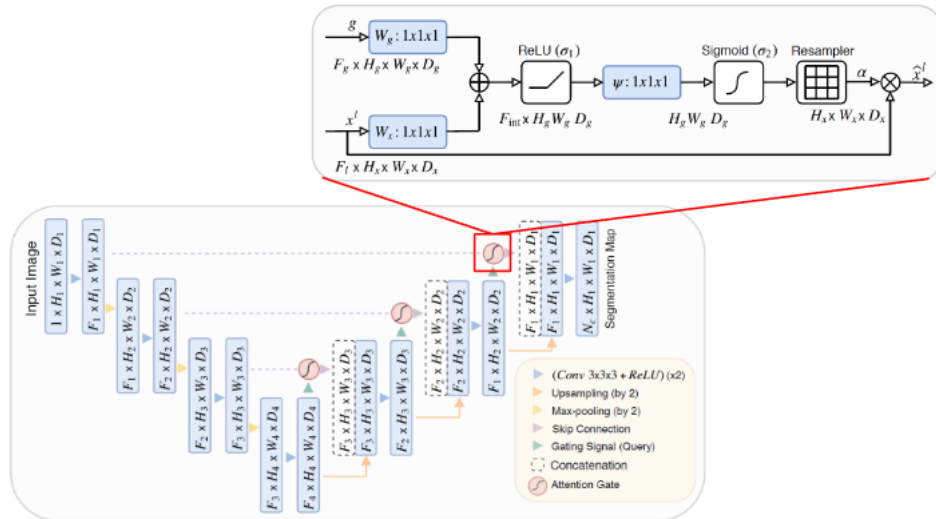
圖十三、左圖為實驗 2 的 ResNet 架構圖，右圖為實驗 3 所用之 SE-ResNet 架構。



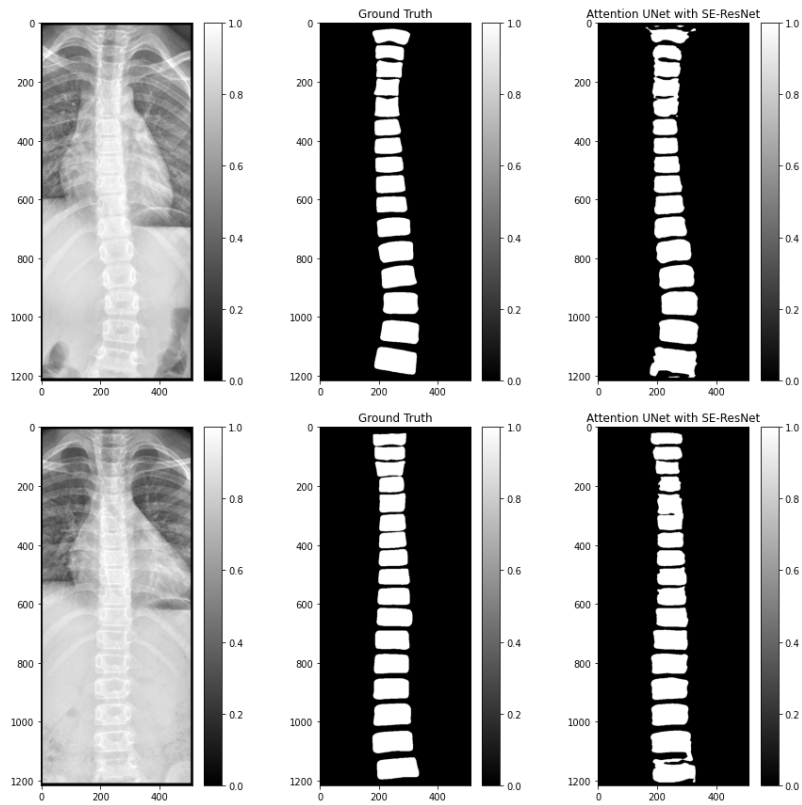
圖十四、兩組為實驗3實驗結果，左為胸部X光圖，中為脊椎 label 圖，右為預測出的結果圖。

4. Attention UNet with SE-ResNet

將實驗3再加上 Attention Gate(圖十五)，經過評估後 K-Fold Average Dice 為 0.86，由實際分割的結果來看(圖十六)，周圍的雜訊幾乎完全消失，而且每一塊脊椎也都更好地被分開。



圖十五、加上 Attention Gate 後的架構圖。



圖十六、兩組為實驗 4 實驗結果，左為胸部 X 光圖，中為脊椎 label 圖，右為預測出的結果圖。

5. 後處理

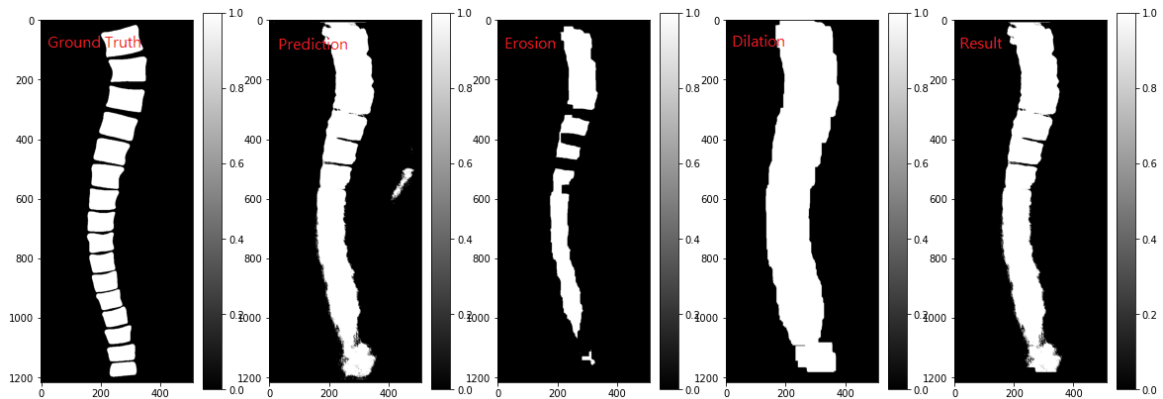
將模型的產出透過傳統影像處理的手段進行調整，這裡不涉及 AI 的參與。

首先對產出結果進行侵蝕(Erosion)，此舉會讓影像中的白色區域縮小，或者說擴大黑色的區域，而這會讓各節脊椎縮小，外圍的小雜訊則會因此消失。

接著進行擴張(Dilation)，此舉會讓影像中的白色區域擴大，各節脊椎會因此再變大。這裡刻意將放大的倍率略大於縮小的程度，因此各節脊椎在理想上會膨脹得比原本還大，也因此將彼此融合在一起，形成一個長條形的白色區域。

此時圖中應當剩下一個大範圍白色長條區域和若干小區域白色雜訊，我們僅挑選面積最大者為脊椎，剩餘區域全部捨棄(變黑色)，並以此作為 Mask 罩在模型輸出的原圖上。如此一來，僅白色區域，也就是各節脊椎會被留下，剩餘區域會自動變為黑色。

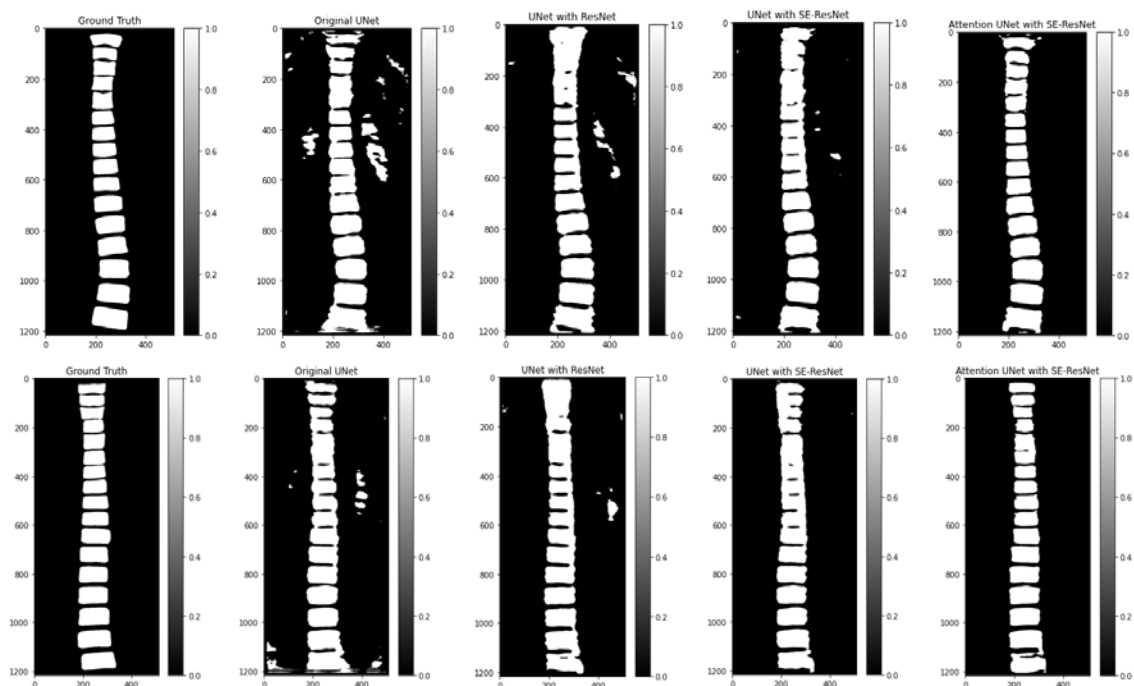
儘管理想上這個過程應該可以作用，但實際上在遇到特別困難的樣本時，很容易在挑選最大白色區域的階段捨棄沒能連接起來的剩餘脊椎，造成該組樣本的 Metric 特別低分，導致訓練後的分數並沒有進步，故最後捨棄。



圖十七、後處理視覺化示意圖。左至右分別為 Ground Truth、預測原圖、侵蝕處理後、擴張處理後(就是 Mask)、處理後結果

八、結論

由不同方法的實驗結果比對(圖十八)來看，我們得出的結論為在下採樣的過程中使用較深層的網路與針對 Channel-wise 及 Pixel-wise 加入的 Attention 機制都能幫助模型更好地萃取特徵以及進行特徵融合。



圖十八、比對的結果圖。

References

1. RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015. p. 234-241.
2. <https://medium.com/ai-blog-tw/deep-learning-residual-leaning-%E8%AA%8D%E8%AD%98resnet%E8%88%87%E4%BB%96%E7%9A%84%E5%86%A0%E5%90%8D%E5%BE%8C%E7%B9%BC%E8%80%85resnext-resnest-6bedf9389ce>
3. HE, Kaiming, et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. p. 770-778.
4. HU, Jie; SHEN, Li; SUN, Gang. Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018. p. 7132-7141.
5. <https://medium.com/@hupinwei/%E6%B7%B1%E5%BA%A6%E5%AD%B8%E7%BF%92-senet-squeeze-and-excitation-networks-52ad0a7fd307>
6. <https://kknews.cc/code/5jm26xk.html>