

CAT-UNet:使用 CNN 注意力模組與 TransUNet 於胸部 X 光影像中 肺腫瘤分割

研究生姓名：資訊工程系 鄭昕恬

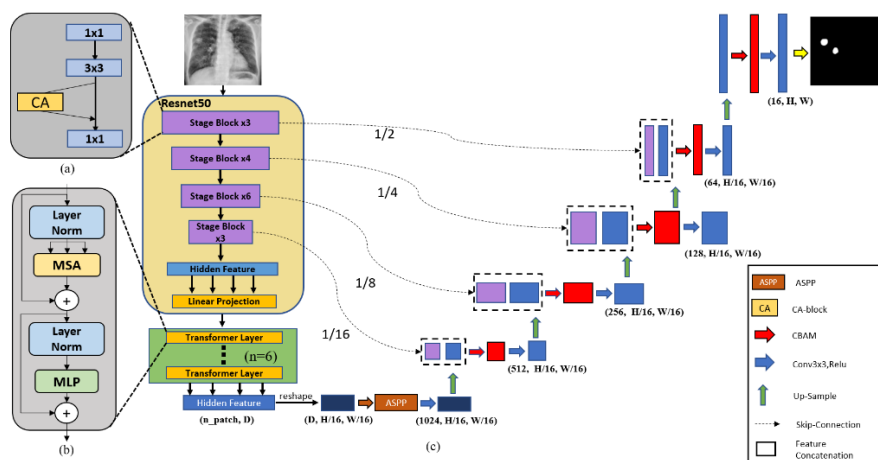
指導教授：國立雲林科技大學資訊工程系 張傳育特聘教授

摘要

胸部 X 光作為放射科中最普遍的檢查之一，被廣泛使用在肺部病灶區域的第一線篩檢，作為檢視不同種類肺部疾病的評判標準。本研究中提出 CAT-UNet，透過 CNN 與 Transformer 的結合使其成為一個 U 型的分割結構，用於進行胸部 X 光影像的肺腫瘤分割，引入 TransUNet 作為架構的基礎，在 CNN 方面使用 ResNet50 作為 backbone，添加 CA(Coordinate Attention) block、Convolutional Block Attention Module (CBAM)，透過通道與空間注意力機制，更準確的尋找到淺層特徵，提升對原始影像的恢復率。在 Transformer 方面則使用 Google 提出的 Vision Transformer (ViT)，增加對於深層特徵提取，並引入 Atrous Spatial Pyramid Pooling (ASPP)，融合多尺度的空洞卷積，捕捉更多影像中的資訊。使用類 U-Net 的方法，將取得的特徵資訊進行上採樣，輸出分割結果並疊加回原始的胸部 X 光影像達到肺腫瘤定位效果，提供可視化的解釋。實驗結果顯示對於正常與肺腫瘤的胸部 X 光影像，Dice 達到 89.06%，Sensitivity 達到 91.62%，Specificity 達到 98.64%，Accuracy 達到 95.15%，優於 U-Net、TransUNet、Swin-UNet，利用本篇論文提出的方法有效降低不同放射科醫師在診斷時產生的誤差，有助於提高醫師診斷決策的穩定性。

關鍵字：肺部腫瘤、胸部 X 光、深度學習

研究方法



CAT-UNet 架構圖

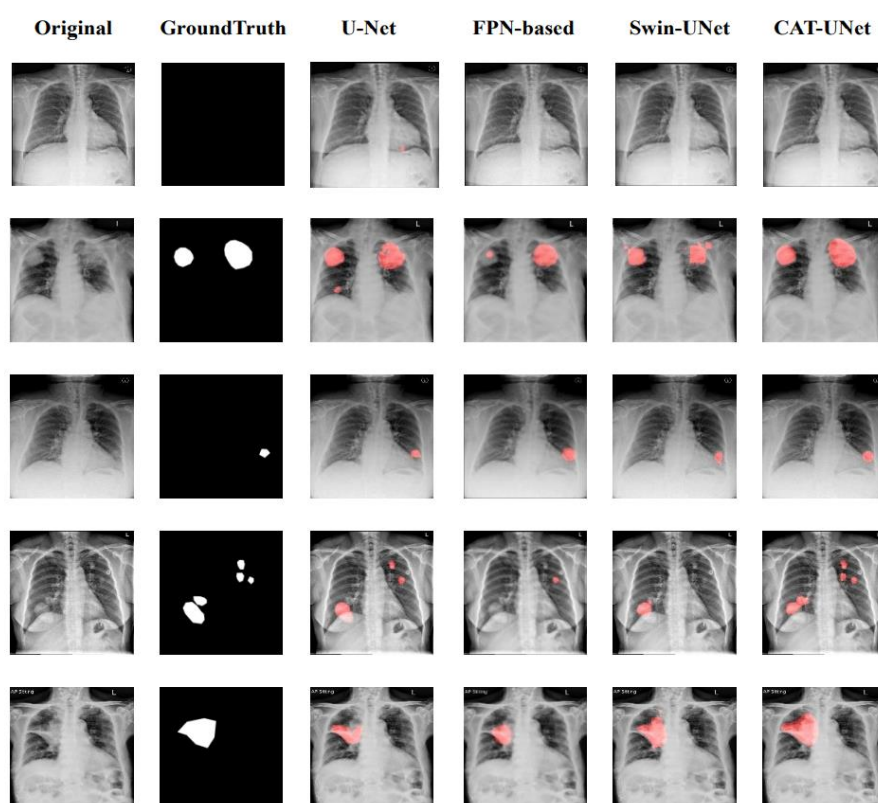
本論文中使用的分割模型參照 TransUNet 的基礎架構來搭建，提出一個應用於胸部 X 光影像肺腫瘤分割任務的模型稱 CAT-UNet。原始的 TransUNet 是一種 Encoder-Decoder 結構的網路，因為融合了 CNN 與 Transformer 兩者的優勢，常被用於醫學影像的分割，而本論文在 TransUNet 的基本架構上，

添加了 Attention Modules 與 Atrous Spatial Pyramid Pooling (ASPP)，提升在分割任務中的效能。

實驗結果

與其它分割網路的比較

Model	Dice	Sensitivity	Specificity	Accuracy
U-Net-based [7]	81.27%	80.40%	98.13%	89.29%
FPN-based [35]	69.14%	51.92%	99.89%	75.91%
Swin-UNet [26]	78.13%	78.16%	95.71%	86.9%
CAT-UNet(Our)	89.06%	91.62%	98.64%	95.15%



可視化結果比較

結論

本論文整合醫學臨床、醫學影像與人工智慧的技術，透過學習專業放射科醫生手動標記的 X 光影像，發展更精確的肺部腫瘤分割的研究，提出了 CAT-UNet，引入多尺度與注意力機制的概念，結合 CNN 與 Transformer 的優勢，建立基於深度學習的自動分割系統，本研究提出的 CAT-UNet 架構分別在 Dice 達到 89.06%，Sensitivity 達到 91.62%，Specificity 達到 98.64%，Accuracy 達到 95.15%，也進而改善了過去繁複又不精確的肺部腫瘤可視化方法，利用分割模型進行病灶區域的分割與定位，更具精準度與說明性，透過本研究的可視化方法，能夠成功地降低誤判率，輔助不同醫師在臨床上的診斷。最後，我們相信醫學影像結合人工智慧的部分，在未來會有更多不同面向的發展，期待在不同的醫學案例中都可以結合深度學習技術進行臨床上的輔助，在智慧醫療中達到更高的頂點。