

國立交通大學資訊工程學系

資訊專題競賽報告

基於深度學習之醫療影像分析-息肉影像分割

(Medical Image Segmentation by Deep-Learning – Polyp Segmentation)

專題題目說明、價值與貢獻自評（限100字內）：

在本項專題中，目標建立出能預測出腫瘤位置的AI模型、尋找優化預測效果的訓練方式(比如各種augmentation手法)，並將模型整合至網頁應用，針對使用者輸入圖像、影片等進行預測，以供需要的人(如醫師)使用。

專題隊員：

學號	姓名	手機	E-mail	負責項目說明	專題內貢獻度(%)
0816125	張紀睿	0933775101	jerryyyyy708@gmail.com	資料前處理 模型訓練 程式撰寫與整合 文獻閱讀 結果分析	100%

本專題如有下列情況則請說明：

1.為累積之成果(含論文及專利)、2.有研究生參與提供成果、3.為大型研究之一部份。

相關研究生資料（無則免填）：

級別年級	姓名	提供之貢獻	專題內貢獻度(%)

【說明】上述二表格之專題內貢獻度累計需等於100%。

指導教授簡述及簡評：

指導教授簽名：

中 華 民 國 一 一 一 年 月 日

一、 關鍵詞

人工智慧、機器學習、深度學習(Deep Learning)、卷積神經網路(Convolutional Neural Network)、資料增強(Data Augmentation)、醫療影像(Medical Image)、圖像分割(Image Segmentation)

二、 專題研究動機與目的

醫療影像 AI 的訓練常常需要大量資料，然而醫療影像的資料集在取得上較為困難，並且使用大量訓練資料需要消耗許多運算資源以及訓練時間，大幅提高了開發的門檻。因此，在本次專題中希望利用有限的訓練資料(息肉影像資料集 Kvasir-SEG[1] 1000 張圖片中的 900 張圖片，如圖 1)，透過進一步的資料增強、模型預訓練等方式，訓練出息肉影像分割的 AI 模型，克服醫療影像 AI 開發門檻較高的困境。

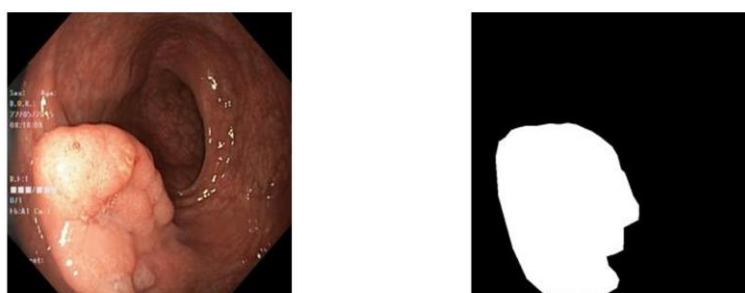


圖 1. Kvasir 資料及範例影像與其標註

另外，為了減少使用者(如醫生)在 AI 操作上可能會遇到的困難，專題最後將把模型串接至網頁應用，提供更方便的操作介面。使用網頁應用也能使整個預測過程更輕量化，使用者不須額外安裝應用程式就能有效利用模型，減少整體儲存空間的消耗。

三、 現有相關研究概況及比較

醫療影像一直都是 AI 常常應用的領域。在分類任務上，有肺部 CT 圖的癌症預測[2]；物件偵測任務上，有超音波影像的疾病偵測[3]。而在本次專題中，則專注於影像分割的任務，並且著重在息肉影像。在現有的息肉影像分割研究中，許多研究都有著不錯的表現。舉例而言，在此論文[4]中利用 PraNet 架構與 Multi-Scale 的方式訓練模型；另一項研究，Polyp-PVT[5]，則是針對腫瘤影像分割訓練了 Transformer 模型。而本次專題研究以 UNet[6]及 DoubleUNet[7]架構為基礎，結合影像增強與預訓練 Encoder 的方式，嘗試利用現有資料集提升模型的預測效果。

四、 專題重要貢獻

本次專題的三項主要貢獻為：1. 針對息肉影像分割，嘗試了許多資料增強的方法，進而尋找有效提升預測效果的策略。2. 針對 UNet 的 Encoder 進行了不同的預訓練方式，並找出能提升整體表現的方法。3. 將 AI 模型串接至網頁應用，提供輕量化且易操作的介面使用 AI 模型協助診斷。

五、 團隊合作方式

本次專題由我獨立完成。專題初期每兩周與教授開會一次，確認研究方向與進度狀況，根據教授的建議、文獻搜尋與結果分析來訂定後續的工作項目並將其實作。模型研究部分告一段落後，再自行進行網頁應用的架設。

六、 設計原理、研究方法與步驟

1. 主要模型架構

專題中使用了兩種 CNN 模型架構，分別為 UNet 以及 Double-UNet(如表 1)，以下分別對其選擇做詳細的說明。

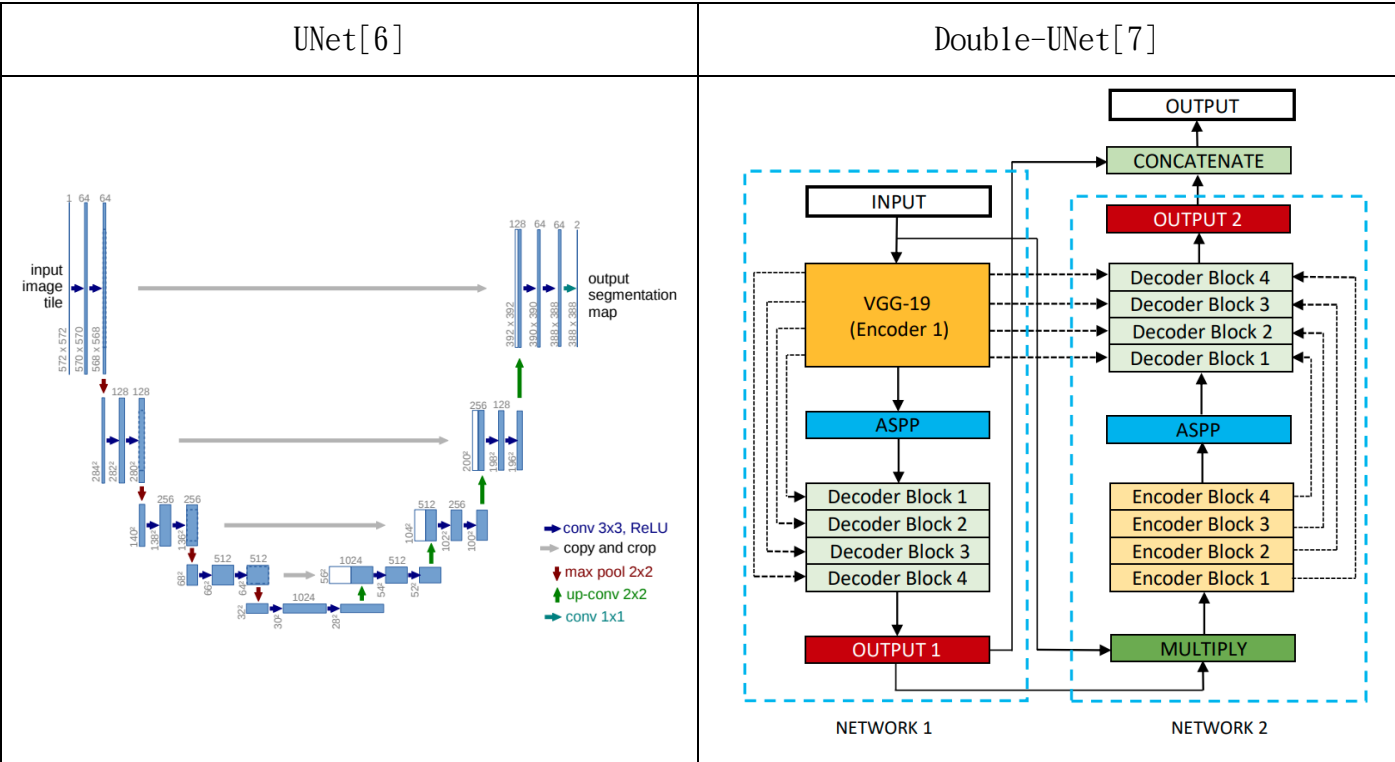


表 1. 模型架構圖

1.1 UNet:

UNet 在醫療影像分割上有很好的表現，由於其架構會將不同層的 feature map 經過 up-sample 後連接起來，可以有效的對不同尺度的特徵進行預測。而在息肉影像資料集中，不同的拍攝角度、息肉的生長程度等等，都會造成特徵有些許的不同，因此

選用 UNet 作為主要模型架構，以針對這些特徵學習。另外，UNet 對於物體的邊緣也能有效地預測，此特性對框選出息肉有很好的幫助。

1.2 Double-UNet:

Double-UNet 為 UNet 模型的延伸版本，結合了兩個 UNet，將第一個 UNet 的輸出與原輸入圖片相乘後，輸入至第二個 UNet，對第一個模型的預測結果再次精煉，以獲得更準確的特徵，並連結兩個 UNet 的 Encoder 以及自身 Decoder 的 feature map，得到更豐富的資訊，以改進單一 UNet 的預測結果。使用 Double UNet 架構會需要更多的訓練時間，但模型的預測效果也會有所進步，為了能更準確地預測出息肉的細節位置，因此後續也採用了這個模型。

2. 資料增強

由於訓練時只採用資料集中 900 張圖片，為了豐富訓練資料以增進模型訓練效果，採用了以下方法進行資料增強。

2.1 基本圖像轉換

首先，將圖片進行隨機的角度旋轉，並隨機進行水平或垂直翻轉，再透過放大、縮小等方式模擬出鏡頭從不同角度、距離拍攝的圖片，並對圖片顏色做些微的隨機震盪，以及不同程度的高斯模糊，藉此營造出不同光線及拍攝狀態的圖片，將原本的 900 張訓練圖片透過資料增強擴增到 2700 張。



圖 2. 基本圖像轉換增強範例

2.2 息肉圖像剪貼

由於訓練模型時，發覺模型對於一張圖片中有多個息肉位置時預測的效果較差，因此設計了以下圖像增強流程：第一步，隨機選取一張訓練圖片，並找出其不含息肉的位置。第二步，隨機選取另一張訓練圖片，將其息肉部分裁切下來。第三步，將第二步裁切下的息肉圖片黏貼到第一步的位置。以此流程增強資料集，使模型能針對單張圖片多個息肉去做訓練，將資料及數量再從 2700 擴增到 3000 張。



圖 3. 息肉圖像剪貼範例

3. Encoder 預訓練方式

本章節的預訓練主要針對 UNet 的 Encoder 進行改良。在專題期間，一共嘗試了三種模型預訓練方法，首先是偵測圖片旋轉角度的 self-supervise learning，但此預訓練並沒有顯著的效果。之後嘗試了 Momentum contrastive learning[8]，也因為資料集規模不大而無法有效幫助模型訓練。最終，嘗試了在 UNet 的 Encoder 部分加入一層 fully-connected layer，並將每張圖片裁切成含有息肉的部份以及不含息肉的部分，輸入 Encoder 做 Classification 分辨圖片中是否包含息肉的方式，使模型 Encoder 能更有效地分辨息肉的特徵。此方法得到了不錯的效果，可針對圖片中容易被搞混的部分讓模型學習，因此選定此方式預訓練模型。

4. 模型訓練

完成 Encoder 的預訓練以及圖像增強後，即可將資料輸入模型進行預測，並由預測結果與 Ground Truth 比較計算 loss 來讓模型學習。Loss Function 選用 IOU Loss，針對預測出有腫瘤的部份與 Ground Truth 中腫瘤的實際部分重疊度去計算，避免學習到太多背景的資訊。訓練完成並確認預測效果後，即可得到最終的模型。

$$IOULoss = 1 - \frac{TP}{TP + FN + FP}$$

5. 網頁應用

完成模型訓練後，將訓練好的模型串接至網路應用。首先，在 server 端載入預測模型，而網頁端有兩個介面，分別提供使用者輸入欲進行預測的圖片或影片。使用者輸入的檔案會被上傳到 server 端的模型內進行預測。若輸入檔案為圖片，輸入的圖片與預測結果會被顯示在同一列以做比對；若輸入檔案為影片，則會將預測結果與原影片結合後顯示在影片窗格中，使用者可以從輸出影片看到即時的預測結果。

七、 系統實現與實驗

本章將分為模型訓練與網頁應用兩部分，針對前者說明實驗上設定的細節以及整體流程；後者則專注於系統實現的部分。

7.1 模型訓練

圖 4 為模型訓練的完整流程圖。在本次專題中，使用 PyTorch 實現模型的建立與訓練，並分別對 UNet、Double-UNet 以下幾種訓練設定：

模型	訓練設定
UNet	原始資料集(900 張訓練影像)
	增強資料集(900+2100 張訓練影像)
	預訓練 Encoder+原始資料集
	預訓練 Encoder+增強資料集
Double-UNet	原始資料集(900 張訓練影像)
	增強資料集(900+2100 張訓練影像)

Optimizer 選用 SGD，超參數的設定為 Batch Size:8, Learning Rate: 1e-2，並訓練 60 個 epoch。

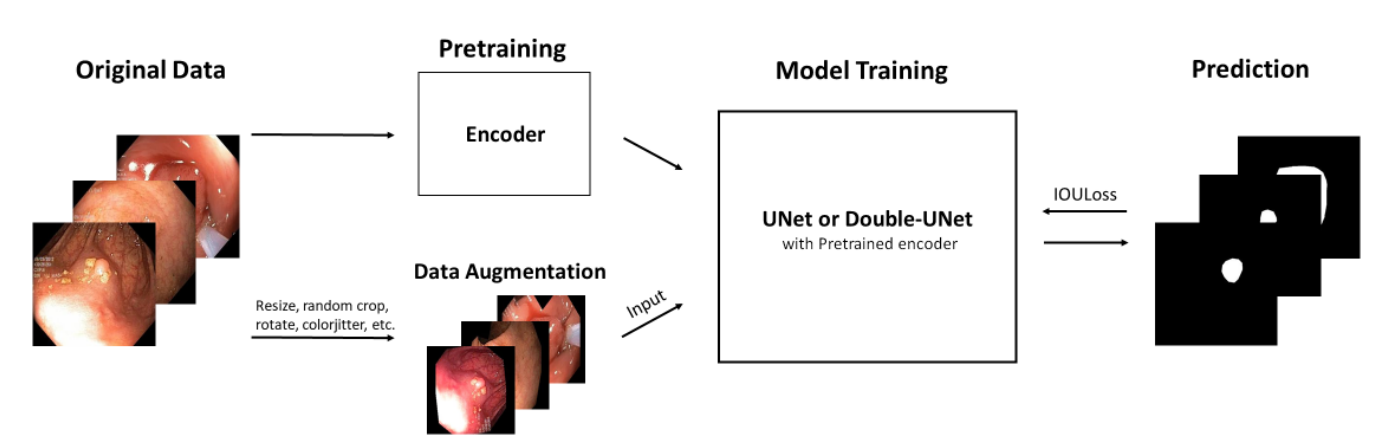
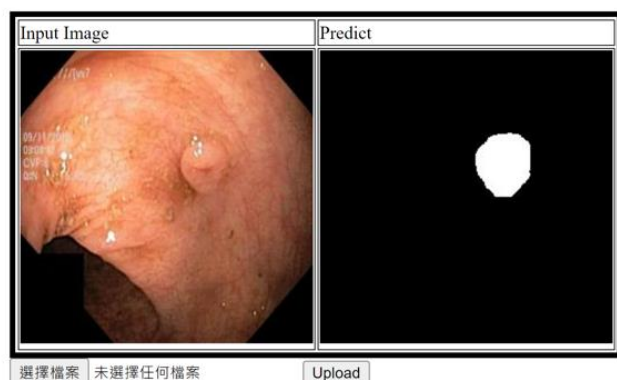


圖 4. 模型訓練流程圖

7.2 網頁應用

在本次專題中，運用 Python 模組 Flask 進行網頁應用的架設，後端會載入訓練好的模型，針對使用者輸入的圖片可直接預測，模型會將預測結果輸出至 server 端的資料夾，再透過網頁顯示結果。若是輸入影片，則會利用套件 opencv 將其拆分為多張圖片，並儲存影片的偵率，將圖片各自進行預測後，再重新與原圖結合為影片並顯示於前端，供使用者查看。透過網頁進行預測的範例見圖 5。

Upload new File



Upload new Video



圖 5. 網頁預測圖片、影片範例

八、 效能評估與成果

從實驗結果(表 2)看來，本次專題所使用的資料增強策略，無論對於哪種模型的訓練效果都有顯著的幫助。應用在 UNet 時，IOU 上升了 0.05，在 Double-UNet 的 IOU 也有 0.032 的提升。在原始資料量較少的情況下，選用這個策略是一個很好的選擇。而 UNet 的 Encoder 預訓練也同樣使其表現有所提升，同時使用預訓練 Encoder 與資料增強更能使表現再往上提升一些，雖然提升的幅度沒有如資料增強那麼明顯，但此種方法預訓練所耗費的時間很短，幾乎不影響訓練上的時間成本。表格中紅字部分為各模型之最好表現，由於是針對息肉位置的偵測，因此以 IOU 為主要評判標準。模型種類的比較上，Double-UNet 由於其模型複雜度較高，預測效果優於 UNet。UNet 經過 Encoder 預訓練與資料增強後，表現雖仍不及 Double-UNet，但在計算資源較有限，無法負荷較複雜的 Double-UNet 時，也不失為一個好的辦法。

Model & Training Method	Accuracy	IOU	DICE
UNet 原始資料集	0.928626	0.592346	0.743991
UNet 增強資料集	0.940864	0.648868	0.787047
UNet 原始資料集 Pretrained Encoder	0.928806	0.604757	0.753705
UNet 增強資料集 Pretrained Encoder	0.937552	0.653285	0.790288
Double-UNet 原始資料集	0.944115	0.677014	0.807404
Double-UNet 增強資料集	0.950565	0.709908	0.830346

表 2. 模型預測結果

九、 結論

在本次專題中，試驗了能有效提升模型表現的資料增強與模型預訓練方法，並對不同模型進行實驗，以驗證此種方法的通用性。未來我將持續進行更多深入的研究，以精進模型的表現，找出更多可以幫助醫療影像 AI 模型訓練的方法。網頁應用架設

方面，讓使用者可以用更方便的方式進行預測，但礙於運算資源的限制，影片的預測速度會較為緩慢。未來也將試著使應用介面更加完善，並嘗試雲端運算等方法來解決運算速度的問題。經過這次的經驗，我對於醫療影像 AI 有了更多的認識，也看到了許多自己在 AI 領域可以持續努力學習的方向，使我有許多成長。

十、參考文獻

- [1] Debesh Jha, Pia H. Smedsrud, Michael A. Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, Håvard D. Johansen. "Kvasir-SEG: A Segmented Polyp Dataset". In: arXiv:1911.07069.
- [2] Bhoj Raj Pandit, Abeer Alsadoon, P.W.C. Prasad, Sarmad Al Aloussi, Tarik A. Rashid, Omar Hisham Alsadoon, Oday D. Jerew. "Deep Learning Neural Network for Lung Cancer Classification: Enhanced Optimization Function". In: arXiv:2208.06353.
- [3] Rodina Bassiouny, Adel Mohamed, Karthi Umapathy, Naimul Khan. "An Interpretable Object Detection-Based Model for the Diagnosis of Neonatal Lung Diseases using Ultrasound Images". In: arXiv:2105.10081.
- [4] Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen and Ling Shao. "PraNet: Parallel Reverse Attention Network for Polyp Segmentation". In: arXiv:2006.11392v4.
- [5] Bo Dong, Wenhai Wang, Deng-Ping Fan, Jinpeng Li, Huazhu Fu, Ling Shao. "Polyp-PVT: Polyp Segmentation with Pyramid Vision Transformers". In: arXiv:2108.06932v5.
- [6] Olaf Ronneberger, Philipp Fischer, Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: arXiv:1505.04597.
- [7] Debesh Jha, Michael A. Riegler, Dag Johansen, Pål Halvorsen, Håvard D. Johansen. "DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation". In: arXiv:2006.04868.
- [8] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, Ross Girshick. "Momentum Contrast for Unsupervised Visual Representation Learning". In: arXiv:1911.05722