



國立中山大學資訊工程學系

碩士論文

Department of Computer Science and Engineering

National Sun Yat-sen University

Master's Thesis

基於深度學習之腫瘤與其器官分割

Organ with Tumor Segmentation Based on Deep Learning

研究生：林昱成

Yu-Cheng Lin

指導教授：蔣依吾 博士

Dr. John Y. Chiang

中華民國 113 年 7 月

July 2024

中文摘要

卵巢是女性之重要性腺器官，主要功能為排卵和分泌女性激素。卵巢腫瘤是女性生殖器常見腫瘤，因卵巢位於骨盆腔深部，腫瘤不容易及早發現，卵巢腫瘤型態可分為：子宮內膜異位囊腫、漿液性卵巢囊腫、黏液性卵巢囊腫、惡性卵巢腫瘤等。

現今卵巢癌檢查方式有：陰道超音波檢查、電腦斷層、核磁共振、抽血檢查(CA-125)等。依據卵巢癌病灶大小以及是否轉移至其他部位，判斷病患嚴重程度，醫師再根據病情預測手術後腫瘤完全清除機率決定是否進行手術。

由於每位病患個體差異以及腫瘤增長程度不同，以及不同醫師根據自身經驗判斷結果不一致，導致判定結果並沒有統一標準。本文使用病患之電腦斷層(CT)影像作為訓練資源，所收集 CT 影像包含 107 位病患，影像張數共 7294 張，並將所有影像資料以病患為單位，以 9：1 方式切分為訓練集與測試集，再從訓練集選取 20%作為驗證集，訓練集包含 77 組病患，共 5247 張影像，驗證集包含 19 組病患，共 1295 張影像，測試集包含 11 組病患，共 4512 張影像。

本研究使用深度學習模型訓練 CT 影像，目的為將病患 CT 影像進行器官部位分割，提高各器官部位之腫瘤辨識，並根據模型所預測之分割部位輸入下一階段模型以偵測對應腫瘤。其中肝臟部位準確度約 0.82；淋巴結部位準確度約 0.65；骨盆腔部位準確度約 0.61。透過先預測部位位置，再進行腫瘤分割預測，提升肝臟腫瘤、淋巴結腫瘤以及卵巢腫瘤之分割結果，其中肝臟腫瘤分割準確度由 0.71 提升至 0.78；淋巴結腫瘤分割準確度由 0.52 提升至 0.61；卵巢腫瘤分割準確度由 0.79 提升至 0.84。模型預測結果可輔助醫師進行臨床判斷，作為腫瘤清除手術執行與否之參考。

英文摘要

The ovary is a crucial gonadal organ in women, primarily responsible for ovulation and the secretion of female hormones. Ovarian tumors are common tumors within the female reproductive system. Due to the ovaries being located deep within the pelvic cavity, these tumors are often difficult to detect early. Ovarian tumors can be classified into different types, including endometriomas, serous ovarian cysts, mucinous ovarian cysts, and malignant ovarian tumors.

Current methods for diagnosing ovarian cancer include transvaginal ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), and blood tests (CA-125). The severity of the condition is assessed based on the size of the ovarian cancer lesion and whether the tumor has metastasized to other parts of the body. Physicians then predict the likelihood of complete tumor removal post-surgery to decide whether to proceed with surgery.

Due to individual differences among patients, varying tumor growth rates, and differences in physicians' judgments based on their experience, there is no standardized method for assessment. This study uses CT images from patients as training data. A total of 7,294 images from 107 patients were collected and divided into training and testing sets in a 9:1 ratio, with 20% of the training set further used as a validation set. The training set consists of 77 patients with 5,247 images, the validation set includes 19 patients with 1,295 images, and the testing set includes 11 patients with 4,512 images.

In this study, a deep learning model was trained using CT images with the aim of segmenting organ regions in the images to improve tumor identification within these organs. The segmented regions predicted by the model were then input into the next stage of the model to detect the corresponding tumors. The accuracy rates achieved

were approximately 0.82 for the liver region, 0.65 for the lymph node region, and 0.61 for the pelvic region. By predicting the locations of the regions first and then performing tumor segmentation, the segmentation accuracy for liver tumors increased from 0.71 to 0.78, for lymph node tumors from 0.52 to 0.61, and for ovarian tumors from 0.79 to 0.84. The model's predictions can assist physicians in making clinical decisions regarding the necessity of tumor removal surgery.

目錄

論文審定書.....	錯誤! 尚未定義書籤。
中文摘要.....	i
英文摘要.....	ii
目錄.....	iv
圖次.....	vii
表次.....	ix
1 第一章 緒論	1
1.1 背景.....	1
1.2 檢查方式.....	2
1.2.1 超音波檢查.....	2
1.2.2 電腦斷層掃描.....	3
1.2.3 核磁共振掃描.....	4
1.2.4 抽血(CA-125 腫瘤指數).....	5
1.3 電腦斷層掃描影像.....	5
1.3.1 卵巢腫瘤(原位癌).....	5
1.3.2 肝臟腫瘤(擴散癌).....	6
1.3.3 淋巴結腫瘤(擴散癌).....	7
1.3.4 網膜腫瘤(擴散癌).....	7
1.3.5 腹膜腫瘤(擴散癌).....	8
1.3.6 腫瘤播種(擴散癌).....	9
1.4 目的.....	10
2 第二章 相關研究	11
2.1 深度學習.....	11
2.2 U-Net 神經網路.....	12
2.3 Attention(注意力)架構.....	12
2.4 卷積神經網路.....	13

2.4.1 轉置卷積.....	14
2.4.2 空洞卷積.....	15
2.5 殘差神經網絡.....	15
2.6 Transformer 架構.....	16
2.7 Swin U-Net 神經網路	17
2.7.1 Swin Transformer Block	18
3 第三章 研究方法	20
3.1 資料處理.....	20
3.2 資料集.....	21
3.2.1 資料增量.....	21
3.2.2 資料增量方法.....	21
3.2.3 資料分配.....	22
3.2.4 資料前處理.....	23
3.3 深度模型訓練.....	24
3.4 3D U-Net.....	24
3.5 2D U-Net.....	24
3.5.1 2D U-Net 模型骨架.....	25
3.5.2 注意力機制.....	26
3.6 Swin U-Net	26
3.6.1 Transformer block.....	27
3.7 模型評估.....	27
4 第四章 研究結果	28
4.1 模型訓練.....	28
4.1.1 3D 模型.....	28
4.1.2 2D 模型.....	29
4.2 影像訓練結果.....	32

4.3 腫瘤預測結果比較.....	36
5 第五章 結論	37
參考文獻.....	38

圖次

圖 1-1 超音波示意圖	2
圖 1-2 斷層掃描示意圖	3
圖 1-3 核磁共振示意圖	4
圖 1-4 卵巢腫瘤示意圖，左邊為原圖，右邊標記部分為卵巢腫瘤	6
圖 1-5 肝臟腫瘤示意圖，左邊為原圖，右邊標記部分為肝臟腫瘤	6
圖 1-6 淋巴結腫瘤示意圖，左邊為原圖，右邊標記部分為淋巴結腫瘤	7
圖 1-7 網膜腫瘤示意圖，左邊為原圖，右邊標記部分為網膜腫瘤	8
圖 1-8 腹膜腫瘤示意圖，左邊為原圖，右邊標記部分為腹膜腫瘤	8
圖 1-9 腫瘤播種示意圖，左邊為原圖，右邊標記部分為腫瘤播種	9
圖 2-1 深度學習神經網路示意圖	11
圖 2-2 U-Net 結構圖	12
圖 2-3 Attention gate 架構示意圖	13
圖 2-4 卷積示意圖	14
圖 2-5 轉置卷積示意圖	14
圖 2-6 空洞卷積示意圖	15
圖 2-7 殘差學習示意圖	16
圖 2-8 transformer 架構示意圖	17
圖 2-9 Swin UNet 架構示意圖	18
圖 2-10 Shifted window 示意圖	19
圖 3-1 單階段實驗流程	20
圖 3-2 資料增量示意圖	22
圖 3-3 資料分配示意圖	23
圖 3-4 U-Net+ResNet 示意圖	25
圖 3-5 U-Net +注意力機制示意圖	26

圖 3-6 MIoU 計算公式	27
圖 3-7 IoU 示意圖	27
圖 4-1 3D 模型訓練結果圖	28
圖 4-2 2D 模型訓練結果圖	30
圖 4-3 器官預測結果	33
圖 4-4 器官預測遮罩	34

表次

表 3-1 各器官部位標記數量	21
表 3-2 各資料集不同器官部位數量統計	23
表 4-1 MIoU 評估結果	29
表 4-2 各類器官部位 MIoU	29
表 4-3 2D 模型訓練結果比較	30
表 4-4 MIoU 評估結果	31
表 4-5 各類器官部位 MIoU	31
表 4-6 較優預測結果	35
表 4-7 較差預測結果	36
表 4-8 模型辨識 MIoU 比較	36

第一章 緒論

1.1 背景

癌症又名為惡性腫瘤，指細胞不正常增生，且增生細胞可能侵犯身體其他部分，由於控制細胞分裂增殖機制失常而引發此疾病。癌細胞除分裂失控外，還會局部侵入週遭正常組織甚至經由體內循環系統或淋巴系統轉移至身體其他部分。女性生殖系統包括內生殖器之卵巢、輸卵管、子宮、陰道，以及外生殖器，其中最易造成癌症之部位為子宮頸、子宮及卵巢。綜觀民眾癌症發生及死亡情形，呈現逐年上升趨勢。

一般卵巢癌[1]死亡率佔婦科癌症中極高比例，其原因為大部份卵巢癌在早期並無明顯症狀，因此當確診卵巢癌時，通常已擴散至腹腔中其他器官，成為晚期癌症，超過半數患者已達到第三期或第四期；此時病徵多為腹部腫大，會造成患者食慾不佳，脹氣等腹部症狀。若腫瘤已壓迫其他器官，常造成其他如大便習慣改變、頻尿等症狀。當腫瘤過大，病人甚至可以自行觸摸感覺腫塊，外觀則呈現明顯腹部隆起，此時常合併腹水發生，嚴重時會導致呼吸困難、疼痛。

卵巢癌治療主要分為手術治療和化學治療，手術通常為卵巢癌首選療法。手術可能涉及切除一側或雙側受影響卵巢，並且在癌症擴散至輸卵管、子宮、鄰近淋巴結和脂肪組織時，也需將其切除。手術有時可以單獨消除卵巢癌，但通常與化療結合使用[2]以減少卵巢癌復發機率。化療可以在手術之前進行，也能在手術之後進行，治療順序和組合取決於卵巢癌分期以及其是否已擴散至身體其他部位。

另一種治療方法稱為熱灌注腹腔化療（HIPEC）[3]，此療法在手術期間進行，通常在卵巢癌晚期或初期被認為不可手術時推薦使用。溫熱化療藥物泵入腹腔，集中在癌症所在位置，從而避免損害健康組織，減少化療相關副作用。

1.2 檢查方式

1.2.1 超音波檢查



圖 1-1 超音波示意圖

超音波為醫學影像常用檢查方式之一，圖 1-1 為超音波檢查之影像。超音波已廣泛運用於診斷晚期卵巢癌，但利用其診斷早期卵巢癌仍然是目前瓶頸，因為早期卵巢癌外形可類似於正常大小、正常外觀之卵巢，會讓早期卵巢癌在單用傳統超音波診斷時相當困難。隨著都卜勒超音波技術進步，改善傳統超音波影像在分辨良性腫瘤與惡性腫瘤能力，包括腫瘤內部結構解析度提高和異常血管結構評估。較常運用之參數為血流阻力指數（Resistance index: RI），當卵巢腫瘤血流 $RI < 0.4$ 時，惡性腫瘤機率增加。但都卜勒超音波對正常大小卵巢之卵巢癌患者診斷仍然有極限。

1.2.2 電腦斷層掃描

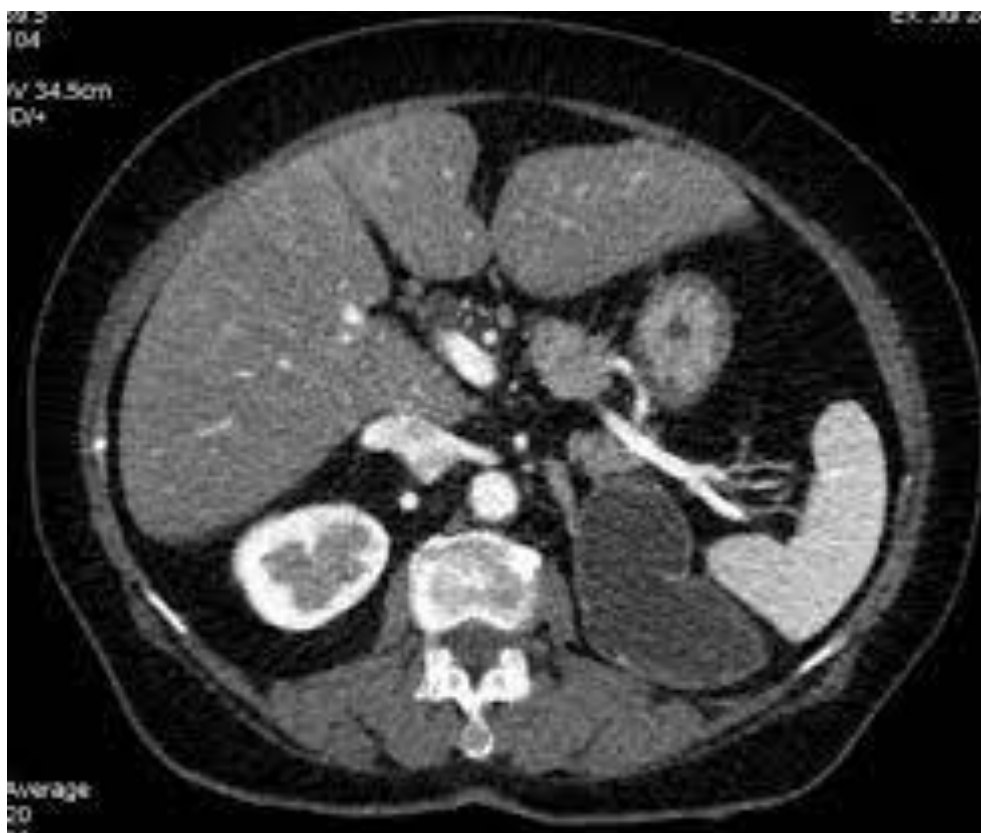


圖 1-2 斷層掃描示意圖

電腦斷層利用 X 光之光束穿透身體，在身體對側以 X 光偵測器接收穿透過身體之 X 光，再以電腦解讀接收器而顯像，最後應用電腦系統計算，重建出該橫截面之組織特性影像，如圖 1-2 所示。由於人體各組織對 X 光吸收能力不同，所表現出特徵也不同。若器官出現異常組織增生（腫瘤），它與正常細胞吸收 X 光能力也會不同，顯像因而產生差異，可藉此判讀身體異狀。在進行掃描之前需要接受靜脈注射顯影劑，以獲得更清晰影像。其優點為速度快、無痛、價格較低，且可應用於全身；缺點為含有微量輻射，對孕婦並不適合。

1.2.3 核磁共振掃描

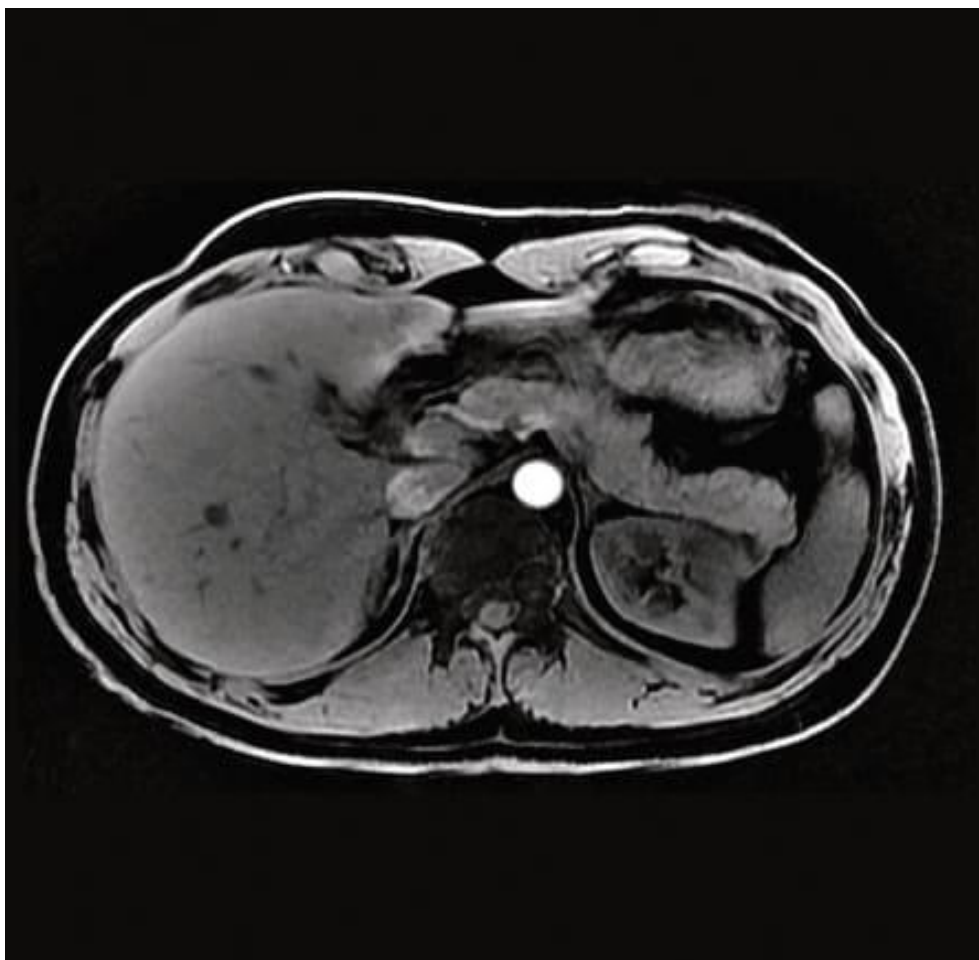


圖 1-3 核磁共振示意圖

此方法利用核磁共振原理，依據所釋放能量在物質內部不同結構環境中不停衰減，通過外加梯度磁場檢測所發射電磁波，即可得知構成此物體原子核之位置和種類，據此可以繪製成物體內部之結構圖像，圖 1-3 為核磁共振掃描之影像。核磁共振掃描優點為對軟組織具極高分辨力，對膀胱、直腸、子宮、陰道、骨、關節、肌肉等部位檢查效果比 CT 優勝；通過調節磁場可自由選擇所需剖面，能得到其它成像技術所不能接近或難以接近部位之圖像，並且不會產生輻射。其缺點為耗時且空間分辨力不足，檢查所需花費也較高。

1.2.4 抽血(CA-125 腫瘤指數)

目前 CA-125 正常值訂在 <35 U/ml，在健康成人中僅 1%大於 35 U/ml。CA-125 為檢查婦女生殖器官之腫瘤標記，當發生卵巢上皮細胞腫瘤、子宮內膜腺瘤、子宮內膜異位、腸胃道癌症等，其 CA-125 指數可能偏高。在現階段，CA-125 可做為卵巢癌治療後追蹤指標，但不適合大規模使用於一般婦女作為卵巢癌之篩檢工具，主要原因為其對早期卵巢癌精確預測率（Positive predictive rate）僅 10%，假陽性及假陰性比率高，故單用 CA-125 並非理想卵巢癌篩檢方式。通常 CA-125 會搭配其他檢查方式(超音波檢查)共同評估。

1.3 電腦斷層掃描影像

1.3.1 卵巢腫瘤(原位癌)

依組織病理分類，卵巢癌可分為：上皮細胞癌、生殖細胞癌以及性腺細胞特定間質細胞癌，如圖 1-4 中標記部分所示。卵巢癌轉移途徑經由腹膜內及血液淋巴轉移，並擴散至大腸、小腸、腹膜、大網膜、脾臟、膀胱、橫隔膜、淋巴等，遠處轉移包括肝、肺、骨骼等部位。

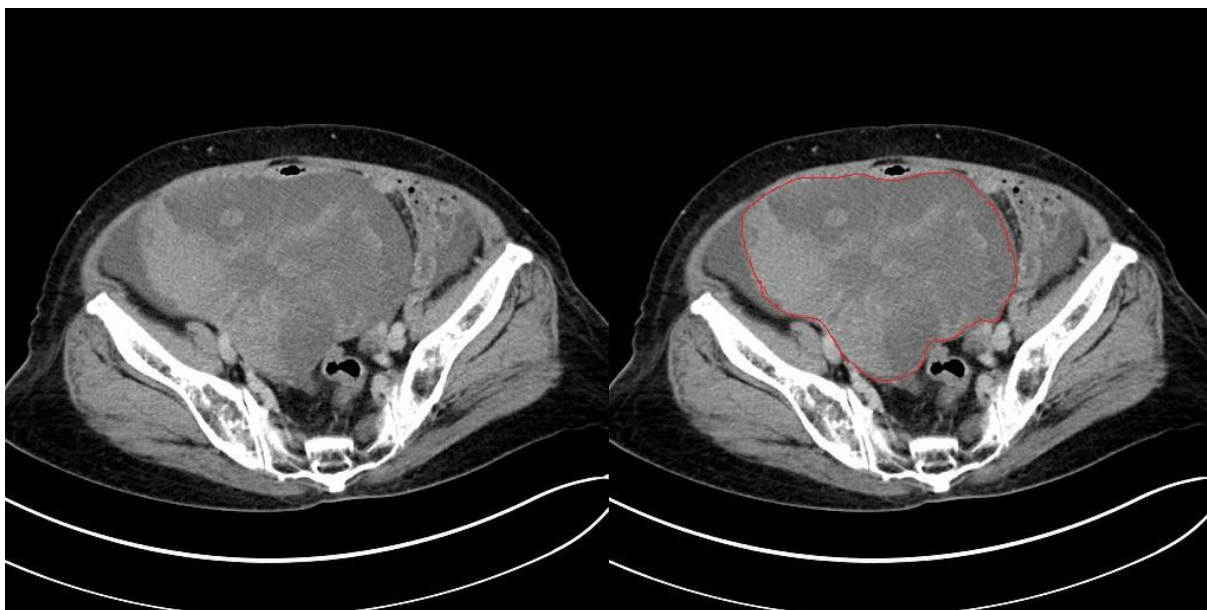


圖 1-4 卵巢腫瘤示意圖，左邊為原圖，右邊標記部分為卵巢腫瘤

1.3.2 肝臟腫瘤(擴散癌)

肝臟腫瘤由卵巢癌轉移至肝臟，如圖 1-5 中標記部分所示。健康肝臟 CT 會呈現較淺顏色，肝臟轉移之患者其肝臟部位會呈現較深區塊，其大小、數量不一。

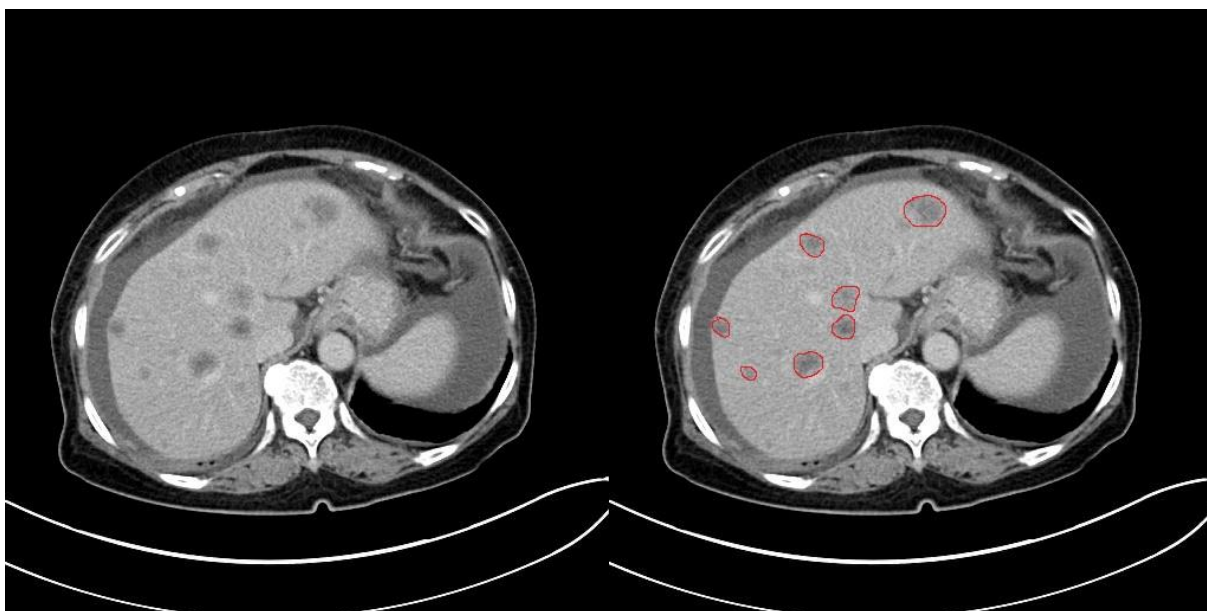


圖 1-5 肝臟腫瘤示意圖，左邊為原圖，右邊標記部分為肝臟腫瘤

1.3.3 淋巴結腫瘤(擴散癌)

淋巴結腫瘤由卵巢癌經由淋巴系統轉移至身體各處淋巴結，如圖 1-6 標記部分所示。其特色是沿血管附近長出球狀腫瘤，血管呈現較淺色，淋巴結轉移則是在血管周圍出現較深色類球狀顆粒。

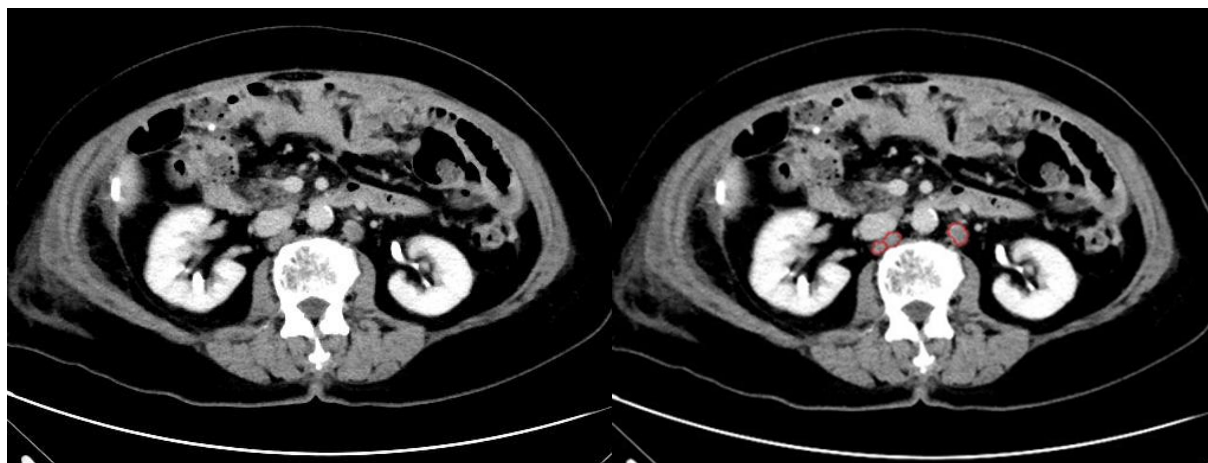


圖 1-6 淋巴結腫瘤示意圖，左邊為原圖，右邊標記部分為淋巴結腫瘤

1.3.4 網膜腫瘤(擴散癌)

網膜由雙層腹膜構成，兩層之間包含血管、神經、淋巴管和結締組織等通過，包括大網膜、小網膜及網膜囊。網膜腫瘤由卵巢癌擴散至腹部，當網膜轉移發生時，會如上圖出現不規則亮暗區域，如圖 1-7 標記部分所示。

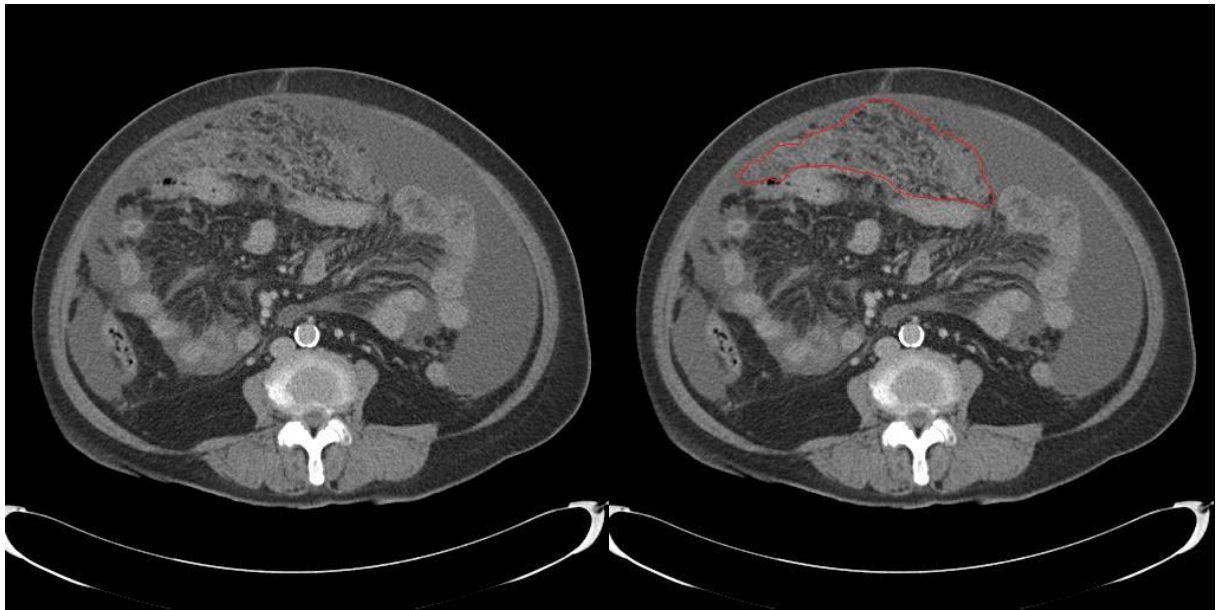


圖 1-7 網膜腫瘤示意圖，左邊為原圖，右邊標記部分為網膜腫瘤

1.3.5 腹膜腫瘤(擴散癌)

腹膜覆蓋於腹壁、盆壁內面和腹腔、盆腔各臟器表面之漿膜，如圖 1-8 標記部分所示。內襯於腹壁、骨盆壁和橫膈下表面腹膜稱壁層腹膜，覆蓋於臟器表面之腹膜稱為臟層腹膜，之間潛在之空間隙稱為腹膜腔，具支援、固定、保護、防禦等功能。腹膜腫瘤由卵巢癌擴散至腹部，當腹膜轉移發生時，邊緣會呈現不平整顆粒狀。

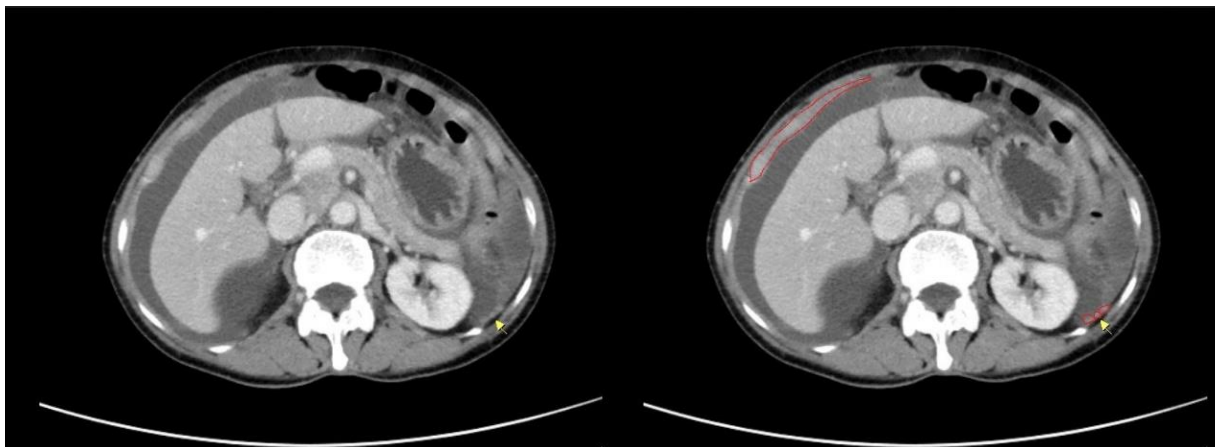


圖 1-8 腹膜腫瘤示意圖，左邊為原圖，右邊標記部分為腹膜腫瘤

1.3.6 腫瘤播種(擴散癌)

當癌細胞侵犯至漿膜外時，可以脫落至腹腔內其他器官表面，引起腹腔腫瘤播種，如圖 1-9 標記部分所示。腫瘤播種為複雜生物過程，好發部位包含大網膜、腸系膜、膀胱直腸陷凹、子宮直腸陷凹等。腫瘤播種由卵巢癌擴散至腹部，主要出現在腸子周圍，其形狀、大小不固定，CT 圖呈現顏色與腸子相近，人眼不容易辨別。

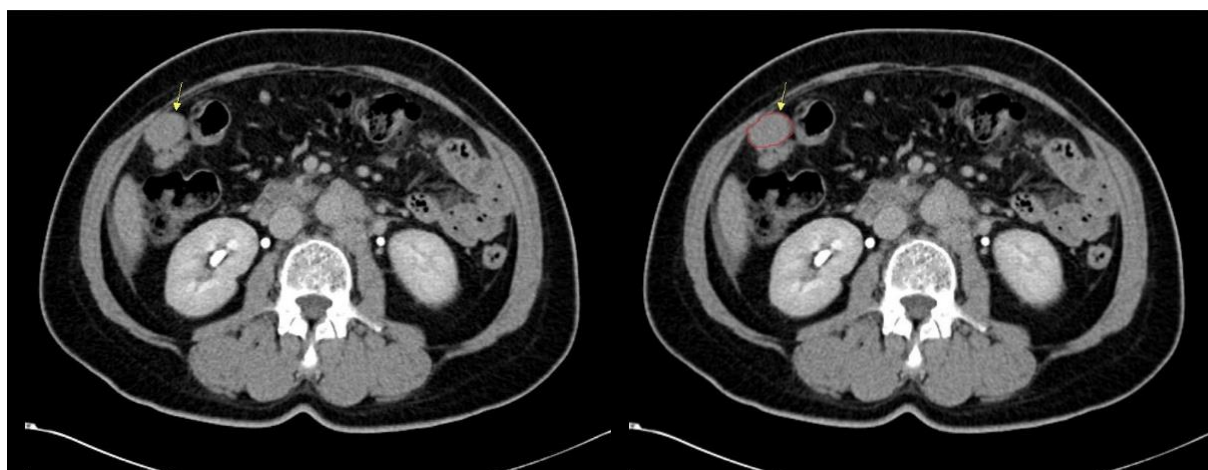


圖 1-9 腫瘤播種示意圖，左邊為原圖，右邊標記部分為腫瘤播種

1.4 目的

電腦斷層影像中各部位腫瘤為醫師判斷治療方法之重要資訊。透過觀察原位癌形狀大小以及增長位置，並檢查其是否有病灶轉移，決定手術執行與否以及手術執行方式。目前有多種檢測卵巢癌方式，每種檢測方法各有優缺點。使用超音波檢測雖然最快速且花費較低，但其成像卻不如其他方法精細；使用核磁共振能獲得高精度影像資料，但成本高昂且檢測時間長；電腦斷層產出影像精度雖不比核磁共振，但已經比超音波精確，檢測成本也低於核磁共振。

不論哪種檢測方法皆需要醫師做最後診斷，目前並無統一診斷準則，因此本研究使用由奇美醫院提供之電腦斷層影像，透過深度學習對影像進行器官分割，並進一步從各器官預測其對應腫瘤，輔助醫師擷取肉眼難以分辨之病灶，再根據模型預測結果協助醫師診斷病患病情，減少主觀判斷誤差，幫助醫師更精準執行手術決策以及治療方案。

原先模型為單階段模型，透過已標記之 CT 影像輸入模型訓練[4,5]，再由訓練完成之模型預測影像中腫瘤種類、位置以及範圍。為提升腫瘤分割準確度，將研究拆分為二階段模型，第一階段分割腫瘤與其好發位置，再將分割結果輸入至第二階段模型，並針對第一階段所預測之分割位置預測其中是否包含腫瘤。本研究為第一階段模型，由於其他擴散腫瘤好發位置、大小等並不固定，因此本研究將針對肝臟部位、淋巴結部位、骨盆腔部位進行分割，透過模型訓練並預測上述部位，將預測結果輸入下一階段模型。

第二章 相關研究

2.1 深度學習

深度學習[6,7,8]為一種以人工神經網路為架構，對資料進行表徵學習之演算法，如圖 2-1 所示。深度學習為多層之神經網路，神經網路會透過節點間相互連接層傳遞資料。當資訊通過圖層時，該圖層中所有節點會對資料執行簡單作業，並選擇性地將結果傳遞至其他節點。其後續圖層皆聚焦於較高層級之功能，直至網路建立輸出為止。隨著硬體發展，其運算力大幅提升，許多以前無法實行複雜運算皆能透過深度學習便能得到結果，深度學習應用也越來越廣，如電腦視覺、語音識別、訊號處理等。深度學習發展促使人工智慧更強大，能更精準完成複雜任務，甚至比人類更強。

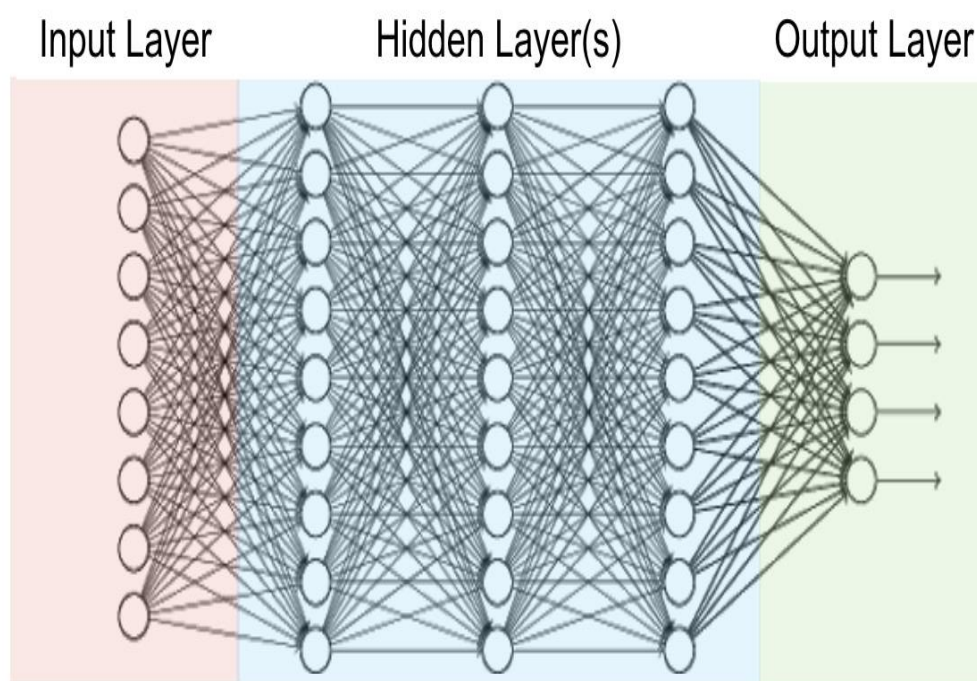


圖 2-1 深度學習神經網路示意圖

2.2 U-Net 神經網路

U-Net[9]有別於傳統神經網路，其採用編碼-解碼之 U 行架構，如圖 2-2 所示。U-Net 特色是使用少量資料便能進行預測。結構圖中左邊降維路線(編碼)稱為 contracting path，右邊(解碼)則為 expanding path。在進入降維或升維前會經過連續兩層卷積，稱為 successive convolution，使神經網路更容易捕捉精確訊息。另一特色為維持 upsampling 之高 channel 數目，使相對位置關係和細部特徵可以充分結合，增加辨識品質。其次，神經網路可以自行以鏡像補充圖片，如此能增加對於在圖片邊緣物件之辨識度，且不占用 GPU 消耗。U-Net 大量應用於生物醫學圖像分割中，如腦圖像分割、肝圖像分割和蛋白質結合點預測等。

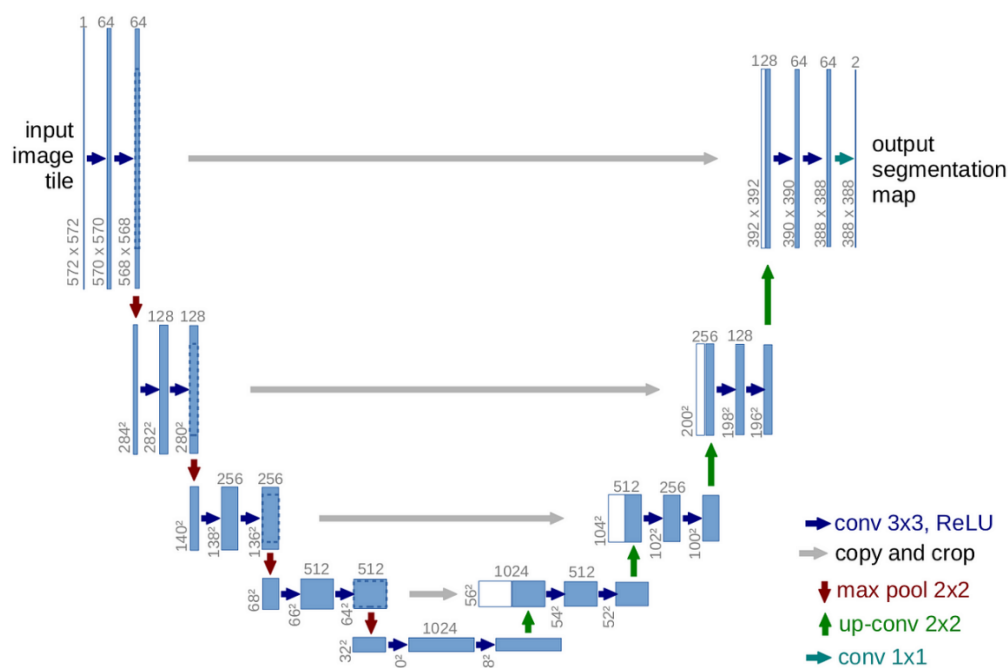


圖 2-2 U-Net 結構圖

2.3 Attention(注意力)架構

注意力機制[10,11]為人工神經網路中一種模仿認知注意力技術，如圖 2-3 所示。此

機制可以增強神經網路輸入資料中某些部分權重，同時減弱其他部分權重，以此將網路關注點聚焦於資料中最重要一小部分。資料中哪些部分比其他部分更重要取決於上下文，可以透過梯度下降法對注意力機制進行訓練。

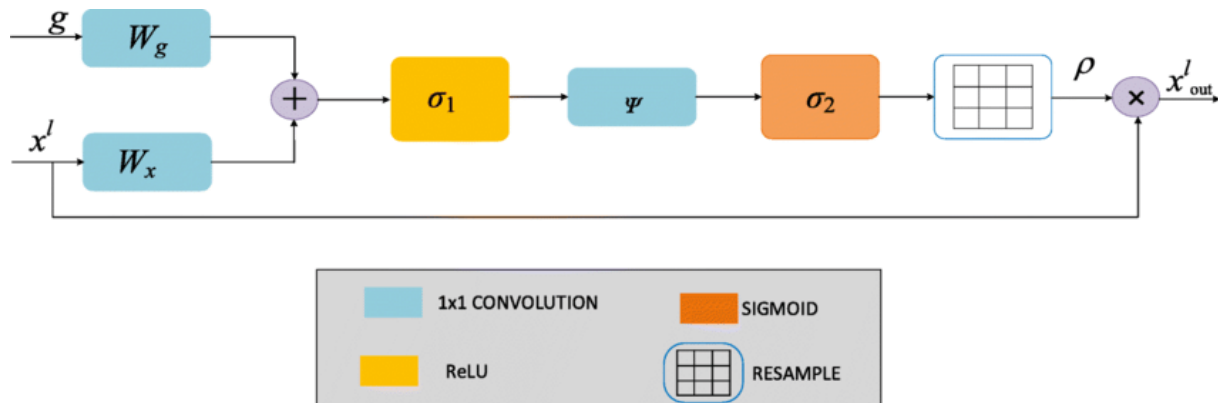


圖 2-3 Attention gate 架構示意圖

在原有 U-Net 模型架構基礎上加入注意力機制[12]可顯著提升模型性能。注意力機制增強特徵選擇能力，使模型動態加權重要特徵，削弱無關特徵，提升細節識別準確性，尤其在物體邊界區域能有更高辨識度。通過過濾冗餘信息，減少噪聲干擾，增強多尺度信息融合，助於處理多尺度物體，提升分割準確性，並使模型在面對各種輸入變化時保持高性能和穩定性。

2.4 卷積神經網路

卷積（Convolution）[13,14,15]是一種在數學和信號處理領域廣泛應用之運算方法，如圖 2-4 所示。卷積主要用於圖像處理和機器學習卷積神經網路中。其基本思想是通過特定運算方法組合兩個函數，產生新函數。在離散形式中，卷積可以定義為兩個離散函數加權求和。在圖像處理中通常使用小矩陣，通過與輸入圖像進行卷積運算提取圖像特徵，稱為卷積核。在卷積神經網路中，卷積操作為核心操作之一。卷積層通過對輸入數據和卷積核進行卷積操作提取圖像或特徵圖之局部特徵。每個卷積核在輸入上滑動，並在每個位置上進行元素乘法和求和運算，得到一輸出值。通過使用多個不同卷積核，網絡可以學習不同特徵，例如邊緣、紋理和形狀等。

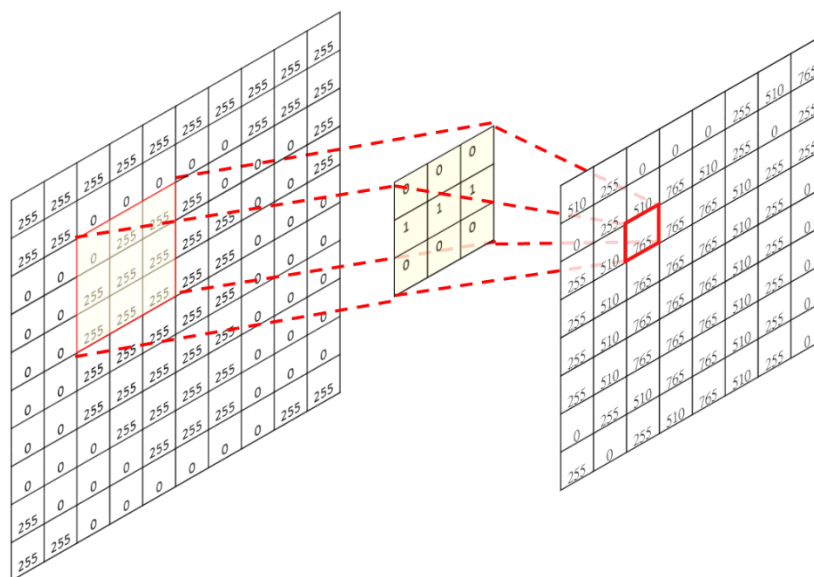


圖 2-4 卷積示意圖

2.4.1 轉置卷積

轉置卷積 (Transpose convolution) [16,17]為卷積操作之變體，通常用於圖像生成、圖像超分辨率和物體分割等任務中，其目的是將低維特徵映射至高維特徵空間。轉置卷積計算過程與普通卷積類似，但使用插值和填充方法，用以將輸入特徵圖擴展更大尺寸，如圖 2-5 所示。轉置卷積可以視為卷積之逆操作，也稱為反卷積 (Deconvolution)。

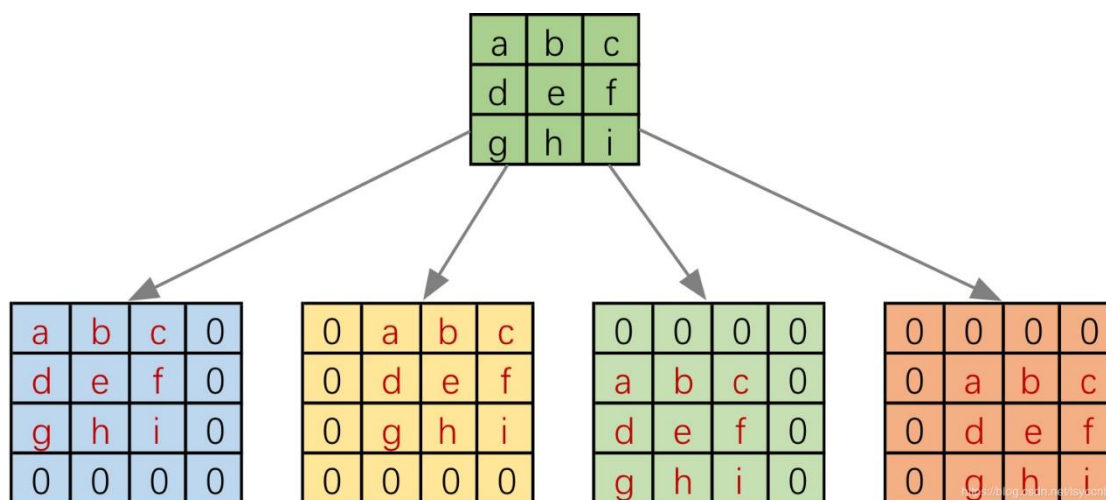


圖 2-5 轉置卷積示意圖

2.4.2 空洞卷積

空洞卷積[18,19]在卷積過程中引入膨脹率 (Dilation rate)，膨脹率控制卷積核中元素之間間距，從而調整卷積核對輸入特徵圖之感受野 (Receptive field)，如圖 2-6 所示。傳統卷積操作具有固定大小感受野，而空洞卷積通過增加卷積核元素之間間距，使得卷積核可以擴大其感受野，同時不改變卷積核大小卻捕捉更廣泛上下文訊息，對處理具不同尺度結構之圖像非常有用。當膨脹率越大，卷積核感受野範圍越廣，能夠捕捉更遠處特徵關係，從而改善圖像分割、物體檢測和邊緣檢測等任務之性能。

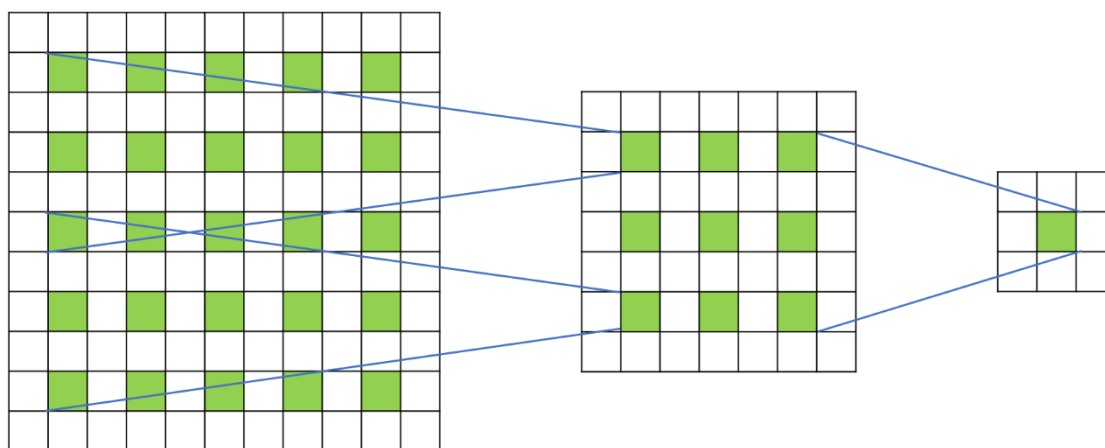


圖 2-6 空洞卷積示意圖

2.5 殘差神經網絡

Residual neural network (ResNet) [20] 為殘差神經網絡，深度神經網路近年在許多機器學習任務中表現出色，但隨著網路深度增加，梯度消失和梯度爆炸等問題也變得更加嚴重。為解決問題，He 等人於 2015 年提出殘差神經網路。殘差神經網路通過引入殘差連接，允許跨層直接傳遞信息，從而使訓練更深層網路變容易。

殘差神經網路通過引入殘差塊 (Residual block) 建構。每個殘差塊包含兩個連續之卷積層，每個卷積層之後進行批量正規化 (Batch normalization) 和非線性激活函數。在殘差塊中，通過添加殘差連接 (Residual connection)，將輸入特徵與輸出特徵相

加，此設計使神經網路學習其預測值之殘差而非直接學習輸入與輸出之間之映射關係，也稱之為殘差學習，如圖 2-7 所示，從而解決梯度消失與梯度爆炸問題。

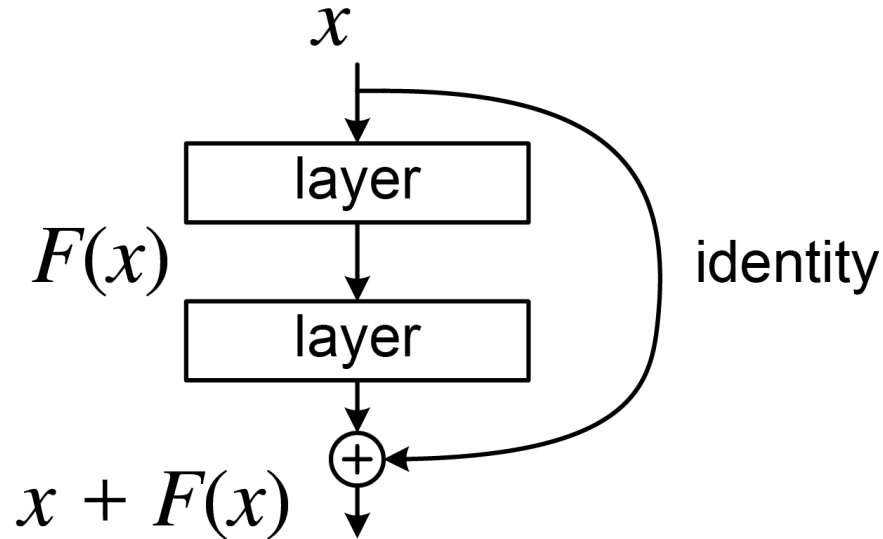


圖 2-7 殘差學習示意圖

2.6 Transformer 架構

Transformer[21,22]模型為採用注意力機制之深度學習模型，如圖 2-8 所示，其特色為不使用循環神經網路（RNN）或卷積神經網路（CNN），而以注意力機制處理輸入序列，並且能在序列中任意位置提供其上下文關係。除此之外，Transformer 架構能進行更多平行進算，進而減少訓練時間。

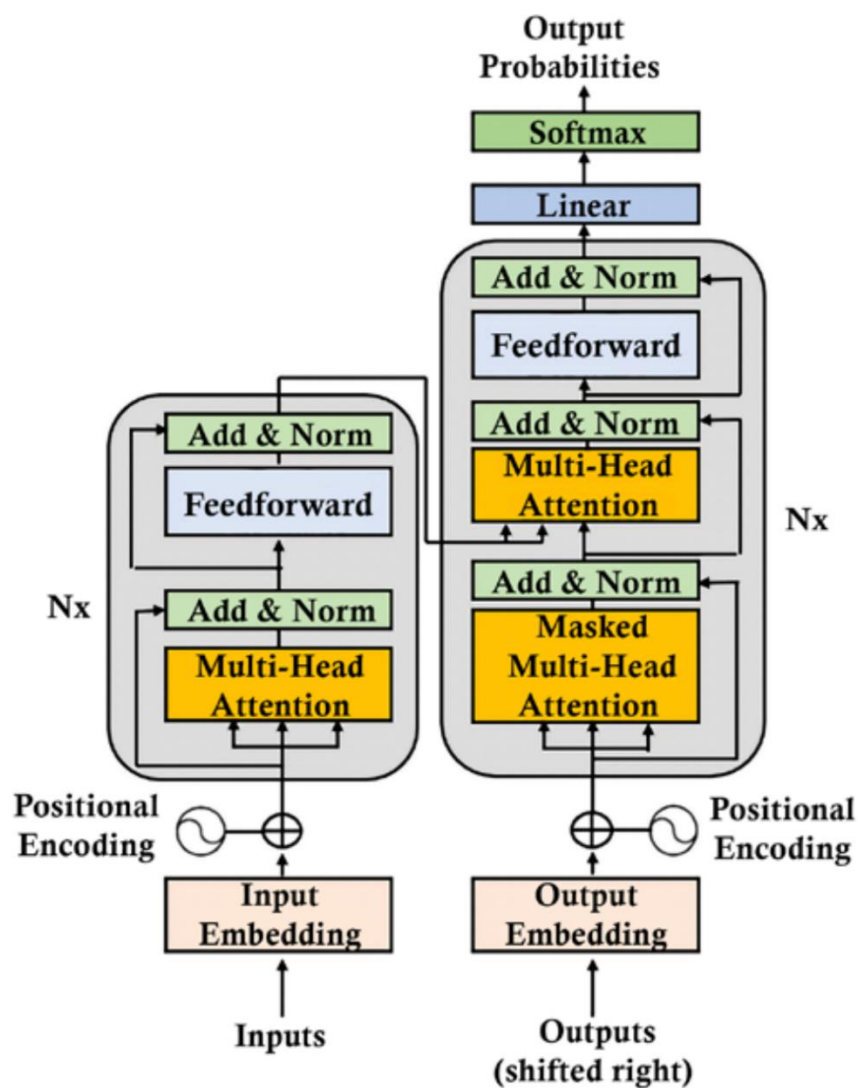


圖 2-8 transformer 架構示意圖

2.7 Swin U-Net 神經網路

Swin U-Net[23]為一種 U 型模型，如圖 2-9 所示，其具備 transformer 架構與 U-Net 之編碼-解碼特色，與 U-Net 最大不同之處在於 Swin U-Net 以 swin transformer block 取代原本 U-Net 中捲積層作為特徵提取器，將影像分割為多個不重疊 patch，並對每個 patch 以注意力計算，模型可以同時保持高分辨率特徵且有效處理大規模高解析度圖像。跨視窗注意力還可以減少模型計算與存儲需求，從而提高模型效率。

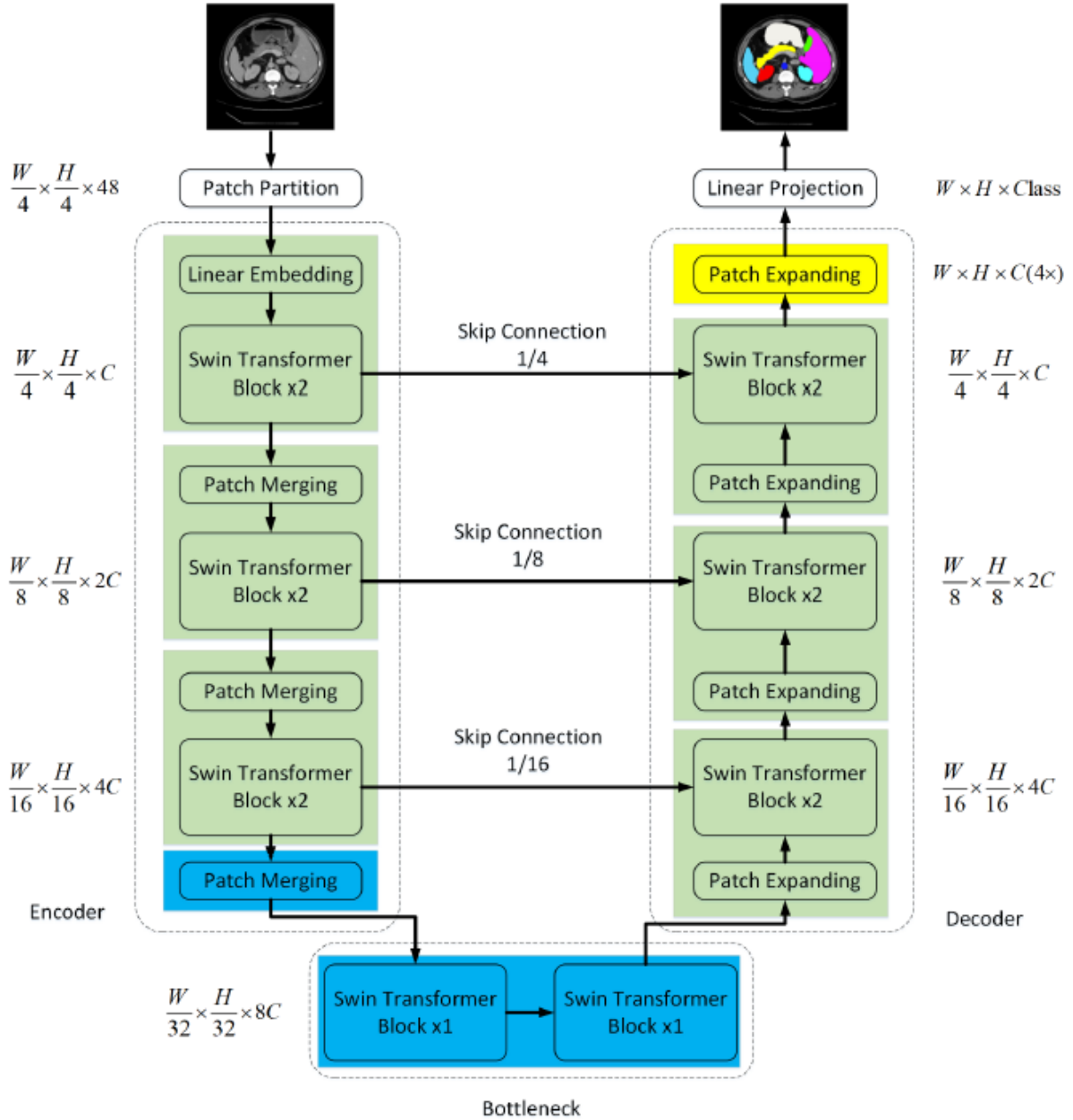


圖 2-9 Swin UNet 架構示意圖

2.7.1 Swin Transformer Block

Swin Transformer Block (Shifted window transformer block) [24] 引入局部注意力機制 (Local attention mechanism)，相較於傳統 Transformer，前者對於圖像處理更具優勢。傳統 Transformer 使用全局注意力機制 (Global attention mechanism)，即每個參數皆可與其他所有參數計算注意力，在處理長序列時會導致計算量和硬體內存開銷非常

大；而 Swin Transformer 使用局部注意力機制，在每個窗口內計算注意力，只關注局部參數，從而減少計算複雜度，同時保留重要局部資訊。

其次，Swin Transformer 採用滑動窗口（Shifted window）機制，如圖 2-10 所示。傳統 Transformer 並無使用窗口概念，所有參數皆為全局參與注意力計算。Swin Transformer 通過滑動窗口機制將輸入劃分為多個不重疊小窗口，在每個小窗口內計算注意力，再通過滑動窗口使窗口之間資訊能夠交流。此方式增強模型知表達能力以及局部結構之捕捉能力，並透過引入層級結構（Hierarchical structure）使 Swin Transformer 在不同層次上使用不同尺寸窗口，進而使模型能夠捕捉不同尺度資訊。此結構類似於卷積神經網絡中金字塔結構，有助於提取圖像中多尺度特徵。

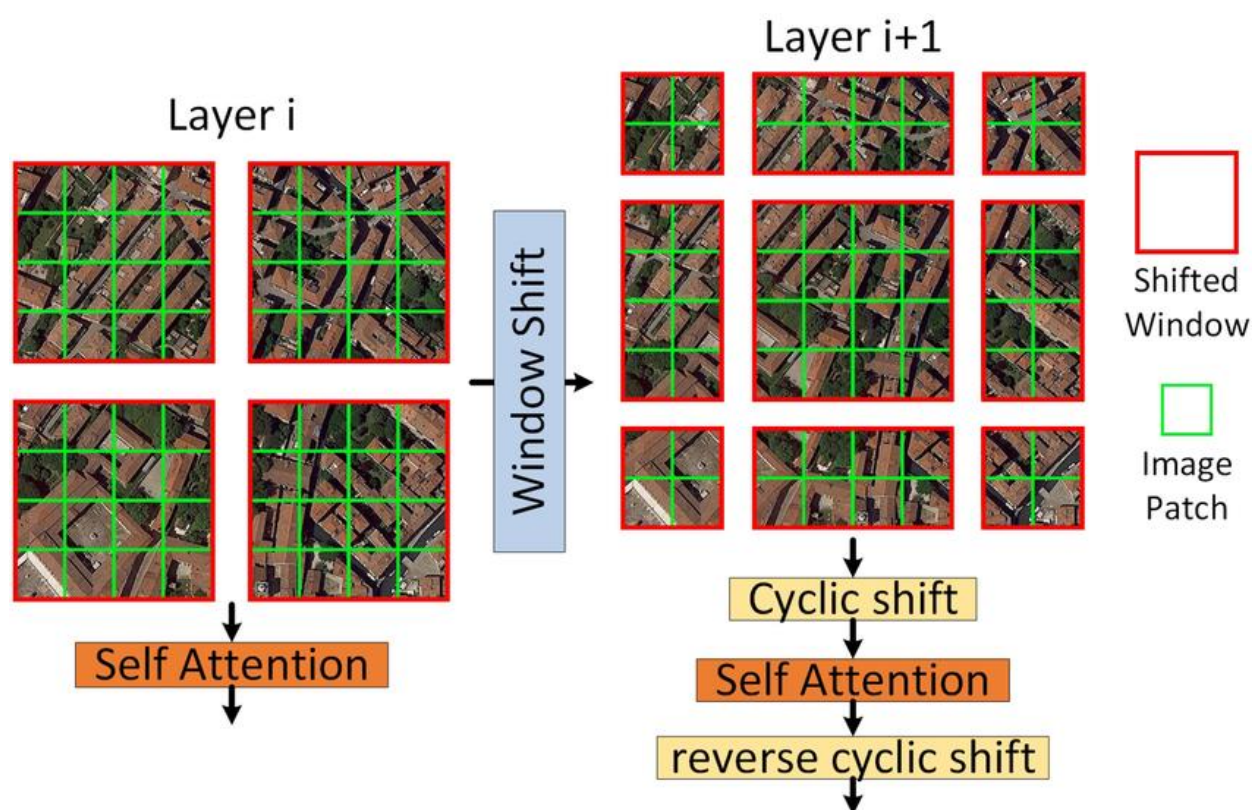


圖 2-10 Shifted window 示意圖

第三章 研究方法

本研究資料來源由台南奇美醫院提供。根據 CT 影像中腫瘤位置標記其病變器官，並繪製為影像遮罩。

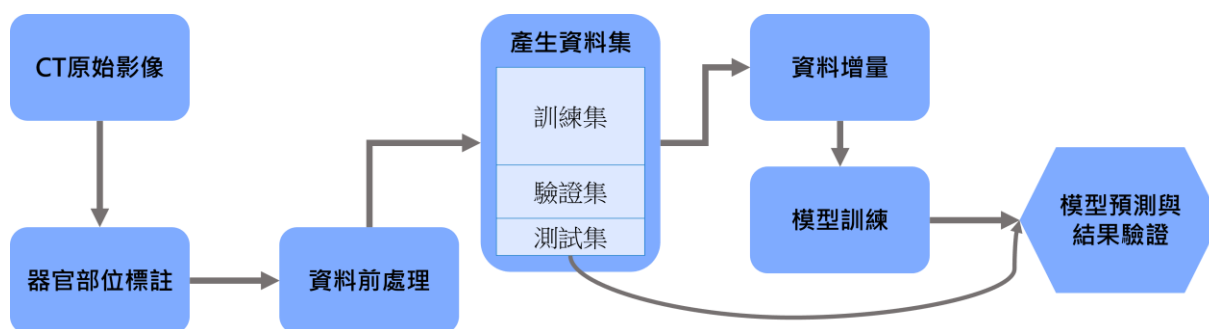


圖 3-1 單階段實驗流程

模型訓練流程圖如圖 3-1 所示，將原始影像資料進行標記，接著對資料進行去噪、歸一化等前處理，再對資料集進行資料分配，分成訓練集、驗證集和測試集，分配完成之資料集會再進行資料增量，增量後將訓練集以及驗證集資料輸入模型訓練，再使用訓練完成之模型預測測試集資料，評估其預測準確性。

3.1 資料處理

由台南奇美醫院提供罹患卵巢癌患者 CT 影像，並由中山大學資工所協助標記患者各部位病症，透過 labelbox 標記器官以及腫瘤，再從 labelbox 輸出標記完成影像，表 3-1 為收集影像中各器官部位之標記總數。由於需要進行 3D 模型訓練，必須先將同一位病人所拍攝多張 2D CT 影像疊加成為一組 3D volume CT[25,26,27]，再放入建立完成之模型進行訓練。

由於各組病患拍攝 CT 切片數量不一致以及硬體資源上限，並讓訓練資料能夠均勻取樣，將各組切片數量固定為 64。影像尺寸從原本 512 調整為 256，最後輸入 3D 模型中 CT volume 尺寸為 64*256*256；2D 模型則維持 512*512 大小。

表 3-1 各器官部位標記數量

標記部位	標記總數
肝臟	3283
淋巴結	7294
骨盆腔	1094

3.2 資料集

3.2.1 資料增量

深度學習模型訓練有賴於大數據以避免嚴重過度擬合（Overfitting），過度擬合指神經網路學習時過於緊密接近特定資料集，以至於無法精準預測特定資料集以外之其他資料集，且由於疾病罕見性、患者隱私及醫學專家對於標記影像要求，進行醫學成像過程所需費用與建構大型醫學圖像數據集較為困難，因此在資料集較少時易產生較大過度擬合問題。在較少數據學習應用中，常使用資料增量解決過度擬合問題，資料增量可增強原始數據集大小及內容，對於圖像進行資料增量包含幾何變換、色彩空間增強、混合影像、隨機擦除與特徵空間增強等方式，資料增量後數據集可提取原數據更多特徵，並從中獲得多樣數據集信息，可藉此訓練更佳深度學習模型，進而減少因數據集較少而導致過度擬合問題。

3.2.2 資料增量方法

由於 3D 模型使用 CT volume 方式訓練，會將同病人連續單張 CT 影像堆疊成類似 3D 影像，一組連續影像視為單組資料集，造成資料數大量減少，對於訓練成效不甚理想，因此對影像資料增量，如圖 3-2 所示。增量方式為向左旋轉 0~30 度、向右旋轉

0~30 度、增加亮度 10%、減少亮度 10%以及高斯模糊，總資料量會提升為原本 6 倍；2D 模型使用單張 CT 切片輸入模型訓練，資料數相較 3D 模型多，其資料增量方式與 3D 模型增量方式相同，增量後為原資料量 6 倍。

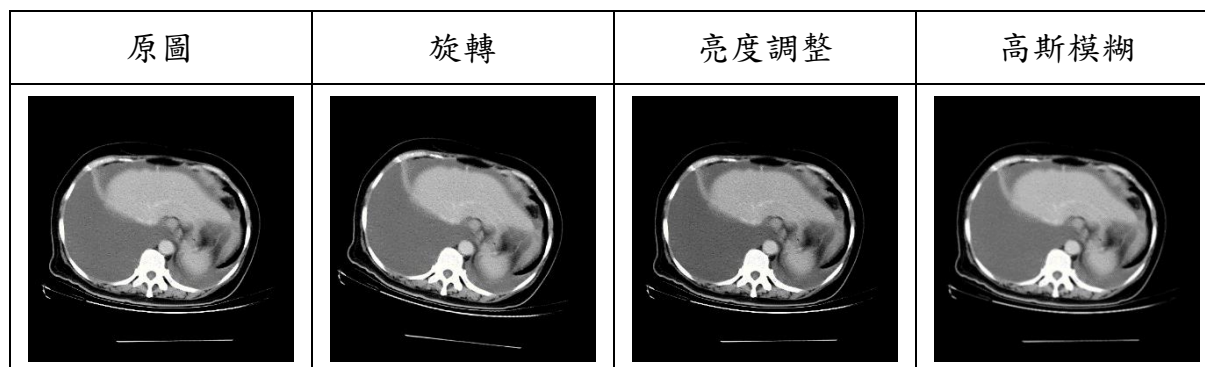


圖 3-2 資料增量示意圖

3.2.3 資料分配

訓練使用資料為 107 組病患所拍攝之胸腔至腹腔連續 CT 影像，總共包含 7294 張 CT 影像，經過資料增量後共 43764 張 CT 影像。由於相同病患其連續 CT 切片間具有高度相關性，因此在切分資料時須以病患為單位分配，資料分配如圖 3-3 所示。將 107 組病患先以 9:1 比例分為訓練資料和測試資料，再將訓練資料以 8:2 比例分為訓練集和驗證集。表 3-2 為資料分配後各資料集所包含不同器官部位之影像張數。分配後訓練集資料包含 77 組病患，共 5247 張 CT 影像，經資料增量後共 31482 張；驗證集包含 19 組病患，共 1295 張 CT 影像，經資料增量後共 7770 張；測試集包含 11 組病患，共 4512 張 CT 影像。

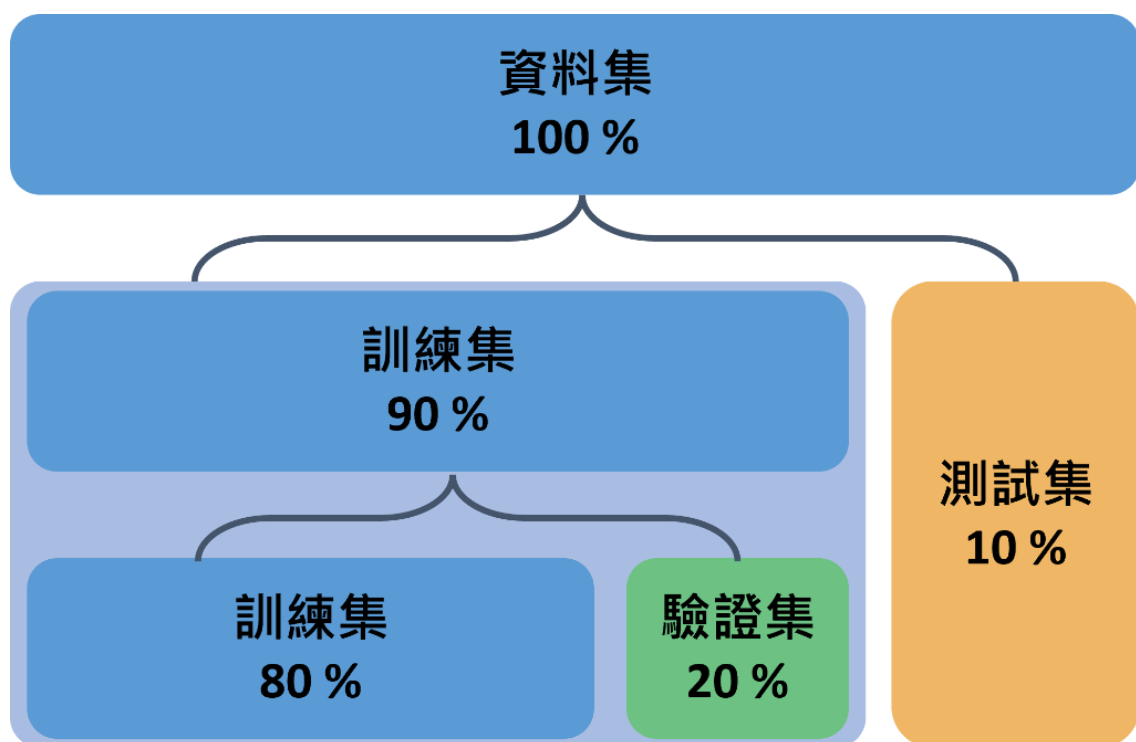


圖 3-3 資料分配示意圖

表 3-2 各資料集不同器官部位數量統計

標記部位	部位數量		
	訓練集	驗證集	測試集
肝臟	2356	593	334
淋巴結	5251	1314	729
骨盆腔	788	197	109

3.2.4 資料前處理

為確保輸入至模型之影像尺寸相同，統一將 CT 影像尺寸設為 512*512，輸入至 3D 模型之影像大小則設為 256*256。接著對影像進行歸一化（Normalization），將其原始像素值除以 255，使像素值介於 0 至 1 之間。對遮罩影像則根據不同標記類別對應 0、1、2、3 像素值，背景像素值設為 0，像素值 1、2、3 則分別為不同種類之器官標

記。

3.3 深度模型訓練

建構模型及訓練框架使用 Keras 和 Tensorflow，顯示卡使用 NVIDIA GeForce RTX4090，深度學習模型使用 3D U-Net、2D U-Net 以及 Swin U-Net。3D U-Net 之學習優化器使用 Adam，損失函數（Loss function）使用 dice loss，訓練次數為 100 epoch，學習率（Learning rate）設定為 10^{-4} ，由於單次輸入資料量龐大，其大小為 $256 \times 256 \times 64$ ，考量硬體上限，因此 batch size 設定為 1；2D U-Net 之學習優化與損失函數皆與 3D U-Net 相同，訓練次數為 100 epoch，學習率設定為 10^{-4} ，batch size 設定為 7，單次輸入資料大小為 $512 \times 512 \times 1$ ；Swin U-Net 之學習優化器使用 SGD，損失函數與學習率皆與 2D U-Net 相同，batch size 設定為 4，patch size 設定為 2，stack number 設定為 3，單次輸入資料大小為 $512 \times 512 \times 1$ 。

3.4 3D U-Net

由於所收集之 CT 影像並非單一張 2D 影像，而是以每位病患為一組，每組包含其病患由胸腔至腹腔之連續 CT 切片，因此研究初期使用 3D U-Net[28]作為 CT 影像之訓練模型，以每組病患為單位輸入。由於每組病患所拍攝之 CT 切片數量並不一致，並且考量硬體內存限制，因此統一將每組病患之 CT 切片數量固定為 64，根據每組病患切片總數除以 64，並以其商數作為選取切片之間隔，避免輸入影像之切片部位過度集中於胸腔而忽略腹腔部位。由於輸入維度增加 64 倍，為符合硬體上限，將原圖長寬縮小為原始大小一半，其輸入維度為 $256 \times 256 \times 64$ 。

3.5 2D U-Net

雖然 3D U-Net 能夠更好地捕捉同一位病患不同 CT 切片間之空間訊息，但受限於硬體上限不得不降低每張影像解析度，同時模型輸入需要以每位病患為單位，相對於

2D 模型以單張影像作為輸入，大幅減少資料量，影響模型訓練成效。因此後續改以單張 CT 影像作為輸入，雖然損失同一組連續 CT 切片間空間資訊，但圖像能以完整尺寸輸入模型，但資料量卻能大幅提升。

本研究前期使用 2D U-Net 作為 CT 影像之訓練模型，並在原本 U-Net 特徵提取部分以 ResNet 取代，由於 ResNet 具有不同深度之神經網路，層數越多代表其辨識能力更好，但運算參數也會因此增加，考量硬體成本以及避免練過度擬合問題選取適宜深度，最後再加入注意力機制，用以提高模型辨識目標準確度。

3.5.1 2D U-Net 模型骨架

為增加 2D U-Net 特徵提取能力，加入 ResNet-50 作為 U-Net 骨架（Backbone），如圖 3-4 所示。在 U-Net 解碼部分，由 ResNet-50 進行特徵提取，再將其輸出至 U-Net，接著 U-Net 對其進行 maxpooling，往下傳遞至下一層，再交由 ResNet-50 提取特徵，重複此動作直至 U-Net 最底層。

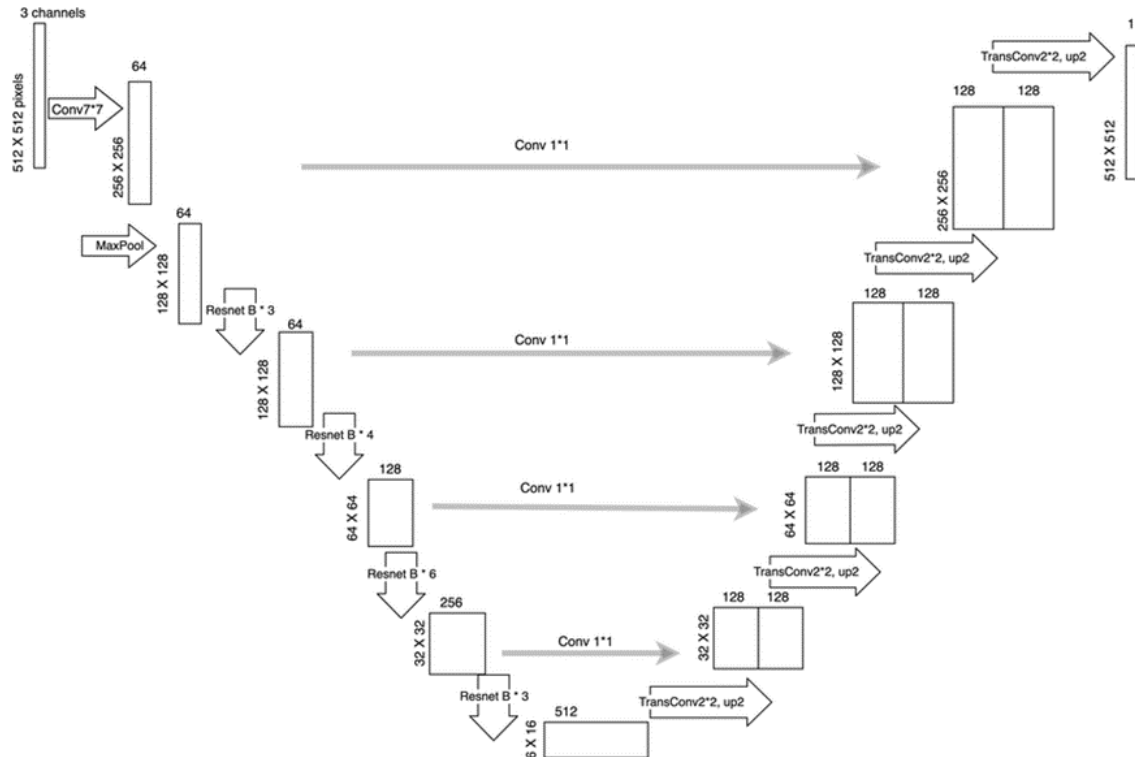


圖 3-4 U-Net+ResNet 示意圖

3.5.2 注意力機制

在 2D U-Net 加入注意力機制 (Attention mechanism)，如圖 3-5 所示。透過注意力機制可以更好分辨目標特徵和其餘不相關特徵，藉此增加模型對於特徵之關注度，提高模型訓練準確度。

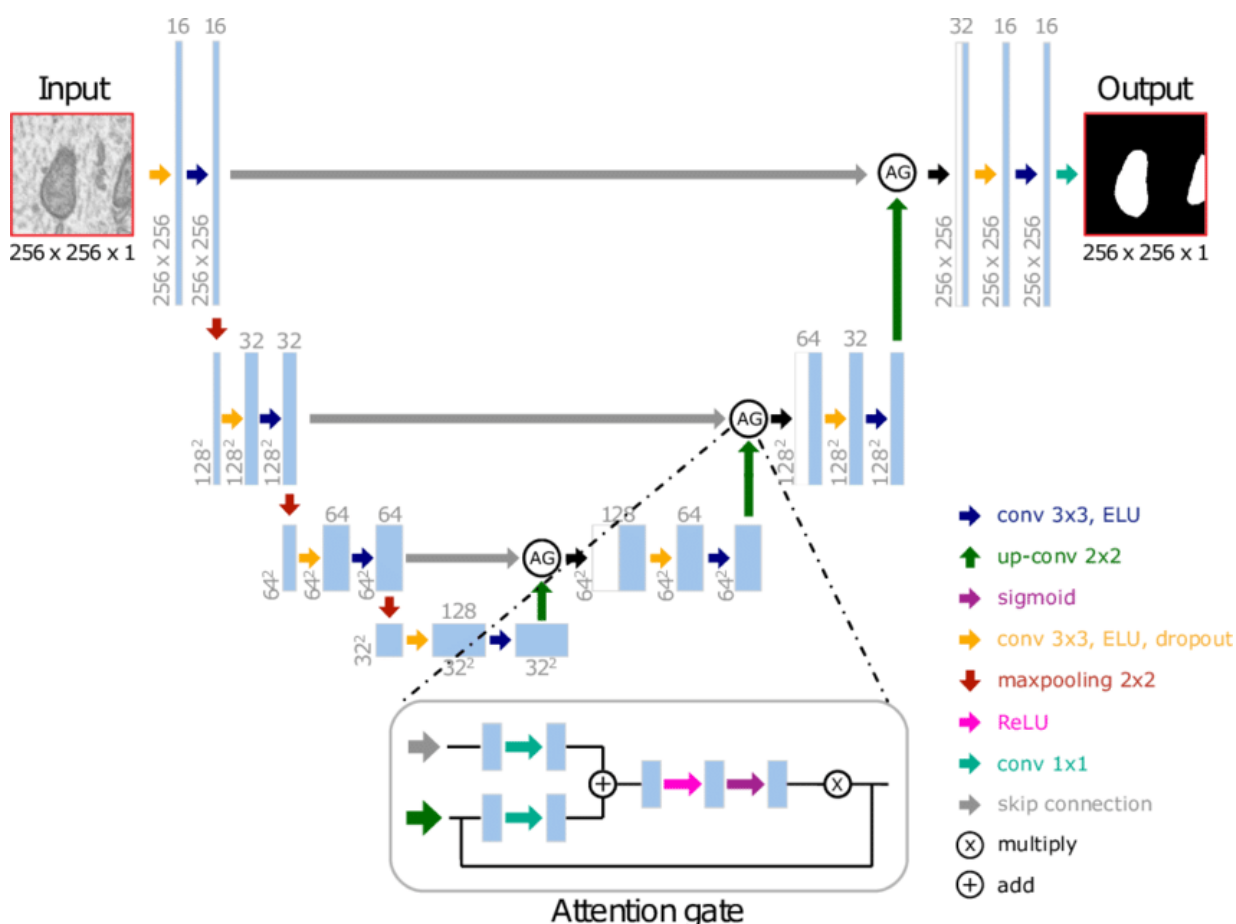


圖 3-5 U-Net+注意力機制示意圖

3.6 Swin U-Net

雖然 2D U-Net 於醫學影像分割之效果顯著，但其模型架構較舊，因此研究後期使用具有 transformer 架構之模型 Swin U-Net 訓練 CT 影像。Swin U-Net 採用與 U-Net 相同結構：編碼-解碼結構，透過降維過程提取不同維度特徵。此外，其採用改良之 transformer 架構，為降低運算複雜度而使用非全局式自注意力，再透過移動窗口方式

使相同影像中不同分割區域間進行自注意力，避免只關注影像局部特徵，如此不僅能關注整張影像所有細節，相比傳統 transformer 架構更大幅減少模型運算時間複雜度。

3.6.1 Transformer block

將 Swin U-Net 中 transformer block 數量由 2 層改為 3 層，增加模型計算量同時也能捕獲更多影像特徵，並將模型 dropout 設定為 0.1 以避免模型出現過度擬合 (overfitting) 問題。

3.7 模型評估

評估模型訓練成果使用 Accuracy 與 MIoU (Mean Intersection over Union) 進行計算，MIoU 計算方式如圖 3-6 所示，當 MIoU 值越接近 1，代表其準確度越高。

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{FN + FP + TP}$$

圖 3-6 MIoU 計算公式

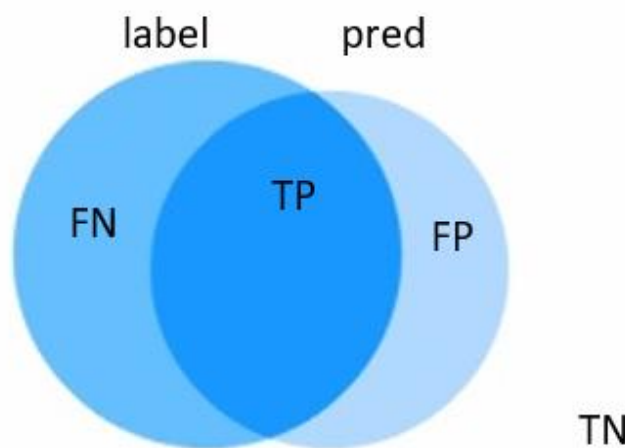


圖 3-7 IoU 示意圖

第四章 研究結果

4.1 模型訓練

4.1.1 3D 模型

本文使用資料集共包含 107 組病患之 CT 影像，經資料增量後共 642 組，使用模型為 3D U-Net，輸入影像尺寸為 $64*256*256$ ，訓練結果如圖 4-1 所示。

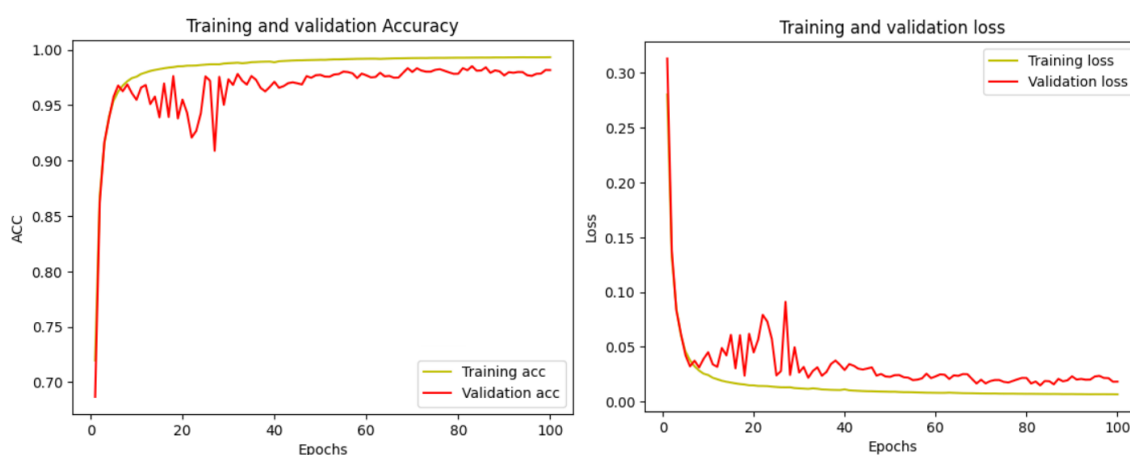


圖 4-1 3D 模型訓練結果圖

完成訓練後利用 MIoU 對模型預測影像之準確度進行評估，表 4-1 為各資料集 MIoU 結果，訓練集 MIoU 約為 61%，驗證集 MIoU 約為 53%，測試集 MIoU 約為 39%；表 4-2 為各資料集中不同器官部位之 MIoU 結果。由於淋巴結部位標記面積相較肝臟及卵巢小，加上 3D 訓練時將每張 CT 切片縮小至 $256*256$ 大小，進一步降低其標記之顯著度，使模型訓練過程中難以有效獲取其特徵，造成淋巴結部位預測結果較不理想。

表 4-1 MIoU 評估結果

訓練集	0.6139
驗證集	0.5271
測試集	0.3863

表 4-2 各類器官部位 MIoU

部位	MIoU		
	訓練集	驗證集	測試集
肝臟	0.7787	0.7143	0.5622
淋巴結	0.4226	0.3424	0.2147
骨盆腔	0.6403	0.5246	0.3818

4.1.2 2D 模型

表 4-3 為不同 2D 模型之訓練結果比較。研究初期採用 U-Net 作為模型訓練，為提升模型準確度，測試加入 ResNet 與注意力機制至模型中，並與 U-Net 進行比較，測試結果得知 U-Net 採用注意力機制以及使用 ResNet 作為骨架效果最佳。研究後期採用 Swin U-Net 作為模型訓練，其訓練結果優於加入 ResNet 與注意力機制之 U-Net，因此最終選擇使用 Swin U-Net 預測 CT 影像。

表 4-3 2D 模型訓練結果比較

model	Accuracy		Mean IoU	
	Training	Validation	Training	Validation
U-Net	96.3	85.6	92.2	79.5
U-Net + ResNet	95.9	84.2	94.3	78.4
U-Net + attention	96.4	86.1	97.1	76.8
U-Net + ResNet + attention	96.3	84.7	95.6	77.2
Swin U-Net	95.2	84.4	92.3	73.1

2D 模型使用資料集共 7924 張 CT 影像，經資料增量後共 43764 張，輸入影像尺寸為 $1*512*512$ ，訓練結果如圖 4-2 所示。

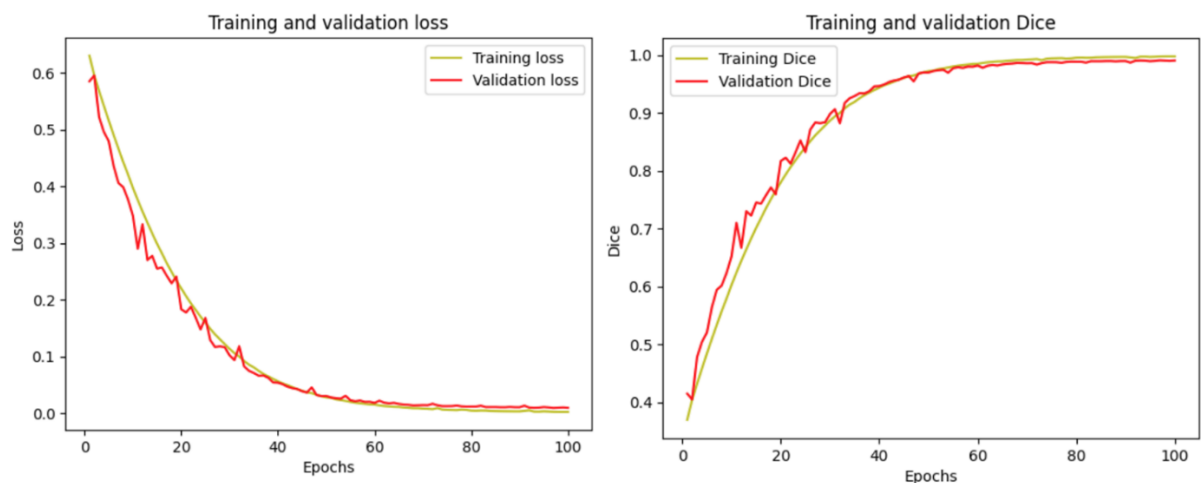


圖 4-2 2D 模型訓練結果圖

完成訓練後利用 MIoU 對模型預測影像之準確度進行評估，表 4-4 為各資料集 MIoU 結果。訓練集 MIoU 約為 93%，驗證集 MIoU 約為 72%，測試集 MIoU 約為 71%；表 4-5 為各資料集中不同器官部位之 MIoU 結果。

表 4-4 MIoU 評估結果

訓練集	0.9260
驗證集	0.7235
測試集	0.7138

表 4-5 各類器官部位 MIoU

部位	MIoU		
	訓練集	驗證集	測試集
肝臟	0.9682	0.8533	0.8162
淋巴結	0.9031	0.7215	0.6510
骨盆腔	0.8867	0.7157	0.6142

4.2 影像訓練結果

由於影像 2D 模型訓練結果優於 3D 模型，因此最後採用 2D 模型作為影像預測方法。圖 4-3 及圖 4-4 為測試組其中一組病患之預測結果，藍色塗層為肝臟部位，粉紅色塗層為淋巴結部位，綠色塗層為骨盆腔部位。

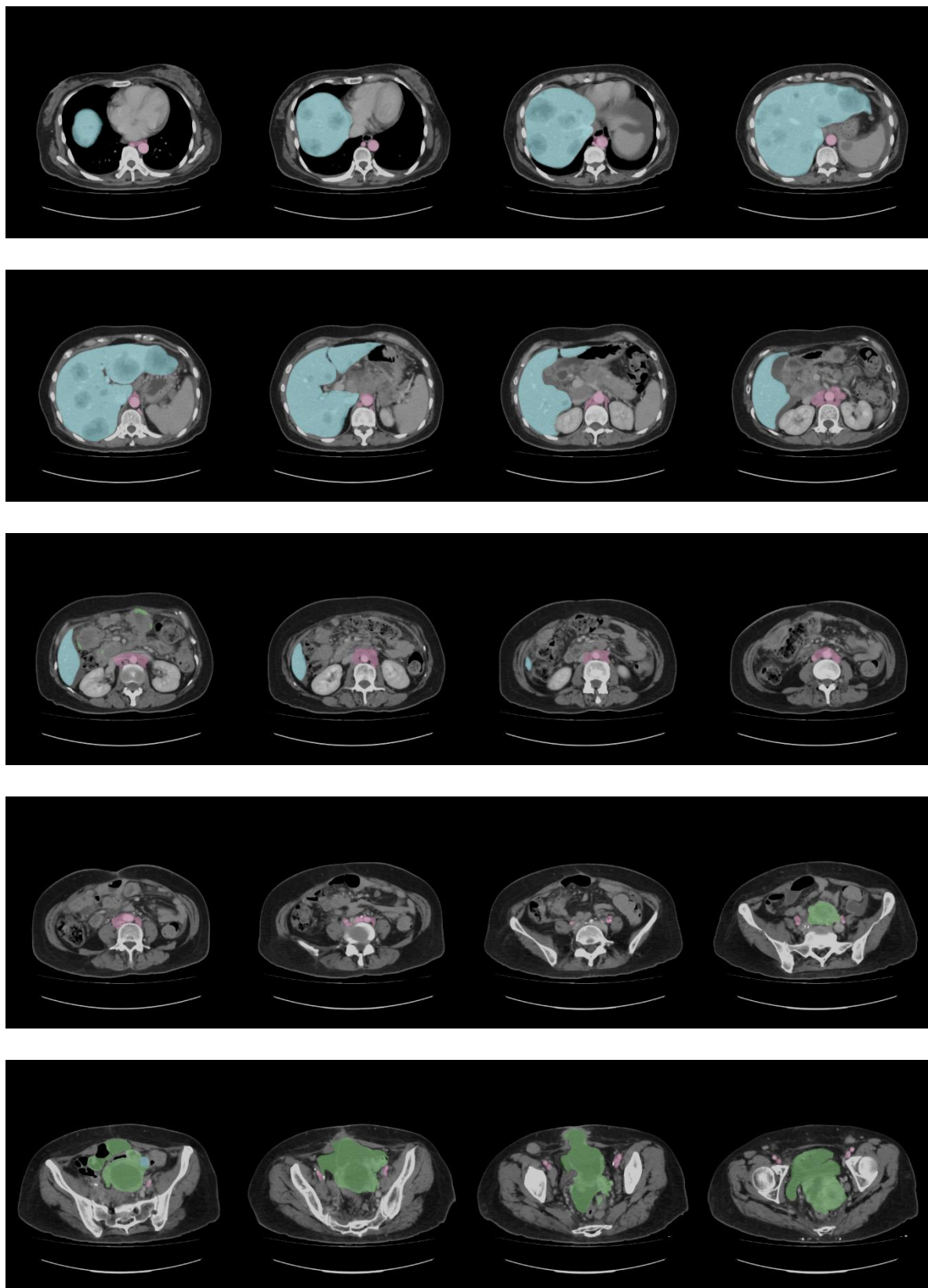


圖 4-3 器官預測結果

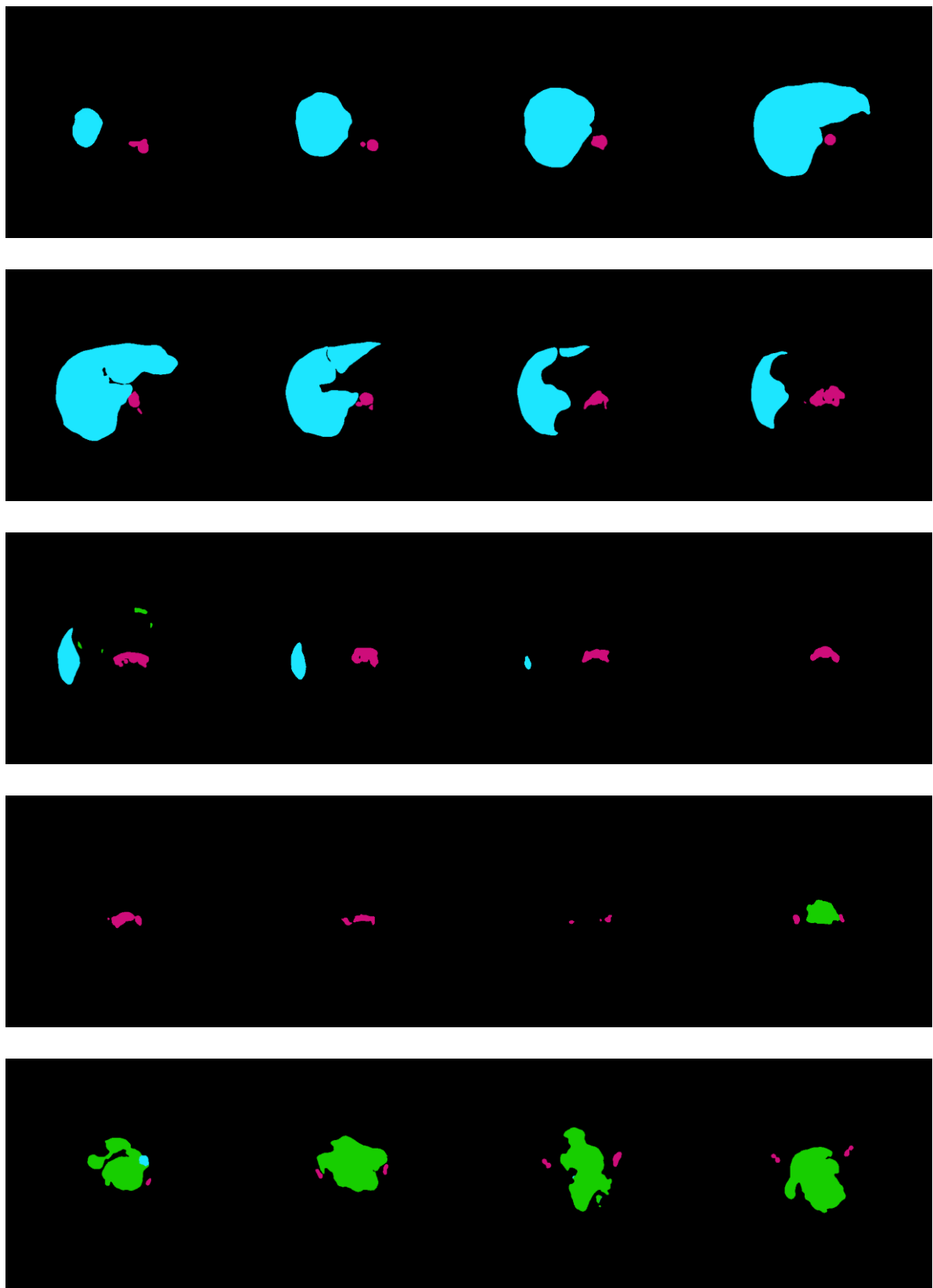


圖 4-4 器官預測遮罩

在大部分情況中，肝臟與淋巴結之預測位置變化幅度不大，骨盆腔位置則較容易預測錯誤，表 4-6 展示預測較優之結果，表 4-7 展示預測較差之結果。

表 4-6 較優預測結果

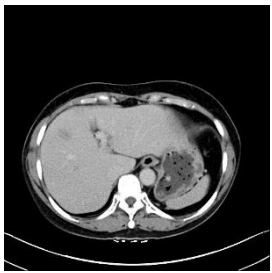
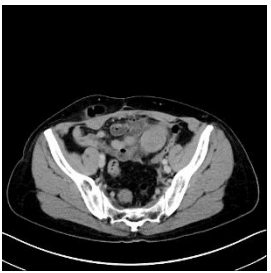
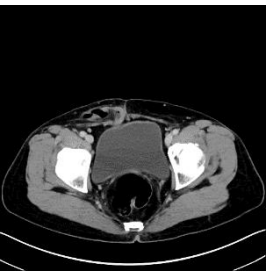
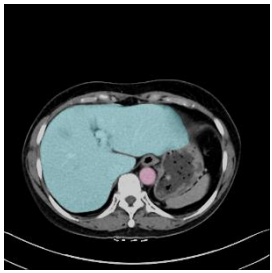
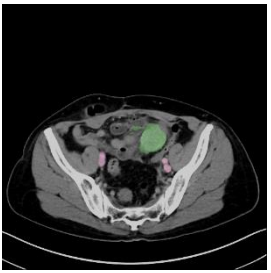
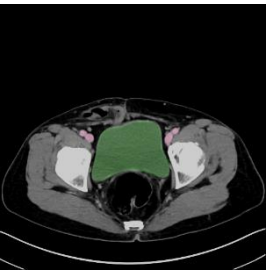
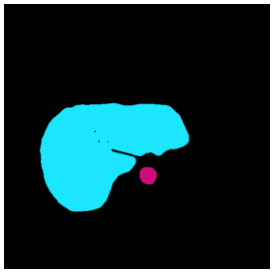


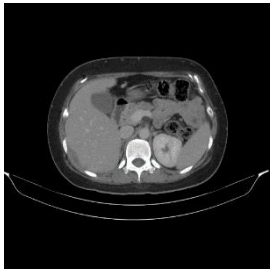

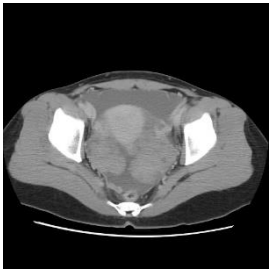
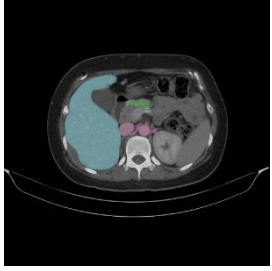
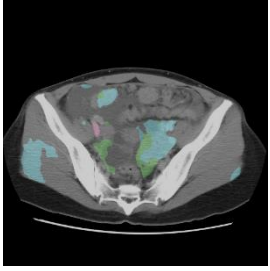
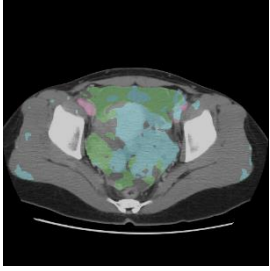

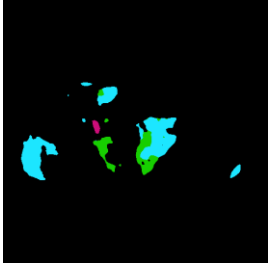
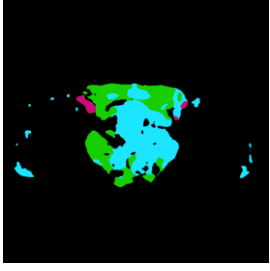
原圖			
模型預測			
遮罩			

表 4-7 較差預測結果

原圖			
模型預測			
遮罩			

4.3 腫瘤預測結果比較

使用本研究模型預測結果作為第二階段模型之輸入，表 4-8 顯示單模型與二階段對於肝臟腫瘤、淋巴結腫瘤、卵巢腫瘤辨識準確度，結果顯示上述三種位置之腫瘤辨識準確度皆有提升。

表 4-8 模型辨識 MIoU 比較

模型	病徵類別		
	肝臟腫瘤	淋巴結腫瘤	卵巢腫瘤
單階段模型	0.7109	0.5171	0.7874
二階段模型	0.7815	0.6117	0.8399

第五章 結論

本研究針對肝臟轉移腫瘤、淋巴結轉移腫瘤、卵巢腫瘤及其患發之器官進行影像分割，其目的是更準確預測各器官中是否存在癌症病灶，透過模型預測相應器官，再針對預測部位進行更細節檢測。研究結果發現，二階段模型相比於單階段模型，其肝臟腫瘤、淋巴結腫瘤、卵巢腫瘤之預測準確度皆有提升，對於尺度較小腫瘤辨別能力也更高，由此可得知二階段模型分割能力優於單階段模型。

考量每位病患多張連續 CT 切片具有高度相關性，為更好捕捉切片之間上下文關係，因此研究採用 3D 模型訓練，但受限於硬體設備，無法將完整尺寸影像資料輸入模型，導致預測結果不甚理想；儘管 2D 模型失去空間相關訊息，但資料數相比 3D 模型大量增加，因此模型訓練結果準確度較高。

由於腫瘤之形狀、大小皆不固定，因此不同病患之器官可能與健康者之 CT 影像有顯著差異，進一步增加模型訓練難度，卵巢部位尤為明顯，因部分病患已透過手術切除卵巢，在訓練過程中容易誤導模型，進而導致訓練結果於無卵巢病患之影像中錯誤預測卵巢。

除以上三種，還有其他擴散於胸腔和腹腔之腫瘤，但由於其患發位置不固定，或患發後與原本無病症時相差巨大，造成標記困難，因此難以透過模型預測其他不固定腫瘤之位置。

未來若能提升硬體上限，並且收集更多組病患資料，對於 3D 模型進行訓練會更加友善，空間維度資訊之捕捉對於連續 CT 影像非常重要，也是 2D 模型無法提取之特徵，因此若沒有硬體限制以及資料稀缺性影響，3D 模型將會是更合適選擇。

參考文獻

- [1] Ursula A. Matulonis, Anil K. Sood, Lesley Fallowfield, Brooke E. Howitt, Jalid Sehouli, and Beth Y. Karlan. "Ovarian Cancer." *Nature Reviews Disease Primers* 2 (2016): article number: 16061.
- [2] Alexi A. Wright, Kari Bohlke, Deborah K. Armstrong, Michael A. Bookman, William A. Cliby, Robert L. Coleman, Don S. Dizon, Joseph J. Kash, Larissa A. Meyer, Kathleen N. Moore, Alexander B. Olawaiye, Jessica Oldham, Ritu Salani, Dee Sparacio, William P. Tew, Ignace Vergote, and Mitchell I. Edelson. "Neoadjuvant Chemotherapy for Newly Diagnosed, Advanced Ovarian Cancer: Society of Gynecologic Oncology and American Society of Clinical Oncology Clinical Practice Guideline." *Gynecologic Oncology* 143, no. 1 (2016): pp. 3-15.
- [3] Panagiotis Filis, Despoina Mauri, Georgios Markozannes, Maria Tolia, Nikolaos Filis, and Konstantinos Tsilidis. "Hyperthermic Intraperitoneal Chemotherapy (HIPEC) for the Management of Primary Advanced and Recurrent Ovarian Cancer: A Systematic Review and Meta-analysis of Randomized Trials." *ESMO Open* 7, no. 2 (2022): article number: 100586.
- [4] Chien-Cheng Lee, Pau-Choo Chung, and Hong-Ming Tsai. "Identifying Multiple Abdominal Organs from CT Image Series Using a Multimodule Contextual Neural Network and Spatial Fuzzy Rules." *IEEE Transactions on Information Technology in Biomedicine* 7, no. 3 (2003): pp. 212-221.
- [5] Chenyang Liu, and Shing-Chow Chan. "A Joint Detection and Recognition Approach to Lung Cancer Diagnosis from CT Images with Label Uncertainty." *IEEE Transactions on Medical Imaging* 39, no. 4 (2020): pp. 1127-1137.
- [6] Mukesh Mishra, and Manish Srivastava. "A View of Artificial Neural Network." In

- Proceedings of the 2014 International Conference on Advances in Engineering and Technology Research (ICAETR)*, 2014, pp. 1-3.
- [7] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. "Deep Learning." Cambridge, MA: MIT Press, 2016.
 - [8] Nir Shlezinger, John Whang, Yonina C. Eldar, and Alexandros G. Dimakis. "Model-Based Deep Learning." *IEEE Signal Processing Magazine* 38, no. 2 (2021): pp. 16-27.
 - [9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." *arXiv* (2015).
<https://doi.org/10.48550/arXiv.1505.04597>.
 - [10] Zhaoyang Niu, Guoqiang Zhong, and Hui Yu. "A Review on the Attention Mechanism of Deep Learning." *Neurocomputing* 447 (2021): pp. 158-175.
 - [11] Jianxin Zhang, Zongkang Jiang, Jing Dong, Yaqing Hou, and Bin Liu, "Attention Gate ResU-Net for Automatic MRI Brain Tumor Segmentation," *Biocybernetics and Biomedical Engineering* 41, no. 1 (2021): pp. 262-276.
 - [12] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y. Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. "Attention U-Net: Learning Where to Look for the Pancreas." *arXiv* (2018). <https://doi.org/10.48550/arXiv.1804.03999>.
 - [13] Zewen Li, Fan Liu, Wenjie Yang, Shouheng Peng, and Jun Zhou. "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects." *IEEE Transactions on Neural Networks and Learning Systems* 32, no. 6 (2021): pp. 2166-2182.
 - [14] Yann LeCun, and Yoshua Bengio. "Convolutional Networks and Applications in Vision." In *Handbook of Brain Theory and Neural Networks*, edited by Michael A. Arbib, pp. 255-258. Cambridge, MA: MIT Press, 2002.

- [15] Matthew D. Zeiler, and Rob Fergus. "Visualizing and Understanding Convolutional Networks." In *Computer Vision – ECCV 2014* (2014): pp. 818-833.
- [16] Guilin Liu, Rohan Taori, Ting-Chun Wang, Zhiding Yu, Shiqiu Liu, Fitsum A. Reda, Karan Sapra, Andrew Tao, and Bryan Catanzaro. "Transposer: Universal Texture Synthesis Using Feature Maps as Transposed Convolution Filter." *arXiv* (2020). <https://doi.org/10.48550/arXiv.2007.07243>.
- [17] Vincent Dumoulin, and Francesco Visin. "A Guide to Convolution Arithmetic for Deep Learning." *arXiv* (2016). <https://doi.org/10.48550/arXiv.1603.07285>.
- [18] Fisher Yu, and Vladlen Koltun. "Multi-Scale Context Aggregation by Dilated Convolutions." *arXiv* (2016). <https://doi.org/10.48550/arXiv.1606.00915>.
- [19] Yunchao Wei, Huaxin Xiao, Honghui Shi, Zequn Jie, Jiashi Feng, and Thomas S. Huang. "Revisiting Dilated Convolution: A Simple Approach for Weakly- and Semi-Supervised Semantic Segmentation." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7268-7277.
- [20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778.
- [21] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention Is All You Need." In *Advances in Neural Information Processing Systems*, 2017, pp. 5998-6008.
- [22] Hanguang Xiao, Li Li, Qiyuan Liu, Xiuhong Zhu, and Qihang Zhang. "Transformers in Medical Image Segmentation: A Review." *arXiv* (2023): <https://doi.org/10.1016/j.bspc.2023.104791>.
- [23] Yizhou Chen, Yuyin Zhou, Ximei Li, and Yu Qiao. "Swin-Unet: UNet-Like Pure Transformer for Medical Image Segmentation." *arXiv* (2023).

<https://doi.org/10.48550/arXiv.2301.23045>.

- [24] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows." *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 10012-10022.
- [25] Lei Song, Haoqian Wang, and Z. Jane Wang. "Bridging the Gap Between 2D and 3D Contexts in CT Volume for Liver and Tumor Segmentation." *IEEE Journal of Biomedical and Health Informatics* 25, no. 4 (2021): pp. 1263-1272.
- [26] Gopi Kasinathan, Selvakumar Jayakumar, Amir H. Gandomi, Manikandan Ramachandran, Simon James Fong, and Rizwan Patan. "Automated 3-D Lung Tumor Detection and Classification by an Active Contour Model and CNN Classifier." *Journal of Computational Biology* 27, no. 3 (2020): pp. 329-342.
- [27] Holger R. Roth, Chen Shen, Hirohisa Oda, Takaaki Sugino, Masahiro Oda, Yuichiro Hayashi, Kazunari Misawa, and Kensaku Mori. "A Multi-Scale Pyramid of 3D Fully Convolutional Networks for Abdominal Multi-Organ Segmentation." In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2019*, pp. 487-495.
- [28] Juarez, Antonio, Raghavendra Selvan, Zaigham Saghir, and Marleen de Bruijne. "A Joint 3D UNet-Graph Neural Network-Based Method for Airway Segmentation from Chest CTs." In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2019*, pp. 663-671.