

國立屏東科技大學資訊管理系

碩士學位論文

基於輕量型密集連接網路之擴大窄頻影像圖像胃癌區
域檢測

Gastric Cancer Region Detection for Magnified Narrow
Band Imaging Images Based on Lightweight Densely
Connected Convolutional Networks

指導教授：許志仲博士

研究生：簡志宇

中華民國 109 年 6 月 8 日

摘要

學號：M10756010

論文名稱：基於輕量型密集連接網路之擴大窄頻影像圖像胃癌區域檢測

總頁數：56

學校名稱：國立屏東科技大學

系（所）別：資訊管理系

畢業年度及摘要別：108學年度第二學期碩士學位論文摘要

研究生：簡志宇

指導教授：許志仲 博士

論文摘要內容：

通常，深度學習需要大規模之訓練資料才能在各種任務上達到良好且穩定之效能。但是，在幾種應用中，例如醫學成像或從新設備取得之成像，能夠取得之訓練樣本遠遠不足。這一問題引起了各方學者討論如何使用深度學習來提高樣本數量視作一個議題。由於 Magnified Narrow Band Imaging (M-NBI) 影像在國內不好取得，能委託製作 Label Data 之醫生也很少，故本論文基於輕量化 (Lightweight) Densely Connected Convolutional Networks (L-DenseNet) 作為骨幹網路為基礎，並且運用特殊之資料預處理方式有效提高訓練樣本。由於僅使用 FCN 進行 Semantic Segmentation 檢測癌症區域速度並不快，為了提升 M-NBI 之癌症檢測速度本論文將檢測網路分為兩個階段進行，在進行 Semantic Segmentation 前先做 Object Detection。Object Detection 主要框架為 One Stage 之 Scalable and Efficient Object Detection Network (EfficientDet)，將 M-NBI 圖像上之癌症區域進行初步預測，並且僅將該預測區域進行 Semantic Segmentation，縮短 FCN 檢測 M-NBI 癌症區域之時間。本計畫之 M-NBI 癌症區域檢測網路框架預期可用於彌

補小資料集資料不足之樣本問題，同時能提高整體檢測之速度。

關鍵字：擴大窄頻影像、目標檢測、語義分割



Abstract

Student ID: M10756010

Title of Thesis: Gastric Cancer Region Detection for Magnified Narrow Band Imaging
Images Based on Lightweight Densely Connected Convolutional
Networks

Total page: 56

Name of Institute: Department of Management Information Systems, National Pingtung
University of Science and Technology

Graduate date: June . 2020

Degree Conferred: Master

Name of student: Chih-Yu Jian

Advisor: Dr. Chih-Chung Hsu

The contents of abstract in this thesis :

Deep learning requires a large-scale training dataset to achieve excellent and stable performance on various computer vision tasks. In medical image processing, the number of training samples is usually insufficient so that we are unable to train a stable deep learning-based classifier. Recently, Magnified Narrow Band Imaging (M-NBI) is widely used to diagnose the early cancerous in Gastric regions. How to obtain a sufficient number of the training samples of M-NBI becomes a serious problem. Since M-NBI images are not easy to obtain in domestic, few experts who can well annotate the raw data. In this paper, we aim to propose an automatic diagnose system to achieve the gastric cancerous region segmentation to reduce the effort of the doctor during diagnosing. It is well-known that the semantic segmentation is relative slow so that it may be unable to detect the cancerous regions in real-time. Meanwhile, recent object detection techniques show the excellent performance with fast process. In this paper, we take the advantages from both object detection and semantic segmentation techniques to build a real-time gastric cancer segmentation system. The main framework of Object Detection in One Stage's Scalable and Efficient Object Detection Network (EfficientDet), which initially predicts the cancer area on the M-NBI image, and only performs Semantic Segmentation on the predicted area, shortening the time for FCN to detect the M-NBI cancer area.

Furthermore, a lightweight Densely Connected Convolutional Networks (L-DenseNet) is proposed as a backbone network and uses special data augmentation methods to fit the requirement of small-scale training set, which our method should be able to overcome the problem caused by a limited number of the training samples as well as improves both accuracy and overall detection. Finally, the experimental results demonstrated that the proposed method can achieve state-of-the-art performance over other methods.

Keywords : M-NBI 、 Object Detection 、 Semantic Segmentation



謝誌

碩士兩年的學程即將結束，在求學過程中非常充實也學習到很多進入職場時該注意的事情，學生非常感謝指導教授 許志仲教授，時常推薦學生去參與研討會與競賽來增加自己的實力和經驗，同時也給學生所多對人應變能力更為嚴謹與流暢。雖然過程十分辛苦，但在這短短的兩年內學生獲得很大的收穫，教授的教誨學生謹記於心。

此外，感謝於論文口試委員張軒庭 教授、康立威 教授、李佳燕教授、李宗鎔 醫師、許志仲 教授給予學生寶貴的建議，使本篇論文能夠更加的嚴謹完善，學生於此僅深致謝。

最後，感謝實驗室學弟慶豪、學妹郁芳，於論文撰寫期間幫忙維護實驗室各項設備。學生亦感謝系辦 黃淑宜小姐，於學生求學階段給予幫助和幫忙，讓學生能順利踏上軌道。學生在此勉勵所有學生，儘管求學過程中時常碰壁，也不要放棄，有時看似高牆的難題，有時換個角度思考也許就會有突破，務必記住遇到困難時請冷靜下來整理好思緒後再付諸行動。

簡志宇 謹致於

國立屏東科技大學 資訊管理系

中華民國 109 年 6 月

目錄

摘要	i
Abstract.....	iii
謝誌	v
目錄	vi
圖目錄	viii
表目錄	x
第壹章 緒論	1
第一節、 研究背景與動機	1
第二節、 研究範圍與限制	4
第三節、 論文架構	5
第貳章 文獻探討	6
第一節、 Narrow Band Imaging (NBI) 技術	6
第二節、 深度學習應用於內視鏡檢測症狀	7
一、 CoLD (Colorectal Lesions Detector).....	8
二、 WM-DOVA (Window Median Depth of Valleys Accumul-ation)....	9
三、 A Real-Time Image Recognition System for M-NBI.....	10
四、 Using Deep Learning for Detecting Gastric Cancer.....	12
第三節、 資料預處理	14
第四節、 Backbone Network Selection	15
第五節、 Object Detection Network Selection	20
第六節、 Semantic Segmentation Network Selection	26
第參章 研究方法	28
第一節、 M-NBI 胃癌檢測研究流程	29
第二節、 研究方法概念	31
一、 使用影像預處理產生更多樣本	31
二、 EfficientDet Selection.....	32
三、 修改 DenseNet 架構	33

四、	修改 FCN 架構	34
五、	Loss Function	35
六、	M-NBI 癌症區域檢測訓練步驟	36
第三節、	評估指標	37
七、	Specificity	37
八、	Sensitivity (Recall)	38
九、	F1-score	38
十、	Accuracy	38
十一、	Average Precision (AP)	38
第四節、	M-NBI 胃癌區域檢測之測試流程 (圖 3-5)	39
第肆章	實驗成果	41
第一節、	實驗環境	41
第二節、	S1 階段 Object Detection 參數設定與實驗結果	42
一、	S1 階段參數設定	42
二、	S1 階段實驗成果	42
第三節、	S2 階段 Semantic Segmentation 參數設定與實驗結果	47
一、	S2 階段參數設定	47
二、	S2 階段實驗成果	47
第四節、	S1 階段與 S2 階段合併之實驗結果	49
第伍章	結論與建議	51
第陸章	參考文獻	52
作者簡介	56

圖目錄

圖 1-1、M-NBI 檢測影像之靜止圖像	2
圖 2-1、內視鏡. (左)傳統內視鏡 (右)NBI.....	6
圖 2-2、M-NBI 圖像 (左) 經醫師診斷無癌症 M-NBI 圖像(中) 經醫師診斷有癌症 M-NBI 圖像 (右)黃色虛線區域為醫生親自標註之癌症區域 (Ground Truth) 遮罩圖	7
圖 2-3、DWT 取輪廓特徵之示意圖 (左) 原始圖像 (右) 特徵輪廓.....	8
圖 2-4、MLP 網路架構圖	9
圖 2-5、息肉遮罩圖. (a) 息肉影像圖 (b) 息肉區域遮罩圖	10
圖 2-6、Energy map method. 運用顏色方塊繪出 polyp 區域.....	10
圖 2-7、用於分析 M-NBI 圖像架構圖在框架 Bag-of-Features 中使用密集採樣 SIFT，然後使用 SVM 進行分類	11
圖 2-8、M-NBI 即時檢測影像. (A)非腫瘤病變檢測結果 (B)腫瘤病變檢測結果	11
圖 2-9、NBI 胃癌檢測影像, (左) 距離較遠時之檢測結果 (右)距離較近時之檢測結果	12
圖 2-10、Random Erasing 資料預處理	14
圖 2-11、EfficientNet 所提出之 Compound Model Scaling	16
圖 2-12、ResNet. Residual block.....	18
圖 2-13、DenseBlock. Dense Connection	18
圖 2-14、BottleNeck Layer.....	19
圖 2-15、DenseNet 與 ResNet 參數和每秒運算速度比較.....	20
圖 2-16、Object Detection 示意圖	20
圖 2-17、Selective Search 示意圖.....	22
圖 2-18、Anchor Box 示意圖	22

圖 2-19、EfficientDet 框架	23
圖 2-20、FPN 架構.....	24
圖 2-21、BiFPN Layer.....	24
圖 2-22、M2Det 網路架構.....	25
圖 2-23、AlexNet 網路架構.....	27
圖 2-24、FCN 與傳統卷積神經網路之差異. FCN 將原先進行全連接層時之動作改 為卷積層，且 FCN 能不受到影像大小進行訓練	27
圖 3-1、M-NBI 胃癌檢測之神經網路模型訓練流程圖	30
圖 3-2、特殊影像預處理示意圖	32
圖 3-3、Lightweight DenseNet, (上)修改後 (下)修改前.....	33
圖 3-4、FCN Lightweight DenseNet	34
圖 3-5、S1 階段 (上) 與 S2 (下) 階段測試流程圖	40
圖 4-1、多框標記(左) 與 單框標記 (右) 示意圖	43
圖 4-2、多框標記 (左) 與單框標記 (右)方法檢測結果，白色背景為癌症遮罩區域	44
圖 4-3、個別 Max Detection Box 數量比較圖	45
圖 5-1、Single Stage Detection Network.....	52

表目錄

表 1-1、歷年因胃癌死亡人數(男女加總)及增減數	1
表 1-2、醫師使用傳統 CADe 檢測與透過 Deep Learning 檢測之比較	3
表 2-1、基於 ImageNet dataset 對 EfficientNet 系列個別最佳化之參數調整結果圖	16
表 2-2、DenseNet 網路架構	19
表 3-1、Object Detection 與 Semantic Segmentation 速度比較	28
表 3-2、評估指標之運算單位	37
表 4-1、實驗環境	41
表 4-2、實驗參數設定	42
表 4-3、個別 Object Detection Network 效能比對圖	45
表 4-4、取最大值座標與平均值座標之檢測結果	46
表 4-5、FCN 實驗參數設定	47
表 4-6、有無做 Data Augmentation 效能比對	48
表 4-7、Semantic Segmentation Network 與 Two Stage Detection Network 效能評估	50

第壹章 緒論

第一節、研究背景與動機

癌症一直是困擾著國民之前幾名惡性腫瘤，癌症在歷年因疾病死亡原因報告上時常榜上有名，但在現今醫學技術下，如果能提早發現初期症狀之話，且依照醫師之專業治療是能夠大幅度之對症下藥降低死亡率 (表 1-1)，其中最常運用到之檢測技術是窄頻影像系統 (Narrow band imaging, NBI) (圖 1-1)，運用高對比之光頻來區分血管與表面黏膜，能加速醫生在行檢測時之速度與提升精確度。

多虧硬體之效能已經達到能觸及深度學習所需要之門檻，同時一些深度學習之台柱發現可以將這門技術運用到醫療領域上，透過深度學習高速之運算速度和學習能力，不僅可以降低傳統人力上之成本，也可以在檢測上得到較快且即時之診斷。

表 1-1、歷年因胃癌死亡人數(男女加總)及增減數

	103 年	104 年	105 年	106 年	107 年
死亡人數(人)	2,350	2,326	2,315	2,304	2,299
增減數(%)	-	-1.02%	-0.47%	-0.47%	-0.21%

資料來源：本論文參考衛生署資料製作

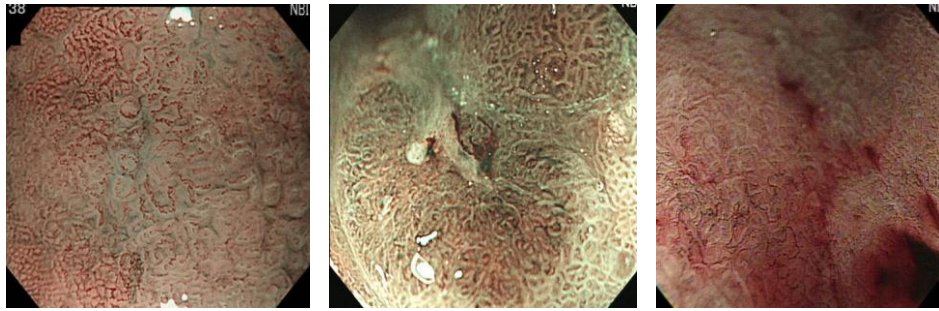


圖 1-1、M-NBI 檢測影像之靜止圖像

另一方面通過內視鏡檢查診斷胃腸道癌之需求正在增加，但是並非所有醫生皆能第一時間使用內視鏡找出患者病變區域（尤其是早期癌症），醫生會使用內視鏡伸入患者體內進行光學檢測，如需要進行腫瘤切除之話也可以利用內視鏡前端之鋼絲刀進行切除，但胃癌內視鏡檢查之需求在增加，而且並不是所有之醫生都能立即確定患者症狀之狀況，尤其是早期癌症症狀較難發現，以及一些假陽性之黏膜狀態容易誤判，因此需要進行多次檢查才能找出病因，過程也會讓患者在檢測過程中有一些不適和精神疲勞感，雖然現今有無痛內視鏡作法減輕患者在進行檢測時之不適感，主要作法是透過靜脈點滴方式施打少量麻藥來達到局部麻醉效果，但是依然無法解決醫生在一天內進行多次反覆之操作檢測儀器可能會因為疲勞而導致判斷錯誤之問題，例如：醫師在進行檢測時，必須同時操作控制器微調內視鏡和盯著檢測螢幕，由於內視鏡顯示出來之畫面速度很快，在全神貫注之情況下很容易造成眼球疲

勞，導致檢測時可能漏掉有異常之區域，如果能在醫生檢測時即時在檢測畫面上提醒醫師該區域有異常，此時醫師可以慢下來進行更詳盡之檢查。

綜上所述，通過使用電腦輔助檢測 (Computer Aided Detection, CADe) 可以有效輔助醫生進行檢測；另外，使用深度學習建構的檢測系統會比傳統電腦輔助檢測方法之精確度 (Accuracy)、敏感性 (Sensitivity) 和特異度 (Specificity) 性能提升 5% 至 7% [1]-[8] (表 1-2)，基於深度學習之優勢，本論文提出基於深度學習的 M-NBI 胃癌區域檢測，來輔助醫師在進行內視鏡檢測時能即時得知當前檢測區域的狀況，並且有癌症區域一定要檢測出來，根據上述目的本論文利用現有深度學習網路架構進行訓練與改良，建構速度夠快且靈敏性 (Sensitivity) 夠高之檢測網路。

表 1-2、醫師使用傳統 CADe 檢測與透過 Deep Learning 檢測之比較[5]

	F1-score	Specificity	Sensitivity
Physician diagnosis (average)	38.7%	33%	44.2%
Deep Learning	43.5%	38.7%	48.1%

第二節、研究範圍與限制

本研究在有限時間與有限設備限制下，僅著重於解決如何在小樣本資料庫下訓練出高性能之網路，以及受到當前設備性能限制僅通過能夠訓練之現有網路框架進行應用和修改。從論文開始至論文結束，所利用之資料集皆是與計畫合作者李宗鏞醫生所提供之 M-NBI 胃癌圖像和親手標記之癌症區域遮罩圖(Ground Truth) 作為訓練圖像，非本論文提出之深度學習網路框架均不參與本研究範圍內。



第三節、論文架構

本論文主要分為四個章節，第一章為緒論，第二章為文獻探討，第三章為研究方法，第四章為實驗成果。各章節簡單概述如下：

第一章 緒論：主要著重介紹現階段癌症醫療檢測 M-NBI 技術之趨勢和硬體設備之成長為深度學習帶來更多討論度，以及深度學習在輔助檢測上之重要性，並且講述本論文之研究動機與目的。

第二章 文獻探討：收集相關論文研究和文獻包括各種深度學習應用在醫療檢測上之方法，以及整理部分涵蓋與本論文研究有相關性之深度學習網路，通過對這些方法之分析與探討後加以解釋。

第三章 研究方法：該章節主要闡述如何運用深度學習網路對 M-NBI 圖像進行檢測，以及整個框架模型之解釋和現階段研究成果。

第四章 研究成果：展示本研究實驗環境和個別檢測網路之實驗成果、模型參數調整以及效能分析。

第五章 結論：本章節會描述本論文之貢獻與未來之研究方向。

第貳章 文獻探討

第一節、Narrow Band Imaging (NBI) 技術

傳統之內視鏡主要使用氬氣當作光源 (圖 2-1)，用肉眼方式觀察白光反射之區域，但用這種方法無法看到較深層組織狀況，而 NBI 主要是用一個特殊濾光盤，過濾出兩種窄頻藍光 (415 nm) 及綠光 (540 nm)，並捨棄紅光之成分，藍色光譜波長將表層黏膜及表面微小之血管呈現藍色圖像，而綠色光譜波可以將深層血管呈現綠色圖像，可以使得血管結構與周圍組織形成明顯之對比，有利於辨別表層黏膜組織病變，目前國內醫療設備上會搭載切換開關來使用傳統內視鏡光譜和 NBI 光譜來為病人進行檢測。

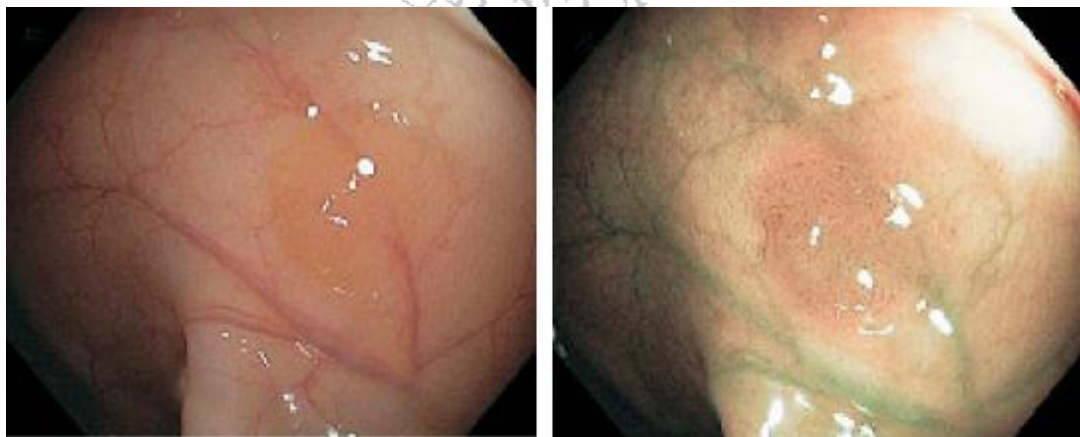


圖 2-1、內視鏡.(左)傳統內視鏡 (右)NBI

資料來源：<http://www.doctor.org.tw/file/66-4-1.pdf>

Magnifying Narrow Band Imaging (M-NBI) 是基於 NBI 技術下進行改良，在內視鏡前端裝上能放大 70 倍之鏡頭 (圖 2-2)，使其能更清楚看見表面黏膜或微血管變化，輔助醫師尋找較細微之病變特徵，這項優點也能應用於深度學習，透過 M-NBI 能學習到更精準之癌症特徵，減少誤判假陽性或假陰性之可能性。

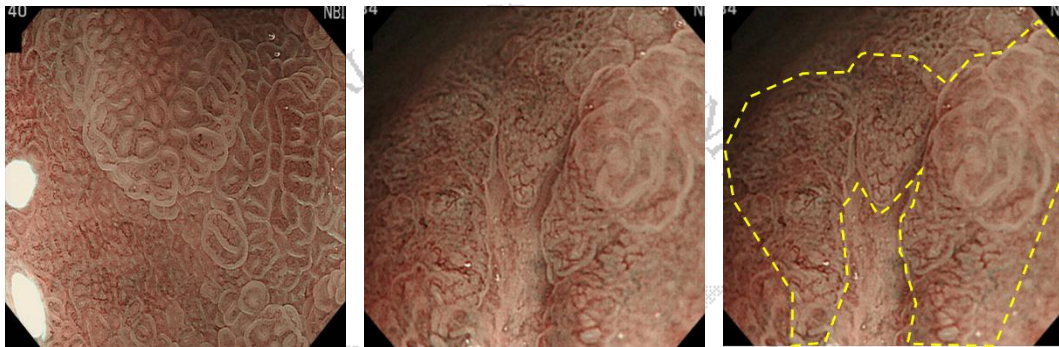


圖 2-2、M-NBI 圖像 (左) 經醫師診斷無癌症 M-NBI 圖像(中) 經醫師診斷有癌症 M-NBI 圖像 (右)黃色虛線區域為醫生親自標註之癌症區域 (Ground Truth) 遮罩圖

第二節、深度學習應用於內視鏡檢測症狀

性能良好之 AI 檢測工具需要包含以下幾種優點 [9]: (1) 檢測病變區域之 Sensitivity 須很高，不可漏掉有癌症區域 (2) False Negative 須很低 (3) 檢測時間需要夠短，現階段為了滿足以上幾項條件，皆是利用截取內視鏡影片中之靜止圖像進行訓練，同時對圖像進行預處理以及有效之利用病變特徵對圖像進行 De-noise 或進行特徵融合。

早期常用之技術像是使用區域增長，一種基於像素之圖像分割之方法，透過提取內腔輪廓並檢測與對照消化道之病理；近些年之方法例如：

一、 CoLD (Colorectal Lesions Detector)

Maroulis [10] 等人以 MATLAB 環境進行開發，利用結腸之黏膜特徵為結腸之表面特徵提供更多訊息，從而將其用於正常或異常之相應圖像進行分類，主要方法是利用每張內視鏡之靜止圖像以 Discrete Wavelet Transform (DWT) (圖 2-3) 計算出二階統計特徵，在以 Multilayer Perceptron neural networks (MLP) 區分正常或異常組織之區域 (圖 2-4)，Accuracy 可達 90%。

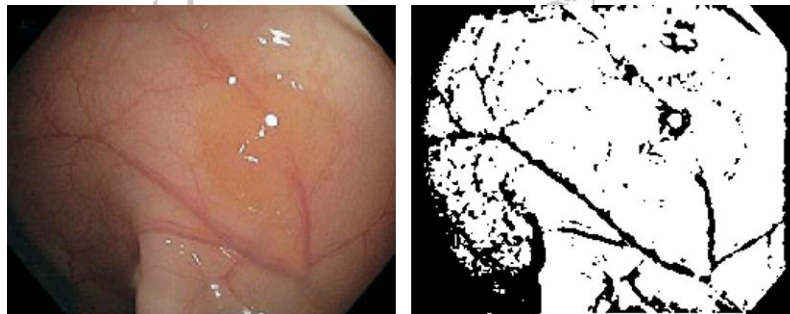


圖 2-3、DWT 取輪廓特徵之示意圖 (左) 原始圖像 (右) 特徵輪廓

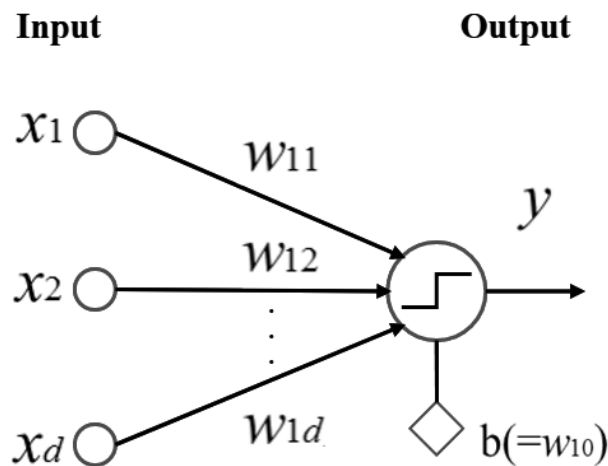


圖 2-4、MLP 網路架構圖

二、 WM-DOVA (Window Median Depth of Valleys Accumulation)

Bernal [11] 等人運用大量息肉內視鏡圖像與手動定義遮罩圖 (圖 2-5) 來建立 Ground Truth 之遮罩影像，進行訓練以達到輔助檢測之目的，基於息肉特徵以及正常區域與息肉邊界差異性創建 Energy Map。主要作法是基於息肉之外觀模型，透過 Model 使用 Valley Information 來表示息肉邊界，在以 Energy Map 定義好之色塊 (圖 2-6) 對異常息肉區域進行 Semantic Segmentation，接著計算遮罩圖和息肉圖之位置精確度，在 Accuracy 上可達 72.33%。

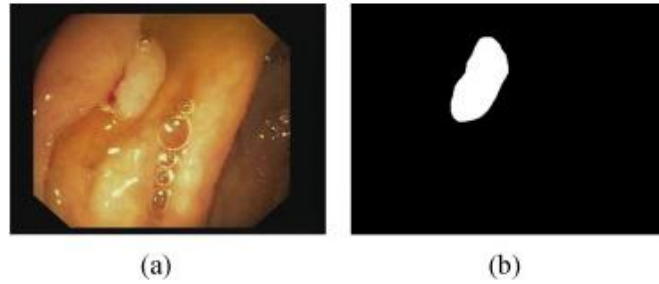


圖 2-5、息肉遮罩圖. (a) 息肉影像圖 (b) 息肉區域遮罩圖 [11]

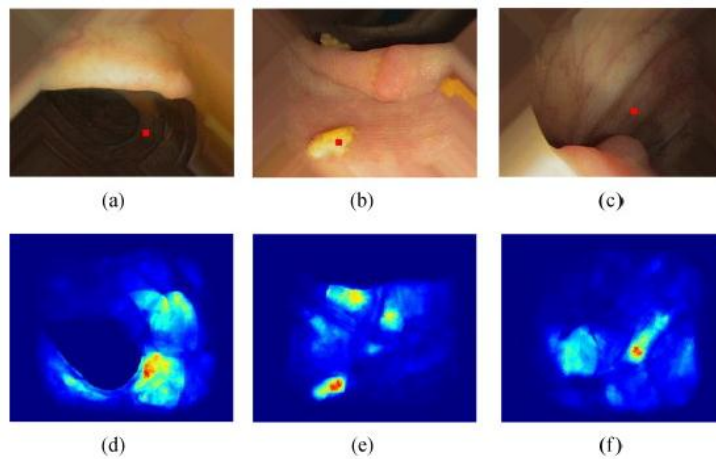


圖 2-6、Energy map method. 運用顏色方塊繪出 polyp 區域
(詳細顏色引用請參閱論文) [11]

三、 A Real-Time Image Recognition System for M-NBI

Kominami [12] 等人基於 Bag-of-Features 與 SVM (圖 2-7) 方法訓練檢測網路用於 M-NBI 即時檢測影像，通過擷取息肉影像特徵，並進行分類用於製作 Bag-of-Features，根據 Bag-of-Features 將影像表示成直方圖，接著使用 SVM 進行分類，能以極短之延遲輸出 2 級診斷，判斷是否有異常息肉於即時影像中 (圖 2-8)。

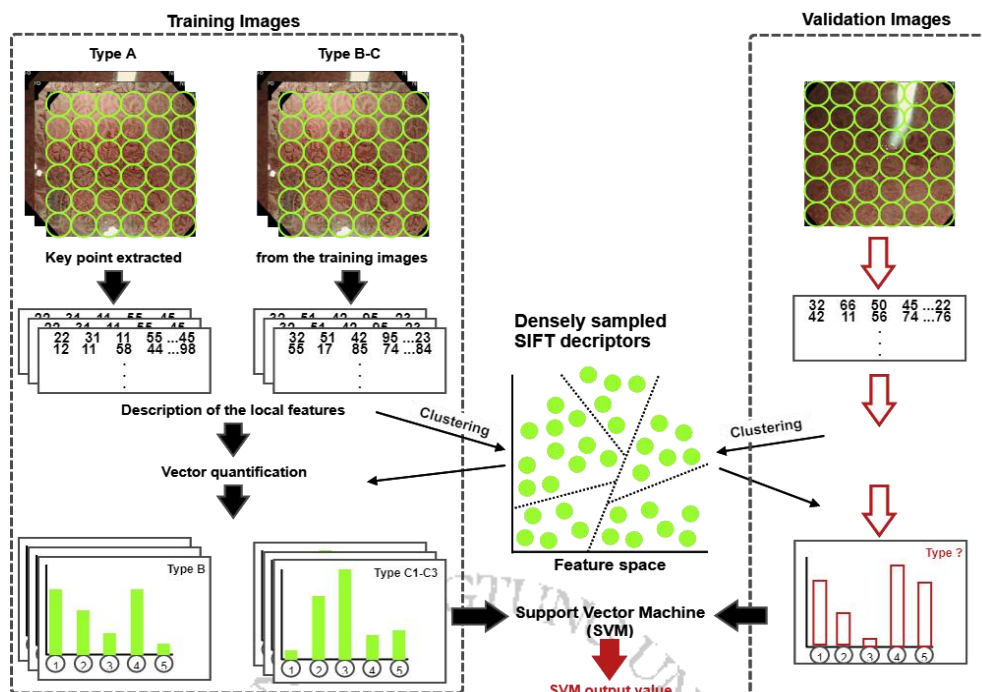


圖 2-7、用於分析 M-NBI 圖像架構圖在框架 Bag-of-Features 中使用密集採樣 SIFT，然後使用 SVM 進行分類 [12]

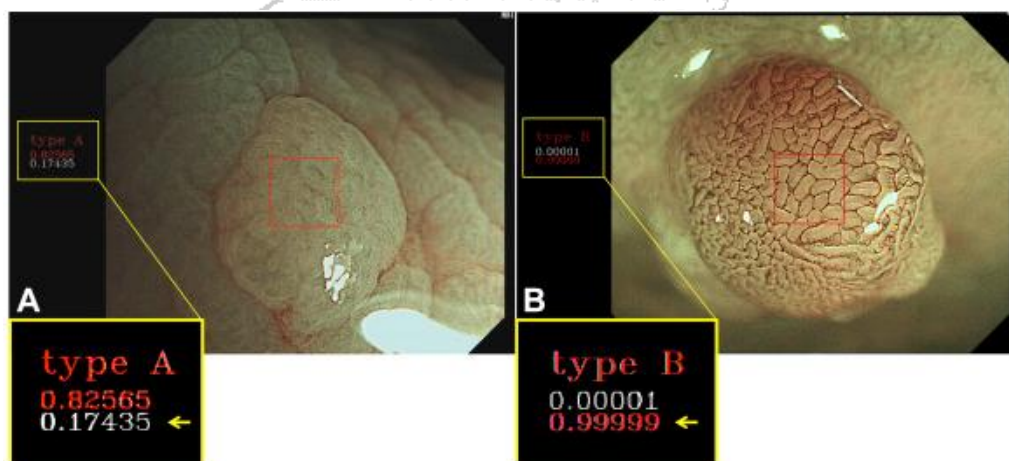


圖 2-8、M-NBI 即時檢測影像. (A)非腫瘤病變檢測結果 (B)腫瘤病變檢測結果 [12]

四、 Using Deep Learning for Detecting Gastric Cancer

關於使用 Deep Learning 檢測胃癌與使用 M-NBI 內視鏡之論文並沒有很多，因此本論文主要以較符合本論文研究方向之深度學習檢測方法做為研究參考，在論文中提到之深度學習專有名詞會在後續章節著重介紹；Hirasawa [13] 等人使用深度學習之 Object Detection 進行即時檢測，利用 Single Shot Multibox Detector (SSD [14]) 之 Object Detection 網路進行胃癌區域訓練，胃癌資料集來源為自家機構親自收集，資料集樣本篩選條件主要以清晰、表面無氣泡和無反光作為挑選基準，總資料集共 15880 張，在實驗成果上 Sensitivity 達到 92.6%、True Positive 達到整體之 30.6%，測試過程如圖 2-9 (綠色方框為親手標記有癌症區域之 Ground Truth，黃色方框為 Predicated Result) 所示，檢測速度上可達每 47 秒讀取 2000 張。

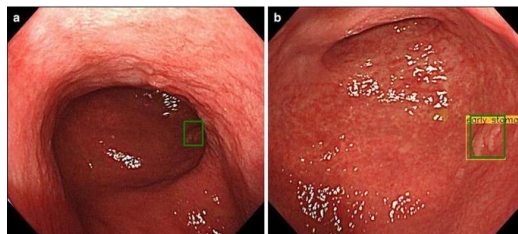


圖 2-9、NBI 胃癌檢測影像, (左) 距離較遠時之檢測結果 (右) 距離較近時之檢測結果 [13]

綜上所述，現階段 CAdE 檢測技術已經能輔助醫師快速尋找病變區域，讓經歷較淺之醫師也能快速掌握病患身體狀況，例如使用 Semantic Segmentation 和 Object Detection 之方法，但部分機器十分昂貴，檢測儀器也不可隨意移動；另外，在 Deep Learning 方法上專攻 M-NBI 內視鏡與胃癌之目標檢測論文也很少，胃癌檢測研究方向主要是以傳統機械學習之方法進行再改良，而 M-NBI 內視鏡方面之病變檢測研究方向主要以 CNN 網路進行區分病症種類，加上想要製作 Deep Learning 需要擁有大量病理樣本，同時醫療影像很難取得，樣本過少會導致檢測結果之精確度極低，這導致本論文在參考相關文獻時，無法直接從成網路進行改良與訓練，必須先著重解決樣本過少之問題後，才能進到下一步挑選適合之網路。

第三節、資料預處理

在深度學習中，為了避免出現 dataset 資料太少導致 Overfitting，通常會進行一些前處理把現有之資料進行幾何之變化，以增加 dataset 之數量，常見之方法是將一張圖片進行旋轉、放大、切割或遮蓋，例如：使用 Radom Erasing Data Augmentation (圖 2-10 [15]) 之方法，通過對圖像進行縮放、部分裁切，搭配固定幾種圖形進行遮罩，最後再將已做初步處理的圖像進行融合遮罩，獲得更多樣本。

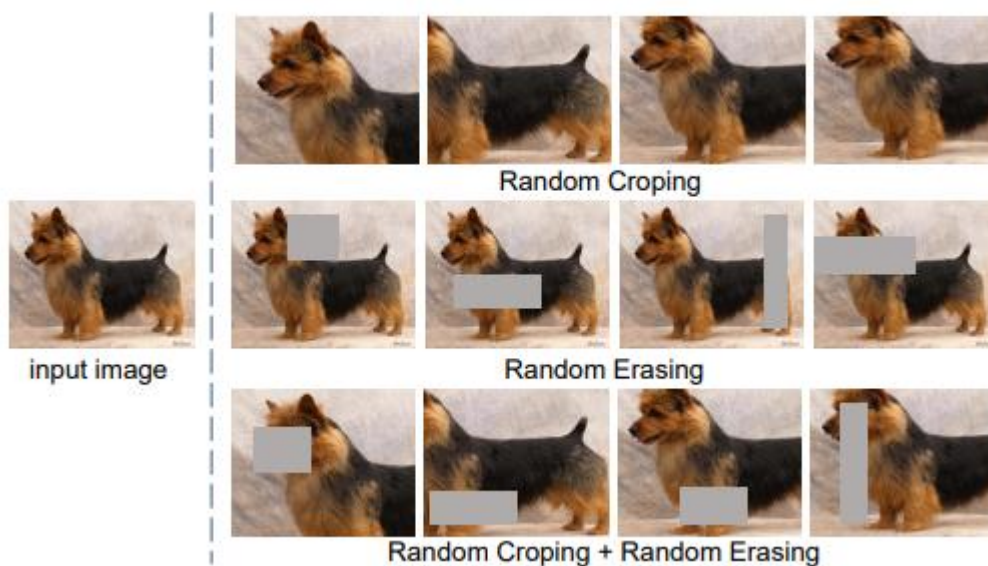


圖 2-10、Random Erasing 資料預處理 [15]

但僅使用這類方法進行資料預處理，能產生之額外樣本其實很有限，樣本過少就無法有效發揮 Deep Learning 的優點，從而導致效能下降。

第四節、 Backbone Network Selection

本論文在 Object Detection 階段使用 EfficientNet [16] 作為檢測網路之骨幹網路，EfficientNet 是基於 NASNet [17] 之概念進行建構，主張讓網路本身自行學習，自動得出較為恰當之參數，減少在訓練階段反覆調整參數之時間成本，另外，該作者提出了 Compound Model Scaling 之算法作為 Efficient Net 之主要網路框架架構，如附圖 (圖 2-11) 中所繪：a) 為常見網路的架構； b) 為對網路進行 channel 之增加； c) 是對網路層數進行加厚； d) 是對輸入之圖像大小進行放大。作者認為對常見網路做加寬、加深或增加圖像解析度不應該是互相獨立，將常見幾個作法因不同任務所需求綜合優化使用，能比對單一維度（寬度、深度、解析度）進行調整更有效提高效能，同時並免過多之參數量，示意圖如附圖 e) 所示，基於 Compound Model Scaling 之算法，作者開發一系列不同尺度之 EfficientNet (表 2-1)，每一個網路皆通過前一個最佳參數所建構而成，例如：在 EfficientNet 最初始之 Baseline 中 (EfficientNet-B0)，經過反覆測試取得最佳參數後，再將其代入 EfficientNet-B0 後，將其命名為 EfficientNet-B1，以此類推最終得到 EfficientNet-B1 到 EfficientNet-B7 之網路，其效能依照網路編號由小至大逐漸變慢，但在其精確度上有很好的提升。

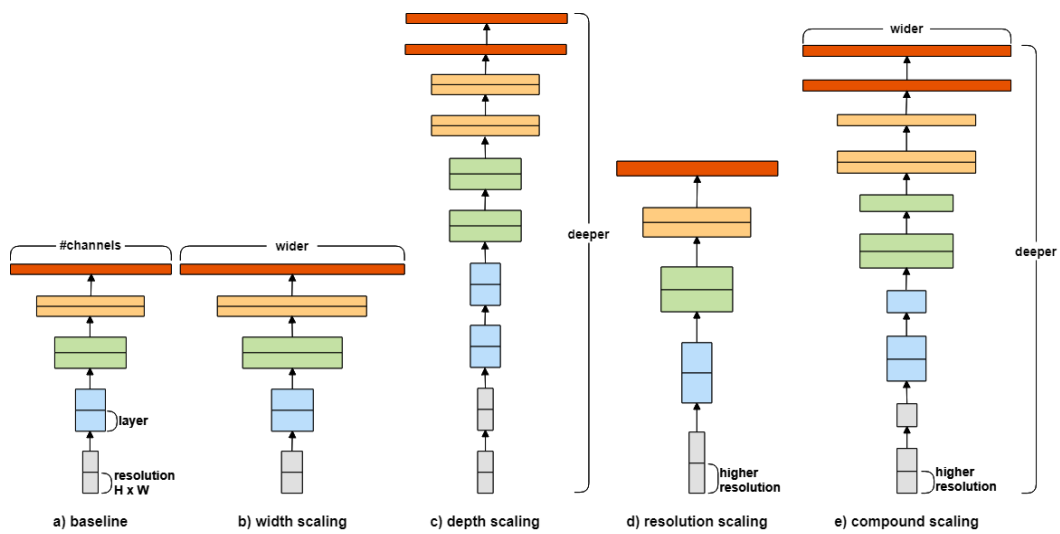


圖 2-11、EfficientNet 所提出之 Compound Model Scaling

表 2-1、基於 ImageNet dataset 對 EfficientNet 系列個別最佳化之參數調整結果圖 [16]

Model	Top-1 Acc	Top-5 Acc	Params	FLOPS
EfficientNet-B0	76.3%	93.2%	5.3M	0.39 B
EfficientNet-B1	78.8%	94.4%	7.8 M	0.70 B
EfficientNet-B2	79.8%	94.9%	9.2 M	1.0 B
EfficientNet-B3	81.1%	95.5%	12 M	1.8 B
EfficientNet-B4	82.6%	96.3%	19 M	4.2 B
EfficientNet-B5	83.3%	96.7%	30 M	9.9 B
EfficientNet-B6	84.0%	96.9%	43 M	19 B
EfficientNet-B7	84.4%	97.1%	66 M	37 B

本論文以目前主流之網路 ResNet [18] 與 DenseNet [19] 為參考依據，目的是用於 Semantic Segmentation 階段之骨幹網路使用。網路架構 DenseNet 與 ResNet 基本上雷同；ResNet 之概念是使用 Residual block (圖 2-12) 將先前訓練之數值加上後在送到下一層，類似將 channel 本身進行累加，缺點是產生之參數太多，用於小資料集網路容易發生 overfitting；DenseNet 之概念是把每一層之 channel 之部分之 Feature Map 進行累加串聯形成 Dense Connection 之形式，類似將 channel 數量進行值之疊加，使 channel 的長度變長，這特點讓 DenseNet 在參數與計算成本更少之情況下能實現比 ResNet 更優之效能，此優點可以有效讓 Semantic Segmentation 階段之網路加速，達到快速檢測之目的。

DenseBlock (圖 2-13) 為 DenseNet 之基本單位，在 DenseBlock (表 2-2) 中，各層之 Feature Map 大小一致，所以能直接進行 Concatenate，但是這樣後面層之 channel 會非常大，所以在原有 DenseBlock 框架中加入 bottleneck layer 來減少參數，主要做法是：在每一層 DenseBlock 中加入 1×1 之 Convolution Kernel 達到參數減少之目的，從而提升計算效率，範例如圖 (圖 2-14) 所示。層與層之間還有一個 Transition Layer 進行連接，能進一步之將 channel 數量壓低，

主要作法是：假設 channel 數為 m ，Transition Layer 可以產生 $\lceil \theta m \rceil$ 個 channel，其中 $\theta \in (0,1)$ 是 Compression Rate，範圍為 0 至 1 之間，當 $\theta = 1$ 時，經過 Transition Layer 的 channel 數量為 $1 \times m$ ，故整體 channel 不變；當 $\theta = 0.5$ 時，經過 Transition Layer 的 channel 為 $0.5 \times m$ ，故整體 channel 減少為原先之 0.5 倍，在效能與精確度上 DenseNet 已經超越了 ResNet (圖 2-15)。

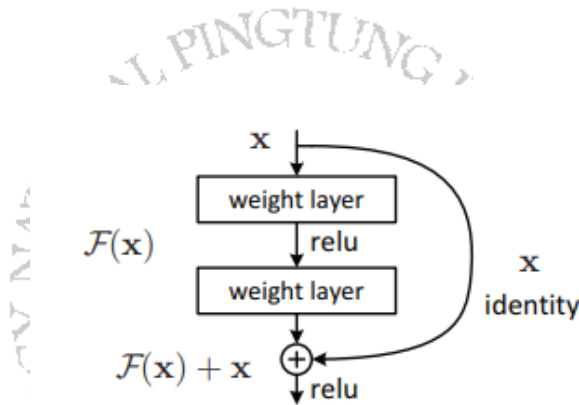


圖 2-12、ResNet. Residual block [18]

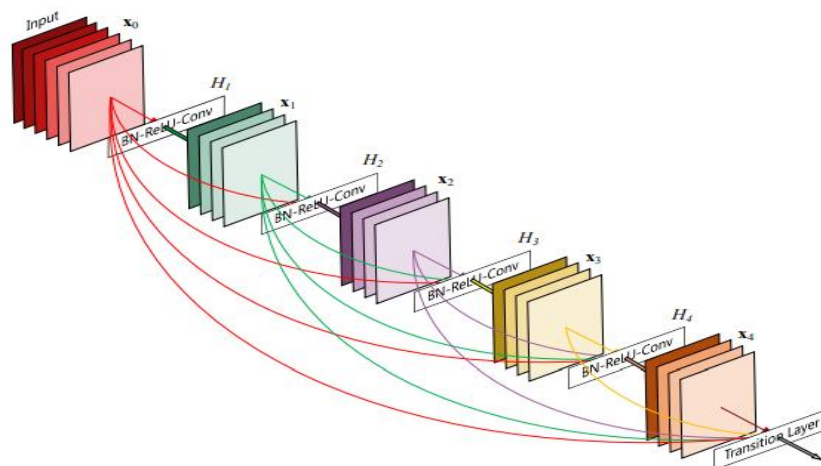


圖 2-13、DenseBlock. Dense Connection [19]

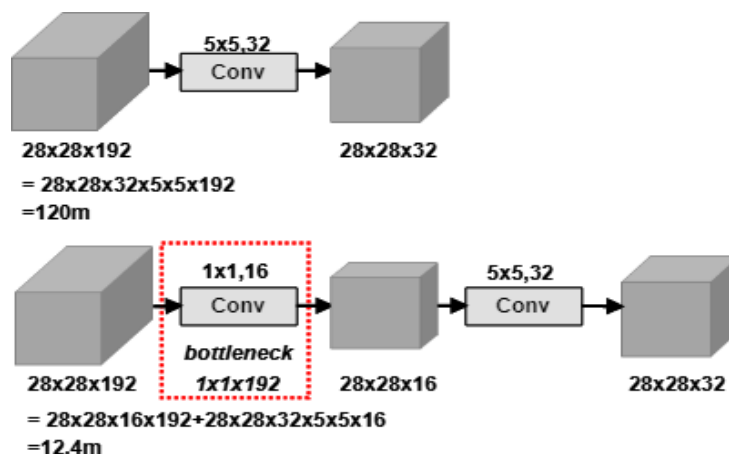


圖 2-14、BottleNeck Layer

表 2-2、DenseNet 網路架構 [19]

Layers	Output Size	DenseNet121
Convolution	112×112	7×7 conv, stride 2
Pooling	56×56	3×3 max pool, stride 2
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layers (1)	56×56	1×1 conv
	28×28	2×2 average pool, stride 2
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layers(2)	28×28	1×1 conv
	14×14	2×2 average pool, stride 2
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$
Transition Layers (3)	14×14	1×1 conv
	7×7	2×2 average pool, stride 2
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$
Classification Layer	1×1	7×7 global average pool
	-	1000D fully-connected, softmax

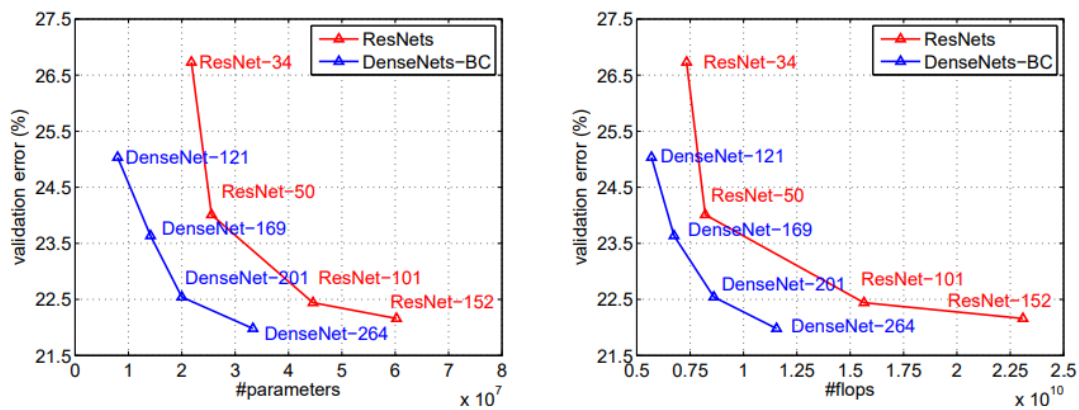


圖 2-15、DenseNet 與 ResNet 參數和每秒運算速度比較 [19]

第五節、Object Detection Network Selection

Object Detection 網路與傳統之 Convolutional Neural Network (CNN) 框架不太一樣，Object Detection 網路具有分類的功能也有將分類目標之位置框選出來之特點，通常是一個四邊之幾何圖形 (圖 2-15)。



圖 2-16、Object Detection 示意圖

為了加快 Sematic Segmentation 階段之檢測速度，在使用 Sematic

Segmentation 前運用 Object Detection 之網路來找出癌症大概區域，並將該區域座標訊息保存至 Sematic Segmentation 階段進行圖像切割使用，運用已初步檢測出來的癌症位置進行 Sematic Segmentation 能有效減少網路檢測整張圖像之時間。

Object Detection Network 又分為 One-stage 和 Two-stage 兩種不同方法之網路框架。傳統上 Two-stage 是在網路進行辨識物件前，先對圖像或特徵圖中紋理或顏色較相近之物件進行 Selective Search (圖 2-16) 產生候選框，然後再對選出來的物件進行物件辨識，由於預先框選出來之候選框非常多，在檢測速度上會有很大的影響，但大量之候選框對影像的物件分類與檢測精確度上會比較好之效能，例如：Faster R-CNN Network [20]、R-CNN Network [21]，。不同於 Two-Stage，One-Stage 對物件之分類與辨識能同時進行，捨棄 Selective Search 這類方法改用 Anchor 機制來進行檢測，通過在 Feature map 上均分成數個格子，然後產生數個固定大小之 box 作為 Anchor Box (圖 2-17)，在針對不同大小之物件下使用不同大小之 Anchor box 進行目標檢測，例如：YOLO [22]、Single Shot Detector (SSD)，雖然速度上比 Two-Stage 快上不少，但對物件位置之精確度上沒有 Two-Stage 來的好，但在整體辨識度上仍在可接受範圍內。

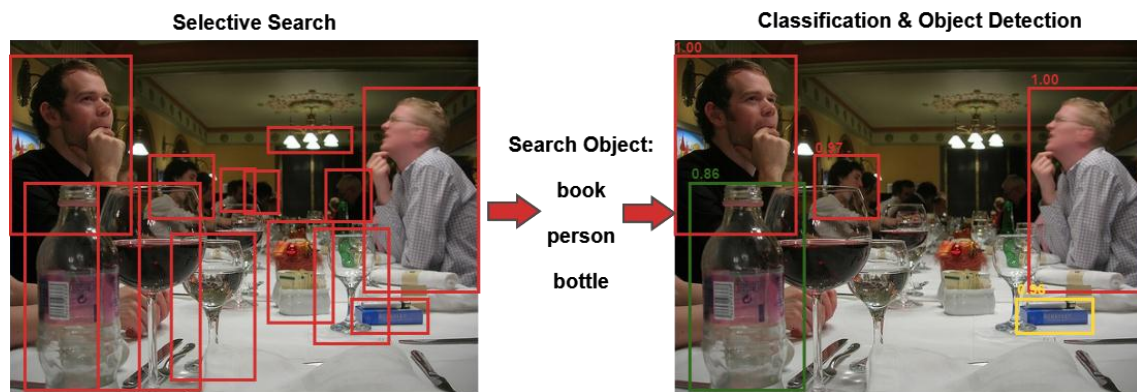


圖 2-17、Selective Search 示意圖

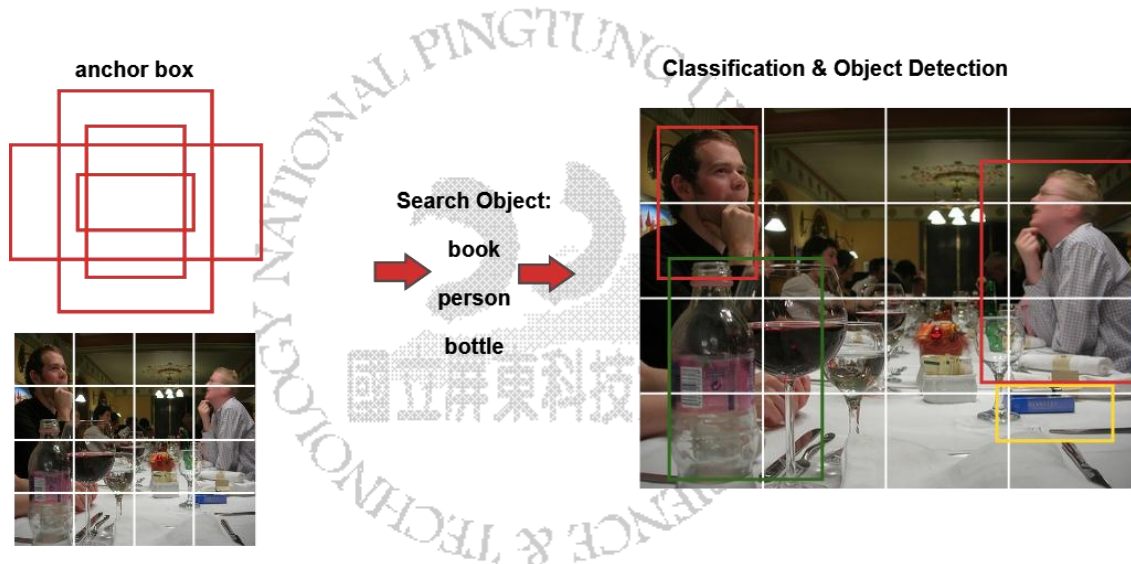


圖 2-18、Anchor Box 示意圖

EfficientDet [23] (圖 2-18)是基於 EfficientNet 所建構而成之 One-stage 檢測網路，骨幹網路使用了 EfficeintNet，從檢測速度快至慢，以及檢測精確度低至高之下又分成 EfficientDet-D1 至 EfficientDet-D7 七種分別對應 EfficientNet 因應最佳化參數調整所建構而成之七種網路。EfficientDet 採用特徵融合 Feature Pyramid Network (FPN) 作為主

要方法，其概念為將深層之特徵與淺層之特徵進行相加融合，針對此概念其他研究後也有發展出如：Path Aggregation Network (PANet) [24] 、NAS-FPN [25] 類似概念之架構 (圖 2-19)，作者為了使 EfficientDet 之特徵融合階段能運用到 EfficientNet 之 NAS 概念，專為其設計 BiFPN 框架作為 EfficientDet 主要特徵融合之方法，作者發現 PANet 之效果比 FPN、NAS-FPN 要來的好，但是該方法因連接所有結點導致計算量龐大，於是基於 PANet 框架進行改良，使用了 Skip-Connection 之概念，設計了 BiFPN (圖 2-20)。

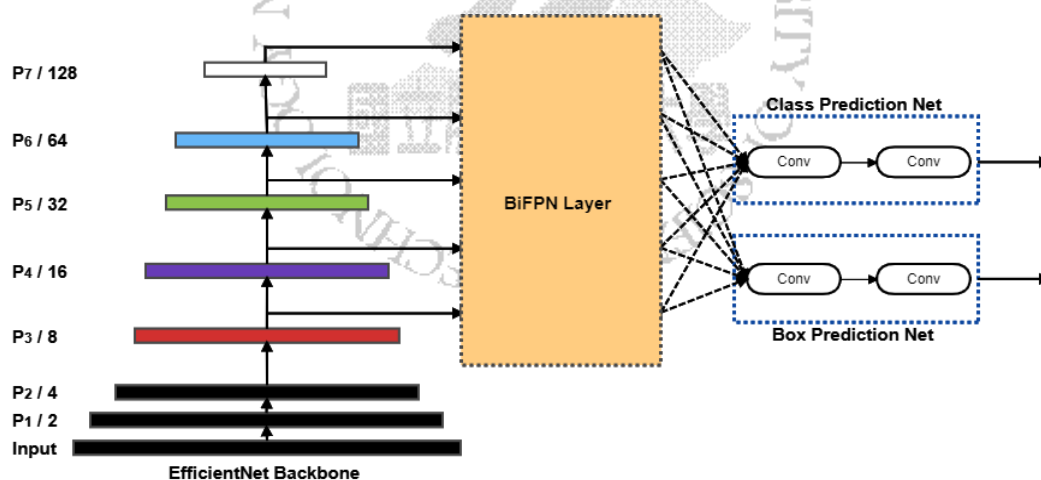


圖 2-19、EfficientDet 框架

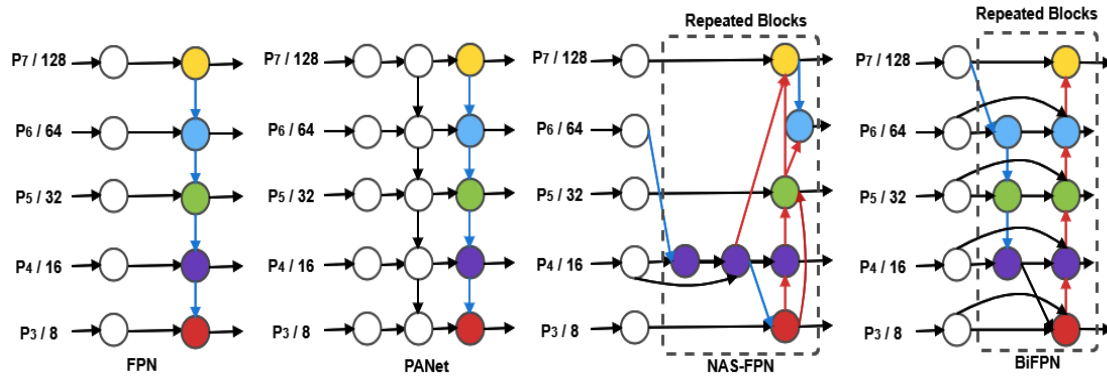


圖 2-20、FPN 架構

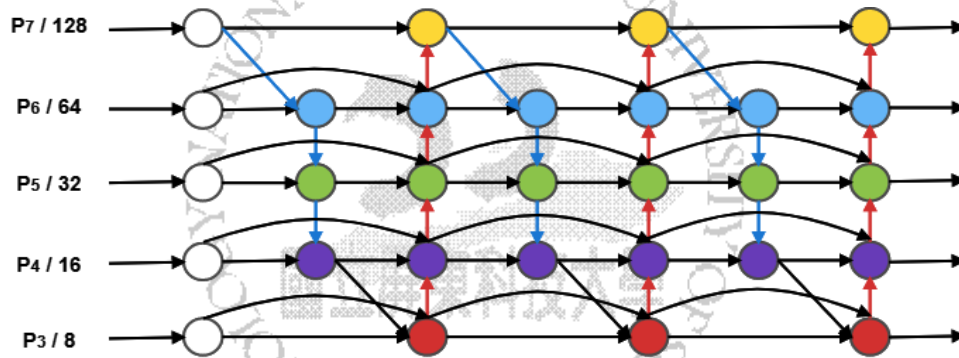


圖 2-21、BiFPN Layer

M2DetNet[26] 是基於 SSD 網路之框架與該論文提出之 MLFPN 特徵融合技巧建構而成，M2DetNet 能有效解決 Object Detection 在尺度變化中帶來之一些問題，該論文主張使用高階之特徵來解決分類之問題，以及使用低階之特徵用於解決物件識別；MLFPN 主要由三個模塊組合（圖 2-21）：(1) Feature Fusion Module (FFM)、(2) Thinned U-shape Module (TUM)、(3) Scale-wise Feature

Aggregation Module (SFAM)；FFM 之功能主要是將不同層之間之特徵矩陣進行合併，由於 FFM 之特徵矩陣是由來自前一層 Backbone Net 不同層所組成，所以在進行 FFM 特徵矩陣合併時，再送到 TUM 區塊時能同時學到高階與低階之特徵；在每一個 TUM 區塊中，皆是由縮減版之 U-net 所建構而成，其網路層數會比 FPN 要來的小，同時 TUM 會輸出輸出不同尺度之特徵，下一個 TUM 可以從上一個 TUM 學習到更為複雜之特徵再繼續進行學習，在取得所有層級 TUM 之不同尺度特徵後，會進行特徵融合以構成 Multi-level Feature Pyramid，在該階段稱為 SFAM。

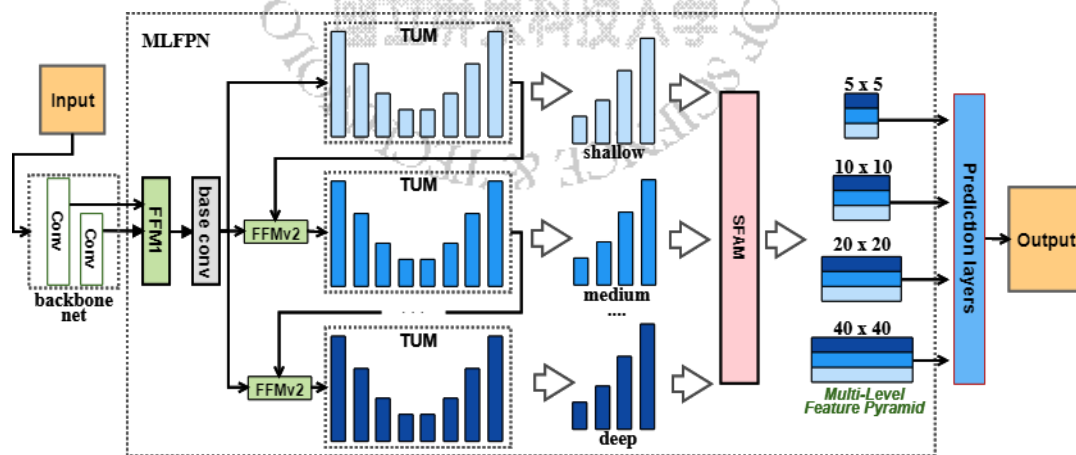


圖 2-22、M2Det 網路架構

第六節、 Semantic Segmentation Network Selection

為了能讓學習完癌症特徵之網路標示出癌症區域，故需要使用別種方法讓預測癌症位置能顯示至 M-NBI 上，這邊以能做 Semantic Segmentation 之 Fully Convolution Network (FCN) [27] 作為主要網路。

通常 Convolution Neural Network 在最後之 Convolution Layer 上進行 Flattening 後，將原先產生之 Convolution Layer 二維矩陣逐一轉為一維矩陣，使其變成一個固定長度之特徵向量。例如：AlexNet [28] 在最後一層 Convolution Layer 上連接 Fully Connected Layer (圖 2-22)，輸出是一個 1000 維之向量，用來計算每一類影像之概率。Fully Connected Layer 與 Convolution Layer 最大之差異就在資料之連接方式與儲存之資訊，Fully Connected Layer 僅會保留對特徵之結果訊息，並且拉成一維向量後，連接至基本之神經網路進行分類；Convolution Layer 則會保有原有的特徵訊息。綜上所述，是可以將這兩類進行互相轉化，而 FCN 就是用這個方法將 Fully Connected Layer 所丟棄之資訊保留起來 (圖 2-23) 並且經過 Deconvolution Layer 後還原回原圖大小，所以 FCN 整個網路都是以 Convolution Layer 組成。

第參章 研究方法

過去使用於癌症檢測之方法，缺點有以下幾種：(1) 部分早期癌症區域無法被正確檢測 (2) 判斷精確度與檢測速度不夠有效率 (3) 無法取得大量樣本，故效能不佳。為了改善這些缺點，能夠運用深度學習之 Semantic Segmentation 網路進行檢測癌症，達到更好之檢測效能，但是 Semantic Segmentation 網路在檢測速度上也不算很快；另外，現階段以發表 Object Detection 網路之檢測速度能比 Semantic Segmentation 網路要來的快 (表 3-1)，基於 Object Detection 這項優點，本論文設計兩階段檢測網路來進行驗證 Object Detection 網路所檢測出來之癌症區域在經過輕量型 Semantic Segmentation 網路進行檢測後，該檢測結果是否與單純使用 Semantic Segmentation 網路做癌症解測之效能差不多。本章節包含研究流程、檢測步驟及常用於醫學檢測之評估指標。

表 3-1、Object Detection 與 Semantic Segmentation 速度比較 [29]

model	backbone	fps	accuracy
Object Detection Network (COCO dataset, 1280 × 1080)			
EfficientDet-D1	EfficientNet	27±5	45% (mAP)
YOLOv3	Darknet53	40±5	31% (mAP)
Semantic Segmentation Network (COCO dataset, 1280 × 1080)			
R-FCN	VGG16	5.9±5	29% (mIoU)
DeepLab v3+	Xception	8±5	85% (mIoU)

第一節、M-NBI 胃癌檢測研究流程

為了輔助醫師能在使用內視鏡時能快速檢測出癌症區域，需要將整個網路架構進行優化，以達到高速度且高準度之要求。本論文使用深度學習之 EfficientDet 與 FCN 作為主要框架網路。本論文研究方法流程如圖（圖 3-1）所示，本研究將 Objection Detection 與 Semantic Segmentation 之階段分開運行，分別為 S1 與 S2，步驟如下：

S1 階段之圖像採用無 Segmentation Map 之影像進行訓練，使用特殊之 Data Augmentation 以達到增加訓練樣本之目的，在進行訓練同時以取得預測分類和目標檢測之結果，再將預測目標座標與分類結果進行記錄保存。

保存之預測目標座標會在 S2 階段之圖像採用有 Segmentation Map 之影像進行訓練進行讀取，經過 Data Augmentation 階段時，導入 S1 儲存之影像預測目標座標與分類結果，將輸入影像對應測目標座標位置進行裁切，並且再一次做特殊之 Data Augmentation 增加樣本，最後將處理後之影像輸入至 FCN 做 Semantic Segmentation，最終將結果輸出。

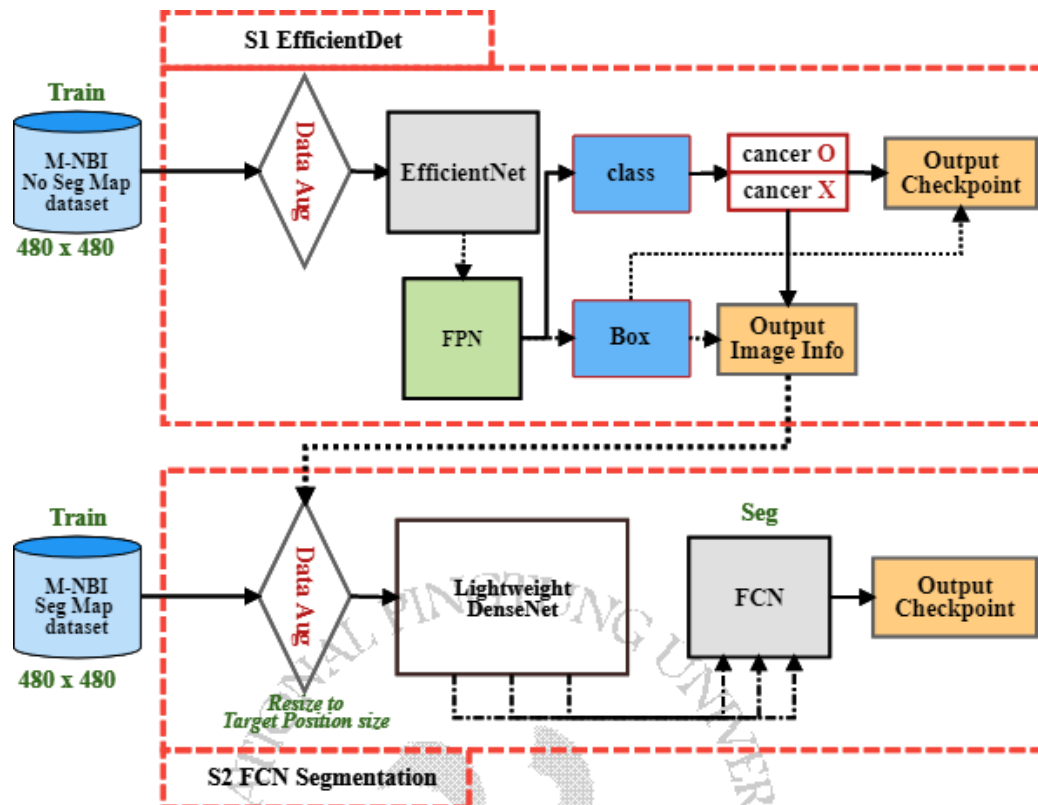


圖 3-1、M-NBI 胃癌檢測之神經網路模型訓練流程圖

第二節、研究方法概念

一、使用影像預處理產生更多樣本

由於傳統之影像預處理方法能產生之額外樣本十分有限。本論文這邊採用一種特殊之預處理方式能有效增加樣本之數量 (圖 3-2)，以下為處理步驟：(1) 將原始未做任何處理 480×480 之訓練圖像做傳統資料預處理 (垂直、水平翻轉、90 度、180 度和 270 度旋轉) 做影像增強、(2) 產生一張遮罩分割圖 (本論文主要以 2048×2048 解析度作為實驗大小)、(3) 在該遮罩分割圖上使用 Radom Crop 操作以獲得大小為 480×480 之子圖、(4) 將經過幾何影像預處理之影像和 Radom Crop 之子圖相乘、(5) 將經過幾何影像預處理之影像和特殊預處理之影像輸出。

使用此影像預處理方法以及搭配傳統預處理對圖像進行幾何旋轉，能將數據集樣本數量提升至 8 倍以上，過程中產生之遮罩分割圖，由於是隨機產生，能讓每一次訓練時輸入的圖像特徵皆有一定程度上之差異。

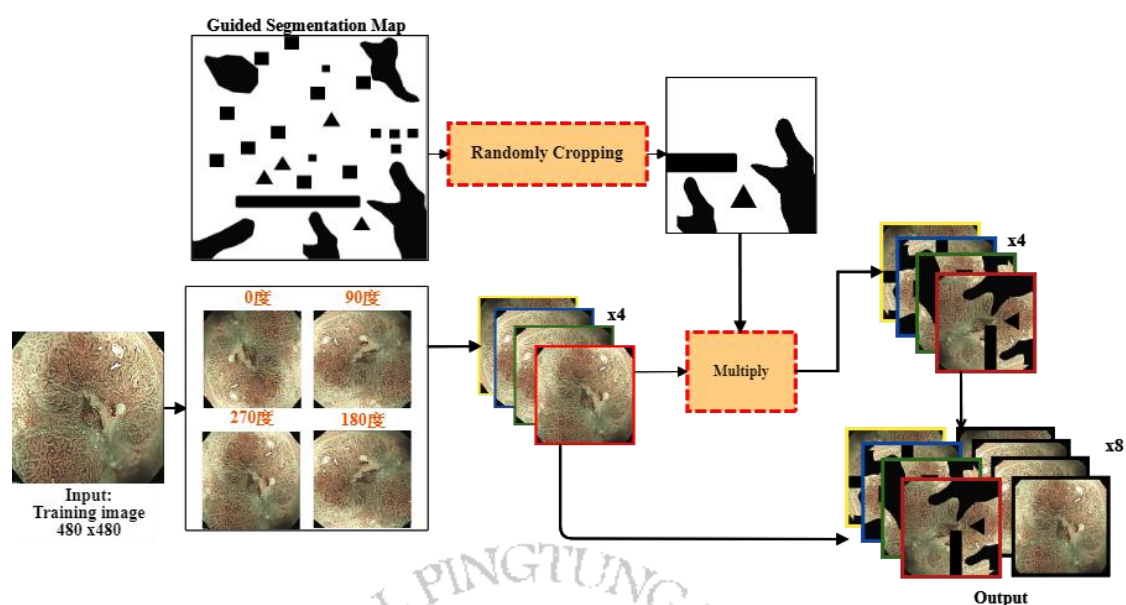


圖 3-2、特殊影像預處理示意圖

二、EfficientDet Selection

EfficientDet 網路主要應用在 S1 階段，其作用性是實現 Object Detection 之技術，同時又能做到 Classification 之目的，由於 S1 階段偵測出之結果會影響 S2 階段，故需要挑選高準確度與高速度之網路，考慮到 EfficientDet-D3 架構網路之後之框架檢測時間較長，其產生過於龐大之參數不利於本研究之小資料集，最後以精確度穩定、速度較快之 EfficientDet-D1 作為主要網路。

三、 修改 DenseNet 架構

DenseNet 主要在 S2 階段使用，目的是作為 FCN 骨幹網路，經過 S1 階段進行初步框選癌症區域後，需要檢測的區域會大幅度之縮小，會使整個流程加快，同時 DenseNet 原先是用來訓練較大的資料集，為了要在小資料集進行訓練時避免出現 overfitting，本論文將原先四個 DenseBlock 減少至三個 (圖 3-3) 能夠有效減少參數數量，同時加快整個訓練速度。

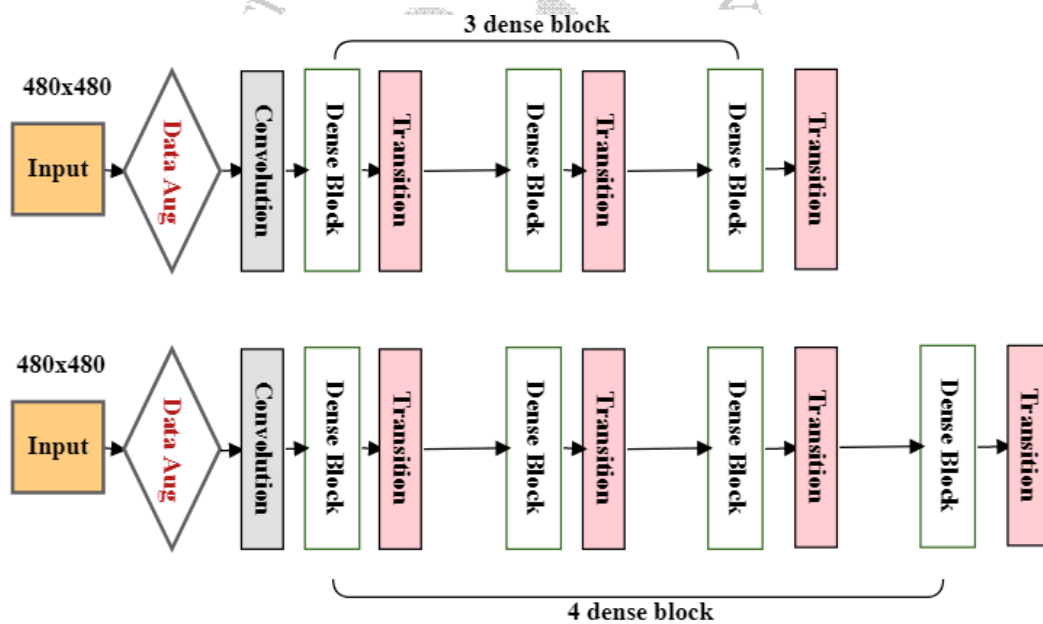


圖 3-3、Lightweight DenseNet, (上)修改後 (下)修改前

四、修改 FCN 架構

FCN 主要在 S2 階段使用，原先 FCN 框架採用之骨幹網路是 AlexNet，故需做一點修改 (圖 3-4)，將 Lightweight DenseNet 取代 AlexNet 部分，並以 Transition Layer 作為 UpSample 特徵訊息擷取層，同時在該階段還要驗證經過 Object Detection 預測出來之結果是否能夠接著做 Semantic Segmentation，在此階段中會將訓練流程分為兩個階段 (L1 與 L2)：L1 是著重讓 CNN 網路做預訓練癌症特徵，L2 階段是使用 L1 階段預訓練結果著重對 M-NBI 進行 Semantic Segmentation 訓練。

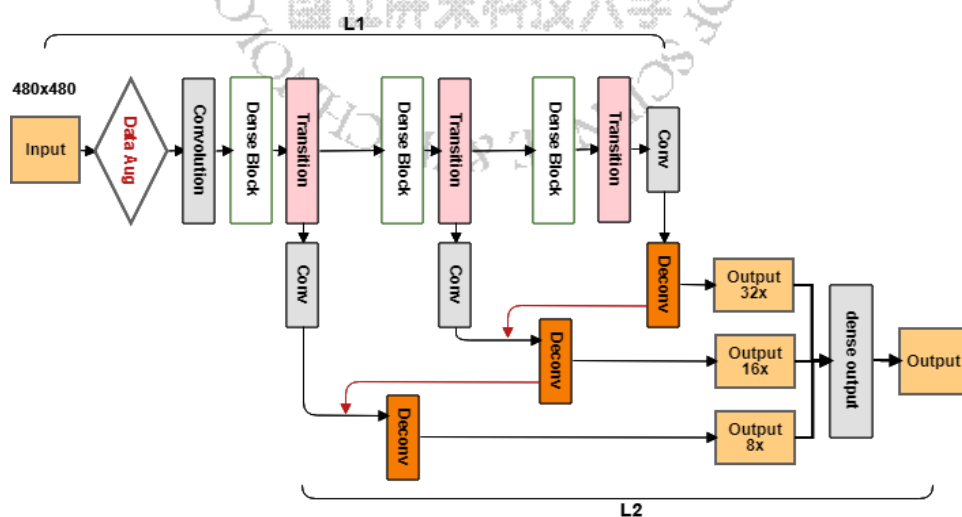


圖 3-4、FCN Lightweight DenseNet

五、 Loss Function

本論文以 Fast Normalized Fusion (1) 作為 S1 階段 Object Detection 時使用之 Weighted Feature Fusion Function；Cross Entropy (2) 則作為 S2 階段使用。

Fast Normalized Fusion 主要是使用在 S1 階段做特徵融合之權重調整，捨棄傳統 Softmax 之指數函數作法，提高整體效率，同時能達到相同的精確度。 I_i 用來表示 Iteration 的矩陣； w 用來表示可以學習之權重，可以是每一個特徵、每一個通道或是每一個像素；為了確保 $w_i \geq 0$ ，會在每次計算完 w_i 後進行 ReLU； ϵ 通常被設定為 0.0001 用來穩定 w 。

$$O = \sum_i \frac{w_i}{\epsilon + \sum_j w_j} \cdot I_i, \quad w_i = \max(0, x) \quad (1)$$

Cross Entropy 主要是用在 S2 階段時進行 Classification 任務。 C 是類別數，表示所有 dataset 中有幾個分類； n 是所有資料的筆數； $y_{C,i}$ 是 binary indicator，通常為 0 或是 1，表示第 i 筆資料屬於第 C 類的真實類別， $p_{C,i}$ 是第 i 筆資料屬於第 C 類預測出來的機率。

$$H = \sum_{i=1}^n -y_{C,i} \log_2(p_{C,i}) \quad (2)$$

六、 M-NBI 癌症區域檢測訓練步驟

S1 訓練階段 (Object Detection) :

- 步驟1: 輸入 480×480 之 M-NBI 圖像
- 步驟2: 圖像進行特殊 Data Augmentation 增加訓練樣本
- 步驟3: 輸出處理後之圖像進入 EfficientNet 開始訓練
- 步驟4: 進入 BiFPN 進行 特徵融合
- 步驟5: 輸出預測分類結果與預測檢測目標結果
- 步驟6: 儲存預測分類結果與預測檢測目標結果

S2 訓練階段 (Semantic Segmentation) :

- 步驟1: 輸入 480×480 之 M-NBI 圖像
- 步驟2: 讀取 S1 階段儲存結果與座標位置
- 步驟3: 依據 S1 階段座標位置進行圖像裁切
- 步驟4: 輸出處理後之圖像進入 DenseNet 開始訓練
- 步驟5: 透過每一層 Transition Layer 抓取特徵訊息
- 步驟6: 將特徵訊息導入 FCN 網路進行 UpSample
- 步驟7: 輸出 Semantic Segmentation 結果

第三節、評估指標

本論文主要以 Precision、Recall、Specificity 和 Sensitivity 作為評估指標，評估公式內部運算單位 Confusion Matrix 如表 3-2 所示，以下公式評估方式皆以所有測試樣本檢測結果之平均為標準。

表 3-2、評估指標之運算單位

Confusion Matrix			
	Actual Class		
		True	False
	Predicted Class		
	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)
True Positive		是癌症圖像且模型判斷是癌症圖像	
False Negative		是癌症圖像但模型沒有判斷出來是癌症圖像	
False Positive		不是癌症圖像但模型判斷是癌症圖像	
True Negative		不是癌症圖像且模型沒有判斷是癌症圖像	

七、 Specificity

Specificity (3) 是用來判斷為無癌症圖像之情況下，有多少是真的無癌症圖像。

$$Specificity = \frac{TN(nocancer)}{FP+TN} \times 100\% \quad (3)$$

八、 Sensitivity (Recall)

Sensitivity (4) 是用來判斷為癌症圖像之情況下，有多少癌症圖像被正確判斷出來。

$$Sensitivity = \frac{TP(cancer)}{TP+FN} \times 100\% \quad (4)$$

九、 F1-score

F1-score (5) 是利用 Specificity 和 Sensitivity 之調和平均數，用來衡量檢測模型之效能表現。

$$F1-score = 2 \times \frac{Specificity \times Sensitivity}{Specificity + Sensitivity} \times 100\% \quad (5)$$

十、 Accuracy

Accuracy (7) 是用來判斷在所有情況下有被正確判斷之有癌症 (TP) 與無癌症 (TN) 之機率為多少。

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

十一、 Average Precision (AP)

Average Precision (8) 主要是用在評估 Object Detection 網路中之評估指標，其數值大小受 Precision 和 Recall 曲線 (PR 曲線) 下之面積與閾值 (Thresholds) 影響。th 為閾值、x 為類別數量、

n 為 PR 曲線中 Recall 維度上 Precision 之最大值數量、 r 為 Recall、 P 為 Precision。

$$\text{mAP} = \frac{\sum_{th=\{0.5,0.55,\dots,0.75\}} AP(IoU_{th}=th)}{x}, AP = \frac{1}{n} \sum_r P_{interp}(r) \quad (7)$$

第四節、M-NBI 胃癌區域檢測之測試流程 (圖 3-5)

S1 Test 階段

步驟1: 輸入無 Segmentation Map 之 M-NBI 胃癌影像測試集

步驟2: 將測試集影像導入 S1 階段完成訓練網路

步驟3: 進行 Object Detection 和 Classification 預測

步驟4: 輸出檢測結果與檢測時間

S2 Test 階段

步驟1: 輸入有 Segmentation Map 之 M-NBI 胃癌影像測試集

步驟2: 將測試集影像導入 S2 階段完成訓練網路

步驟3: 進行 Semantic Segmentation 預測

步驟4: 輸出檢測結果與檢測時間

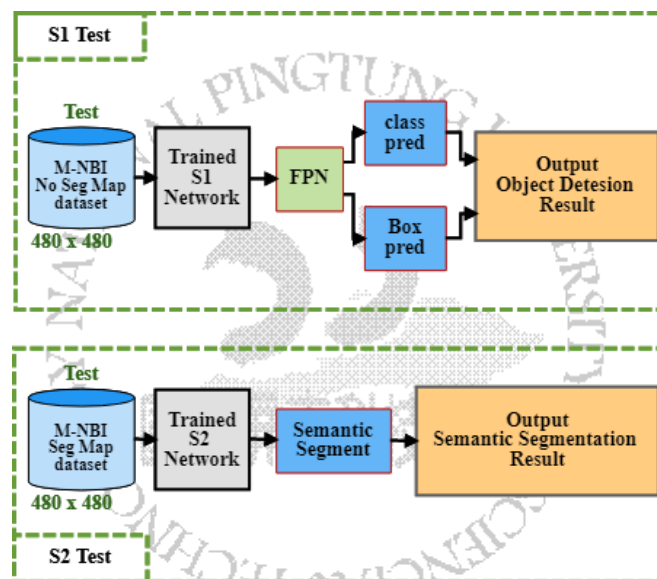


圖 3-5、S1 階段 (上) 與 S2 (下) 階段測試流程圖

第肆章 實驗成果

在這個章節中主要展示本論文之實驗成果，總共分為 5 個小節：

(1) 實驗環境、(2) 實驗過程中使用哪些小技巧、(3)展示 S1 階段實驗成果、(4)展示 S2 階段實驗成果、(5)展示 S1 與 S2 網路合併後之實驗結果。

第一節、實驗環境

本論文實驗環境如表 4-1 所示，其中電腦硬體之處理器為 Intel(R) Core™ i9-9900K CPU@3.6GHz、記憶體為 32GB、圖形處理器為 2 張 EVGA GeForce RTX2080Ti 11GB；電腦作業系統環境為 Linux Ubuntu 16.04 & 18.04；撰寫程式語言為 Python 3.6.2；執行環境為 Cuda 10.2、Cudnn 7.6.4、Tensorflow-gpu 1.15 與 Pytorch 1.0.1。

表 4-1、實驗環境

硬體設備與執行環境	規格
作業系統	Linux Ubuntu 16.04 & 18.04
處理器	Intel(R) Core™ i9-9900K CPU@3.6GHz
圖形處理器	EVGA GeForce RTX2080Ti 11GB EVGA GeForce RTX2080Ti 11GB
記憶體	32GB
撰寫程式語言	Python 3.6.2
執行環境	Cuda 10.2、Cudnn 7.6.4、Tensorflow-gpu 1.15、 Pytorch 1.0.1

第二節、S1 階段 Object Detection 參數設定與實驗結果

一、S1 階段參數設定

本論文使用 Object Detection Network 作為 S1 階段之檢測網路，在任務期間除了使用 EfficientDet-D1 做癌症檢測，也將 EfficientDet-D3、M2DetNet 和 Faster-RCNN-Net 等近年發表之網路作為效能評估基準。EfficientDet 之參數設定如表 4-2 所示，資料集為 113 張有癌症 Ground Truth 標記之 M-NBI 圖像，訓練資料與測試資料分別為 100 張與 13 張。

表 4-2、實驗參數設定

Parameter	Input size	Optimizer	Learning Rate	momentum	Activation Function	iteration
Value	480	Adam	10^{-4}	0.997	Relu	1000

二、S1 階段實驗成果

本論文主要以 EfficientDet-D1 作為主要檢測網路，主要目的是將癌症區域進行 Object Detection，並且擷取預測癌症區域供第二階段之 Semantic Segmentation 網路使用，在該階段訓練過程中，主要以 15 個 epochs 或 30 個 epochs 做為效能評估間距；本論文對 Ground Truth 圖像進行兩種不同之標框方式（圖 4-1，紅框部分為 Ground Truth 區域，背景白色區域為癌症區域遮罩圖），以下兩

種方法簡稱為多框標記 (Ann1) 和單框標記 (Ann2)。多框標記主要以有癌症內部區域進行標框，讓網路著重於癌症特徵之學習，減少檢測網路學習到 Ground Truth 區域之非癌症區域；單框標記為較常見之標框方式，以目標位置之外圍區域進行標框，會將非癌症區域一同框選為 Ground Truth；單框標記為本論文任務初期使用之方法，初步訓練之結果發現檢測網路很容易將無癌症區域視為有症狀，以及使用極大之檢測框進行框選 (圖 4-2)，故想出多框標記方法讓檢測網路針對癌症特徵進行學習，由於 Ground Truth 區域之學習範圍不同，單框標記在精確度上會領先多框標記一些。

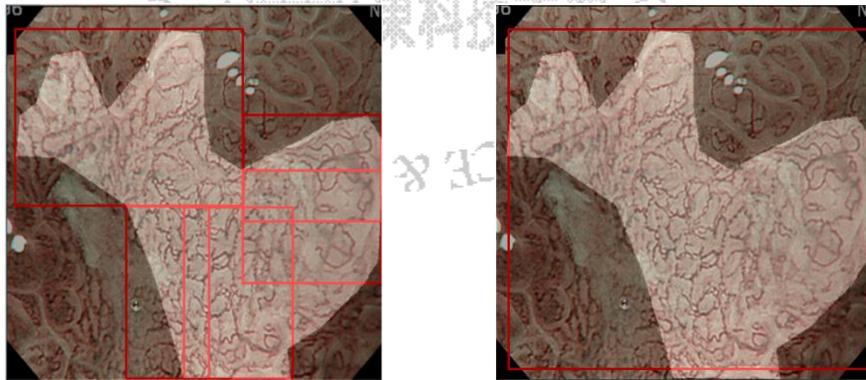


圖 4-1、多框標記(左) 與 單框標記 (右) 示意圖

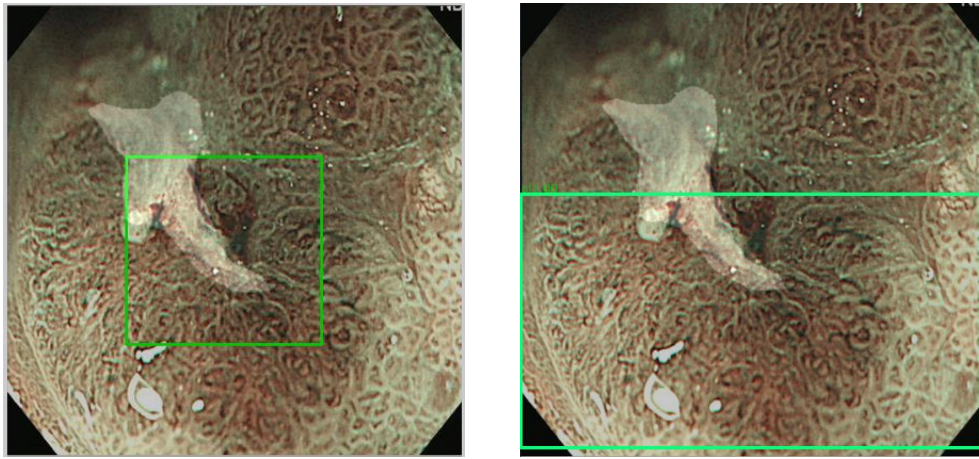


圖 4-2、多框標記 (左) 與單框標記 (右)方法檢測結果，白色背景
為癌症遮罩區域

為了確認標框之方式是否大幅影響檢測網路之效能，本論文在個別評估網路上皆用了多框標記與單框標記，在 One-Stage 與 Two-Stage 檢測網路之 Max Detection Box 設定 1 至 100 之間情況下 (圖 4-3)，同時 AP 閥值 IoU (Intersection over Union) 設為 0.5，在 Detection Box 較多之情況下，多框標記得出之效能比單框標記方法好上不少，Detection Box 較少之情況下單框標記比多框標記效能好一些，閥值 IoU 設為 0.75 情況下更是明顯，而在 Two-Stage 框架底下 Detection Box 之數量設定大或是小，其單框標記效能會比多框標記更佳；此外，根據表 4-3 結果，本論文使用 EfficientDet-D1 作為檢測網路，訓練能比較快得到收斂，其檢測速度也能比 EfficientDet-D3 與 M2-DetNet 還要快。

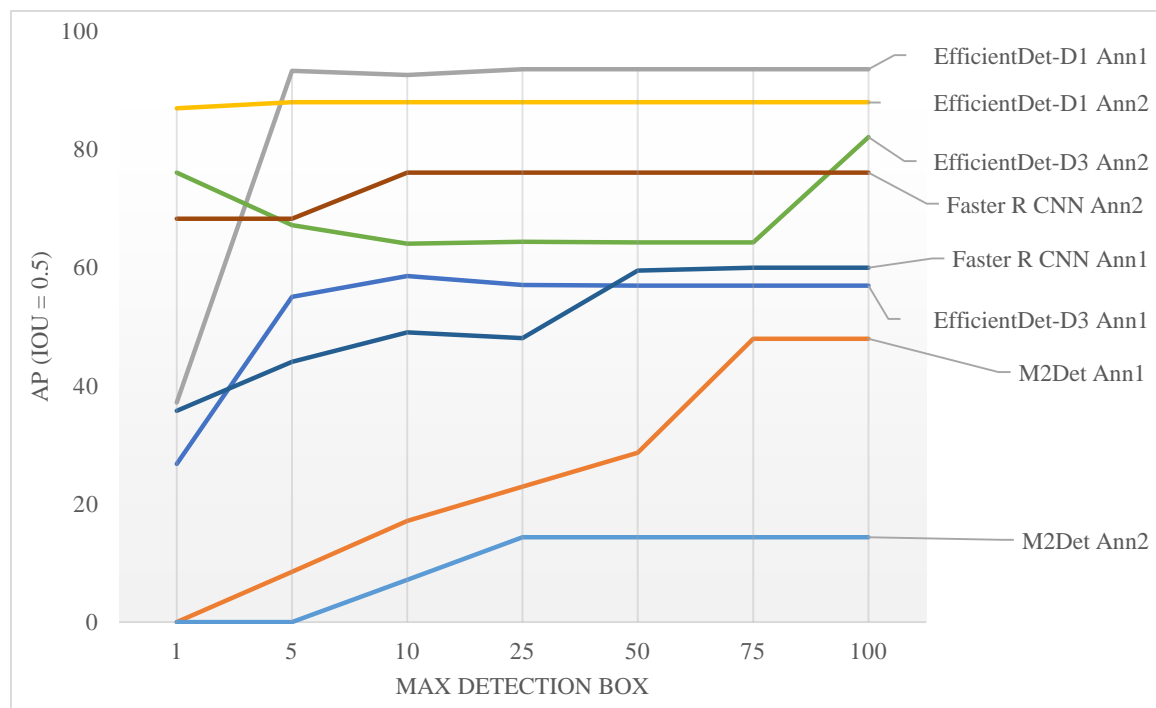


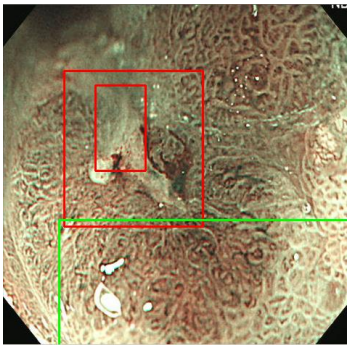
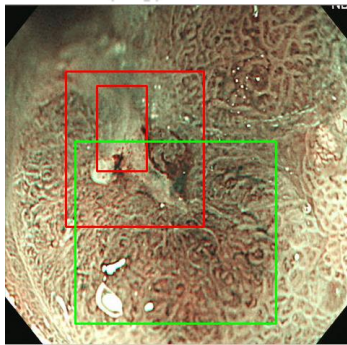
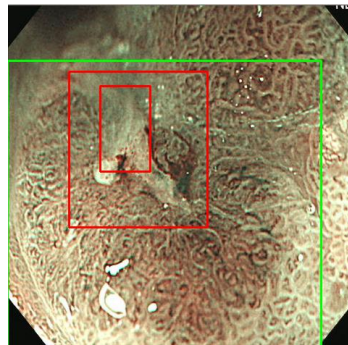
圖 4-3、個別 Max Detection Box 數量比較圖

表 4-3、個別 Object Detection Network 效能比對圖

Epoch = 30, Test data = 20 張						
Model	Backbone	Sensitivity (IoU 0.5)	F1-score (IoU 0.5)	AP (IoU)		FPS (VAL)
				0.5	0.75	
One-Stage (Detection Box=100)						
M2Det Ann1	ResNet-101	28.6%	43.6%	47.9%	34.5%	18±5
M2Det Ann2	ResNet-101	33.2%	48.4%	38.5%	7.1%	18±5
EfficientDet-D1 Ann1	EfficientNet-B1	93.7%	96.2%	93.5%	79.9%	70±5
EfficientDet-D1 Ann2	EfficientNet-B1	98.9%	98.8%	87.9%	4.8%	70±5
EfficientDet-D3 Ann1	EfficientNet-B3	81.1%	89.4%	56.9%	35.9%	29±5
EfficientDet-D3 Ann2	EfficientNet-B3	9%	16.2%	64.1%	35.9%	29±5
Two-Stage (Detection Box=100)						
Faster R CNN Ann1	ResNet-101	51.4%	66%	59.9%	36.4%	17±2
Faster R CNN Ann2	ResNet-101	76.9%	85.8%	76.1%	74.6%	17±2

為了增強多框標記方法在標框上之範圍，本論文針對閥值 IoU 分數前幾個較高之 Detection Box 進行合併處理，以加大檢測癌症可視區域，主要方法以 Detection Box 之座標取平均值或取最大值作為擴大可視範圍之方法（表 4-4、綠色方框為檢測結果，紅色方框為 Ground Truth），在取前 5 個 Detection Box 作為實驗情況下，取座標最大值能得到最大之可視範圍，但會框選到大範圍非癌症區域；取座標平均值情況下能得到較小之可視範圍，但會漏掉很多有癌症區域。

表 4-4、取最大值座標與平均值座標之檢測結果

Single Detection	Average Detection	Max Detection
Detection Box= 5		
		

第三節、S2 階段 Semantic Segmentation 參數設定與實驗

結果

一、S2 階段參數設定

本論文使用 Semantic Segmentation Network 作為 S2 階段之檢測網路，FCN 之參數設定如表 4-5 所示，在該階段實驗中分為兩個階段訓練網路，分別為 L1 階段與 L2 階段，L1 階段作為預訓練分類癌症使用，L1 資料集為 4149 張做為癌症分類訓練集，76 張做為癌症分類測試集；L2 階段作為 Semantic Segmentation 使用，L2 資料集為 170 張有標記與無標記 M-NBI 圖像作為訓練資料集，其中 99 張有癌症，71 張無癌症；20 張有標記與無標記 M-NBI 圖像作為測試資料集，其中 14 張有癌症，6 張無癌症。

表 4-5、FCN 實驗參數設定

Parameter	Input size	Optimizer	Learning Rate	momentum	Activation Function	iteration
Value	480	Adam	5×10^{-5}	0.9	Relu	1000

二、S2 階段實驗成果

為了提高 Semantic Segmentation 檢測網路之深度與複雜度，本論文將 FCN Net 原先之骨幹網路 AlexNet 更換成 DenseNet，為

了減少過多的參數，本論文將 DenseNet 進行輕量化設計，同時為了彌補訓練資料集不夠之問題，加入本論文提出之 Data Augmentation 方法進行訓練，其更改結果在效能如 表 4-6 所示，epoch 低於 1000 之前，其精確度皆低於傳統資料預處理，主要因為使用遮罩對 Ground Truth 進行隨機遮蓋，會增加網路訓練難度，但在 epoch 高於 1000 以上時模型開始收斂，故精確度有了顯著之提升，如結果所示，使用本論文提出之資料愈處理方法會比傳統資料愈處理方法精確度還要高 7% 上下。

表 4-6、有無做 Data Augmentation 效能比對

Epoch=1200, Input Size=480 × 480, Test data = 20 張							
Model	My Data Augmentation	Sensitivity	F1-score	Accuracy	TP	FN	FPS
Lightweight DenseNet	X	42.8%	59.1%	60%	6	8	60
Lightweight DenseNet	V	64.2%	72.5%	70%	9	5	60
Epoch=900, Input Size=480 × 480, Test data = 20 張							
Lightweight DenseNet	X	40%	57%	55%	6	9	60
Lightweight DenseNet	V	28.5%	42.5%	45%	4	10	60

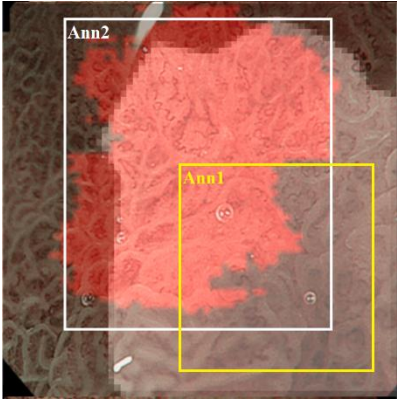
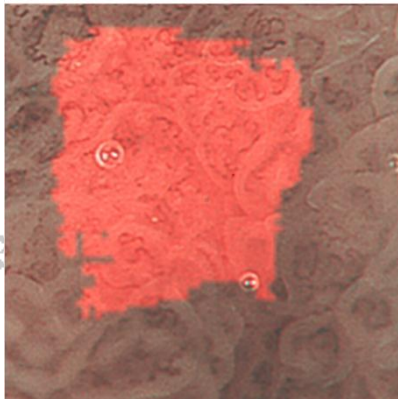
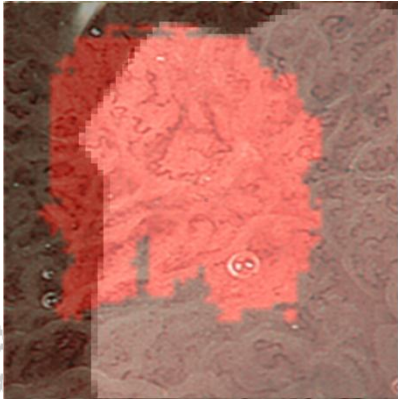
第四節、S1 階段與 S2 階段合併之實驗結果

本階段為驗證 S1 階段之檢測結果送至 S2 階段檢測網路是否能與單純做 Semantic Segmentation 網路得到相同效能，評估方法為將檢測結果之 Detection Box 作為 S2 階段網路之測試資料在與做 Semantic Segmentation 網路結果做效能比對。

從結果上來看 (表 4-7)，使用兩階段 (S1 與 S2 合併) 檢測網路之結果做 Semantic Segmentation 檢測癌症比單純做 Semantic Segmentation 網路效能好一些，但在 Segmentation 圖像呈現上有些遜色，本論文認為原因是 Semantic Segmentation 網路會學習癌症與非癌症之特徵進行區分差異，在影像為完整之情況下，Semantic Segmentation 網路能學到比較準確之特徵，這點使用單框標記會有比較好之效能，反之，如果圖像是使用多框標記方法，透過 Object Detection 網路切割過後再送至 Semantic Segmentation 網路進行檢測，由於圖像尺度會大幅縮小，其癌症特徵連貫性也會消失。基於本論文研究假設上，以驗證 Object Detection 之檢測結果做 Semantic Segmentation 是可行且效能不會落差太多甚至可能更好，在精確度評估上使用兩階段檢測網路之效能比單純使用 Semantic Segmentation 網路還要高 17% 至 20%。

表 4-7、Semantic Segmentation Network 與 Two Stage Detection Network 效能評估

Model	Sensitivity	F1-score	TP	FN	Accuracy
Semantic Segmentation Network	64.2%	72.5%	9	5	70%
Two Stage Detection Network-Ann1	83.3%	85.3%	10	2	87.5%
Two Stage Detection Network-Ann2	84.6%	91.6%	11	2	89.7%

		
Semantic Segmentation Network	Two Stage Detection Network-Ann1	Two Stage Detection Network-Ann2

本論文在所有實驗結果中發現幾項很難避免之問題，在測試階段裡有少數幾張 M-NBI 容易被誤判為無癌症 (False Positive)，主要原因是圖像上有很嚴重之反光或圖像過暗以及氣泡過多導致部分癌症區域因無法被正確檢測導致無法對圖像上之異常區域做框選或進行上色，以及部分假陽性 (False Negative) 特徵容易被檢測網路判斷為有癌症。

第伍章 結論與建議

本論文基於深度學習之 M-NBI 胃癌區域檢測，利用現有深度學習網路架構進行訓練與改良，透過階段性訓練達成兩階段快速檢測之目的，第一階段主要以 Object Detection 進行初步大範圍預測癌症區域以及分辨出 M-NBI 圖像是否有癌症特徵，S2 階段使用 Semantic Segmentation 檢測網路，將 S1 階段檢測結果進行 Segmentation，驗證將 Object Detection 之預測結果直接做 Segmentation 是可實現之方法，在 S1 階段中 EfficientDet 之 AP 可達 90% 上下，FPS 在 70 上下，S2 階段中 Lightweight DenseNet FCN 之 Accuracy 可達 70%，FPS 在 60 上下，將兩階段網路合併後之 Accuracy 可達 89%，FPS 在 50 上下。

由於現階段之兩階段網路檢測成功比率未達 9 成以上，速度也沒有比單純使用 Semantic Segmentation 網路還要快，未來預計對整體網路進行調整後將 S1 階段與 S2 階段網路進行合併 (圖 5-1)，在進行 Object Detection 過程中，將預測框進行輸出導入至 Semantic Segmentation 框架，讓網路能進行 Object Detection 同時做 Semantic Segmentation，由於僅對 Object Detection 之預測框做 Semantic Segmentation，速度上會比起兩階段網路與單純做 Semantic

Segmentation 來進行檢測癌症來的快。

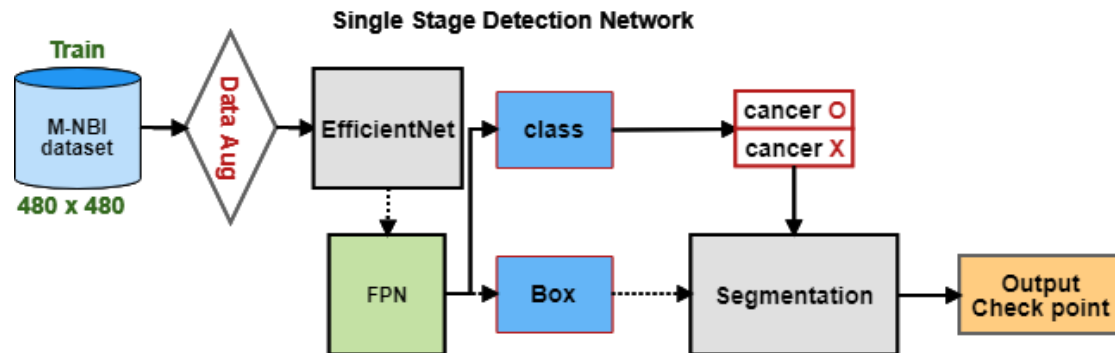


圖 5-1、Single Stage Detection Network

第陸章 參考文獻

- [1] Uedo, Noriya, et al., "Role of narrow band imaging for diagnosis of early-stage esophagogastric cancer: current consensus of experienced endoscopists in asia-pacific region," *Digestive Endoscopy*, Vol. 23, pp. 58-71, 2010
- [2] Ezoe, Yasumasa, et al., "Magnifying narrowband imaging is more accurate than conventional white-light imaging in diagnosis of gastric mucosal cancer," *Gastroenterology*, Vol. 141, pp. 2017-2025, 2011.
- [3] Lee, Tsung-Chun, et al., "Computer-aided diagnosis in endoscopy: a novel application toward automatic detection of abnormal lesions on magnifying narrow-band imaging endoscopy in the stomach," *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Vol. 35, pp. 4430-4433, 2013.
- [4] Kanesaka, Takashi, et al., "Computer-aided diagnosis for identifying and delineating early gastric cancers in magnifying narrow-band imaging," *Gastrointestinal*

- endoscopy*, Vol. 87, pp. 1339-1344, 2018.
- [5] Rajpurkar, Pranav, et al., "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.
 - [6] Yalçın, Nursel, Seyfullah Alver, and Necla Uluhatun. "Classification of retinal images with deep learning for early detection of diabetic retinopathy disease," *Proceedings of IEEE Signal Processing and Communications Applications Conference (SIU)*, Vol. 26, pp. 1-4, 2018.
 - [7] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," *Nature*, Vol. 521, pp. 436-444, 2015.
 - [8] Chen, Liang-Chieh, et al., "Encoder-decoder with atrous separable convolution for semantic image segmentation," *Proceedings of the European conference on computer vision (ECCV)*, pp. 801-818, 2018.
 - [9] Alagappan, Muthuraman, et al., "Artificial intelligence in gastrointestinal endoscopy: the future is almost here," *World journal of gastrointestinal endoscopy*, Vol. 89, No. 4, pp. 816-817, 2018.
 - [10] Maroulis, Dimitrios E., et al., "CoLD: a versatile detection system for colorectal lesions in endoscopy video-frames," *Computer Methods and Programs in Biomedicine*, Vol. 70, No. 2, pp.151-166, 2003.
 - [11] Bernal, Jorge, et al., "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Computerized Medical Imaging and Graphics*, Vol. 43, pp. 99-111, 2015
 - [12] Kominami, Yoko, et al., "Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy," *Gastrointestinal endoscopy*, Vol. 83, No. 3, pp. 643-649, 2016.

- [13] Hirasawa, Toshiaki, et al., "Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images." *Gastric Cancer*, Vol. 21, No. 4, pp. 653-660, 2018.
- [14] Liu, Wei, et al., "Ssd: Single shot multibox detector," *European conference on computer vision*, pp. 21-37, 2016.
- [15] Zhong, Zhun, et al., "Random erasing data augmentation," *arXiv preprint arXiv:1708.04896*, pp. 13001-13008, 2017.
- [16] Tan, Mingxing, and Quoc V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, 2019.
- [17] Zoph, Barret, et al., "Learning transferable architectures for scalable image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8697-8710, 2018.
- [18] He, Kaiming, et al., "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.
- [19] Huang, Gao, et al., "Densely connected convolutional networks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700-4708, 2017.
- [20] Ren, Shaoqing, et al., "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, pp. 97-99, 2015.
- [21] Girshick, Ross, et al., "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580-587, 2014.
- [22] Redmon, Joseph, et al., "You only look once: Unified, real-time object detection," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779-788, 2016.

- [23] Tan, M., R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection. arXiv," *arXiv preprint arXiv:1911.09070*, 2019.
- [24] Liu, Shu, et al., "Path aggregation network for instance segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8759-8768, 2018.
- [25] Pang, Jiangmiao, et al., "Libra r-cnn: Towards balanced learning for object detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 821-830, 2019.
- [26] Zhao, Qijie, et al., "M2det: A single-shot object detector based on multi-level feature pyramid network," *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, pp. 9525-9266, 2019.
- [27] Long, Jonathan, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431-3440, 2015.
- [28] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," *Proceedings of Advances in neural information processing systems*, pp. 1097-1105, 2012.
- [29] Zhao, Zhong-Qiu, et al., "Object detection with deep learning: A review," *Proceedings of IEEE transactions on neural networks and learning systems*, Vol. 30, No. 11, pp.3212-3232, 2019

作者簡介

作者姓名：簡志宇

性別：男

學歷：國立屏東科技大學 資訊管理系 碩士班

國立屏東科技大學 資訊管理系 學士班

研究興趣：人工智慧、醫療輔助檢測、資料分析

競賽經歷：

1. Competition results for VCIP 2019 Grand Challenge
第 3 名
2. (Aidea) 尋找病媒蚊孳生源-積水容器影像物件辨識
第 11 名
3. Embedded Deep Learning Object Detection Model Competition
第 5 名
4. Automatic Detection challenge on Age-related Macular degeneration
第 4 階段第 3 名、不分階段小組名次第 4 名
5. 國立屏東科技大學 108 學年度傑出學生表現
獎助金 30000 元
6. 國立屏東科技大學 108 學年管理學院學生學術獎助
獎助金 5000 元