# 肺腺癌病理切片影像之腫瘤氣道擴散偵測競賽
# II：運用影像分割作法於找尋STAS

## 壹、環境

- 作業系統：Ubuntu 18.04.6 LTS
- 語言：Python 3.8.12
- GPU: NVIDIA-RTX2080TI
- GPU Driver Version：440.33.01
- CUDA Version：10.2
- 套件(函式庫)：mmdetection
- 預訓練模型：
  CBNetV2中的Improved HTC模型，backbone為swin transformer，使用dual backbone方式訓練，表格1為模型細節，以下也附上模型的checkpoint 和configuration。
  - 預訓練 checkpoint
  - 預訓練使用的configuration

表格 1模型細節

| Backbone | Lr Schd | box mAP (minival/ test-dev) | mask mAP (minival/ test-dev) | #params |
|---|---|---|---|---|
| DB-Swin-B | 20e | 58.4/58.7 | 50.7/51.1 | 235M |

- 額外資料集：pretrain on ImageNet-22k

## 貳、演算方法與模型架構

【模型架構1: CBNetV2】

在這一次的比賽中，我們使用了CBNetV2 神經網路架構進行訓練，這個架構的特色是能適用於各式現存模型，將他們進行組合共同訓練在影像識別任務上取得良好成效。我們在這個比賽選定 swin transformer 作為模型的backbone，CBNetV2會在訓練過程中組合多個相同的backbone，由其中一個 backbone 處於領導地位，偕同複數個支援地位的backbone一同訓練。領導和支援的backbone透過論文提出的dense higher-level composition方式連接，每個backbone淺層stage的輸出都將作為領導地位backbone深層stage的輸入的一環(圖1)，配合論文提出的 assistant supervision方法有效提升領導backbone的預測表現。

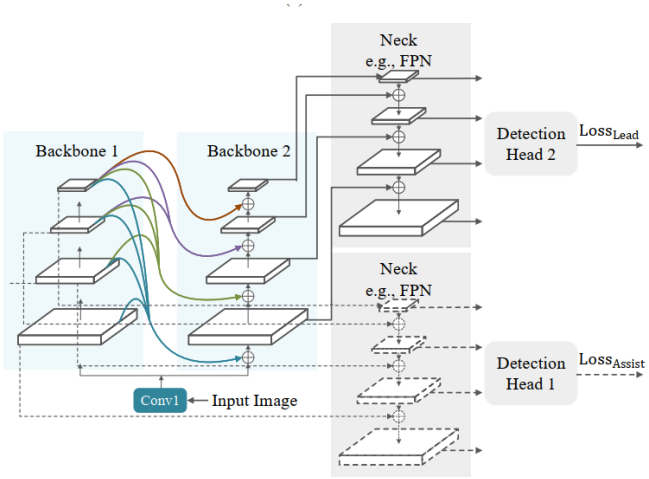



圖 1 Dense Higher-Level Composition

【模型架構2: Hybrid Task Cascade】

在 CBNetV2 所提出的訓練模式之外，因為這次比賽提供的資料同時涵蓋了object detection 和 semantic segmentation 的標註，因此我們選定了能同時訓練的這兩項任務的 Hybrid Task Cascade 架構進行訓練。該架構融合了 Mask RCNN 和 Cascade RCNN 兩個模型的優點，提出漸進式細化的級聯管道，在訓練的每個 stage，邊界框迴歸和 segmentation 的預測都以多任務方式組合，同時不同 stage 的 mask 分支（用來預測 semantic segmentation 的分支）之間用直接的信息流串聯。

【backbone 模型: swin transformer】

swin transformer 當初在設計的時候，主要是為了解決兩個問題。首先是物體尺寸變化大，在不同場景下 Vision Transformer 性能未必好，圖像分辨率高。第二個難點是像素點多，Transformer 基於全局自注意力的計算導致計算量較大。為了解決以上兩個問題，swin transformer是一種包含sliding window，具有層級設計的transformer結構。

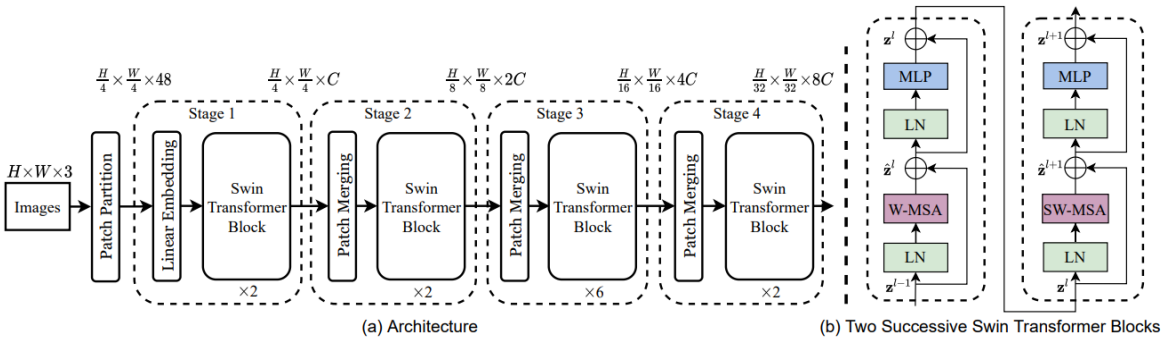



圖 2 Swin Transformer示意圖

- 如圖2所示，swin transformer會對輸入圖片進行patch embedding，亦即先將圖片切成一個個圖塊，並嵌入代表位置資訊的embedding。
- 每個stage由patch merging和多個block組成。
- patch merging模塊的功用為在每個stage開始時降低圖片分辨率。
- block具體結構如上面右圖所示，主要由LayerNorm，MLP，window attention 和 shifted window attention組成。

## 參、資料處理

【調整輸入格式】

將資料轉換成coco data format（json），用segmentation提供的annotations，其中的bbox 設成segmentation 的min xy, max xy，接下來把object detection的資料轉換成mmdetection需要的middle format（pkl）來訓練。

使用convert_STAS.py 將改變資料格式，使用方式可以參閱我們 github repository的 README.md。

【Testing Time Data Augmentation】

(a) MultiScaleFlipAug

透過將同一張照片翻轉、縮放後投入模型得到複數預測結果，將所有預測結果匯總篩選出更高機率的預測框。

其中我們在multi scale augmentation上採用了以下5種scale的圖片投入模型後的輸出彙總成最終結果。[(858, 471), (943, 518), (1000, 565), (1115, 612), (1200, 660)]。

(b) 將nms改成soft_nms，並將max bbox per image改成比賽方提出的200

## 肆、訓練方式

我們在這次比賽找到能同時訓練semantic segmentation 和 object detection任務的模型Hybrid Task Cascade，配合CBNetV2的訓練模式，達到了約0.88的預測成果。

Hybrid Task Cascade 架構具有兩個分支，一個負責 semantic segmentation，一個負責 object detection，為了充分利用該模型雙任務協同訓練的特長，我們將 semantic segmentation 標註中的每一個病變的上下左右界找出來，創造出與 semantic segmentation 標註對應的bounding box投入object detection分支。我們訓練了約50個epoch，其中learning rate初始值為5e-5，分別第6、第10 epoch會乘上0.1變小。

## 伍、分析與結論

我們在 CBNetV2 預設的參數配置之上，調整以下超參數，觀察結果決定最終配置。

【針對object detection分支修改anchor box ratio】

(a) 更改anchor box ratio

我們認為更多樣的 anchor box ratio 能更有效地捕捉不同長寬比的病變部分，因此我們將anchor box ratio的數量從 3 個改成 5 個，並將標記資料的 anchor box ratio 從小到大排列取第 [16.6%, 33.3%, 50.0%, 66.7%, 83.3%] 位置的數值，其分別為 [0.78, 0.92, 1.0, 1.2, 1.41]。

(b) anchor box size

我們統計了所有 bounding box 的 hegiht 、width 的數量，以此獲得 bounding box 的大小分布。如表格 2、3 所顯示。我們發現大多數的bounding box 長寬大小都集中在100左右。在height 的統計中，height 為100單位長度的bounding box 總共佔據了84%。而 width 為 100 單位長度的bounding box 總共佔據了 84 %。因為發現多數train data中的bbox size偏小, 所以我們將base size從8改成4，更好地適應這次的任務。

表格 2 bounding box width frequency

| 大小 | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800以上 |
|---|---|---|---|---|---|---|---|---|
| 數量 | 1532 | 929 | 205 | 74 | 13 | 17 | 8 | 2 |
| 比例 | 0.550 | 0.333 | 0.073 | 0.026 | 0.004 | 0.006 | 0.002 | <0.001 |

表格 3 bounding box height frequency

| 大小 | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 |
|---|---|---|---|---|---|---|---|---|
| 數量 | 1628 | 891 | 163 | 55 | 23 | 15 | 5 | 2 |
| 比例 | 0.584 | 0.319 | 0.058 | 0.019 | 0.008 | 0.005 | 0.001 | <0.001 |

另外，為了設計anchor size我也統計了(height + width /2)後各百分位數的值。

【Loss】

在CBNetV2中，針對Hybrid Task Cascade 這一模型，預設的loss type為GioU。我們在嘗試了CioU、FocalEIoU、FocalCIoU等等的Loss後，使用CioU作為最終定案。

我們的實驗數據紀錄如下，實驗情境為以80%的資料訓練，20%的資料驗證。

(a)更改不同的 Loss

| Loss 種類 | 表現 |
|---|---|
| GIoU | 0.930 |
| CIOU | 0.934 |
| FocalEIoU | 0.933 |
| FocalCIoU (gamma = 2，加強好的bbox) | 0.934 |
| FocalCIoU (gamma = 0.5，加強難的bbox) | 0.928 |

(b) 調整output roi head的loss weight

在模型的預設值中，各種針對不同大小的output roi head都權重皆為10.0，然而我們在分析預測結果後發現模型是判斷小物件時較不精準，因此嘗試將大的bounding box 權重改為5，小的改為20，以此讓模型側重小物件的訓練。但表現不如原先好，因此最後仍沿用舊的20。

(c) 採用OHEM sampler

比較了訓練過程使用random sampler和OHEM sampler的表現，我們選定OHEM更好地平衡正負樣本的比例，讓模型在訓練過程能專注於比較難的bbox。

(d)刪去Loss function中預測「類別」的Loss

Object detection分支中，Loss function包含兩大部分，負責物件類別的classification loss和負責位置資訊的regression loss。因為在這次的比賽中，需要預測的物件僅有一類，因此我們移除classification loss。而在Hybrid Task Cascade的架構下，segmentation分支原本就不包含classification loss，故維持不變。

【Training Time Data Augmentation】

(a) 基本資料擴增

- multiscale image size：[(376, 686), (520, 950)]
- randomflip: 機率：0.5
- 標準化
  - mean=[123.675, 116.28, 103.53]
  - std=[58.395, 57.12, 57.375]

(b) albumentation

因為發現圖片間色調有明顯改變，因此我們增加了color jitter，也使用一些blur等技巧讓模型能適應不同圖片。以下是我們使用的配置。

- RandomBrightnessContrast
  - 機率：0.1
  - 亮度範圍：[0.1, 0.3],
  - 對比範圍：[0.1, 0.3]
- 兩種不同顏色變換模式隨機擇一
  - RGBShift
    - r_shift_limit=10
    - g_shift_limit=10
    - b_shift_limit=10
  - HueSaturationValue
    - hue_shift_limit=20
    - sat_shift_limit=30
    - val_shift_limit=20
- 三種不同模糊模式隨機擇一
  - Blur,blur_limit=3
  - MedianBlur, blur_limit=3
  - MotionBlur, blur_limit=6

(c) Mosaic and mixup：0.2

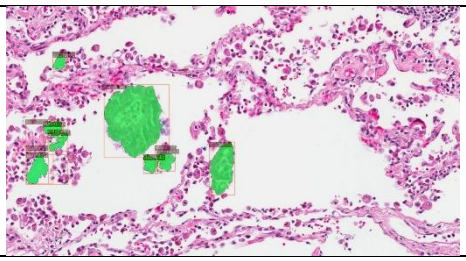
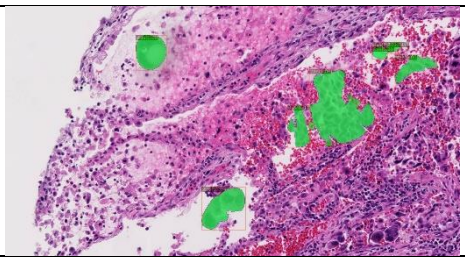
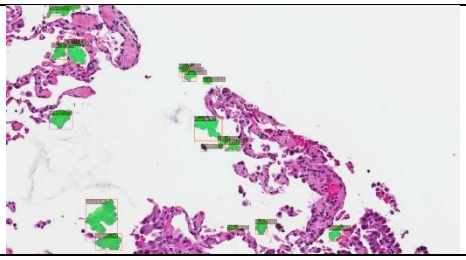
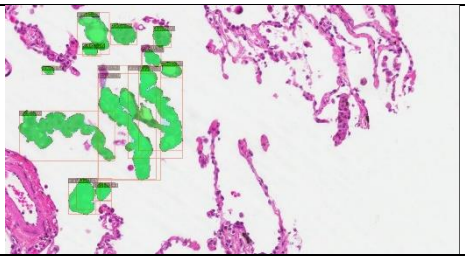
我們在借鑒了yoloV4的論文後，決定嘗試該篇論文使用過的這兩個資料擴增技巧，成功提升了我們的模型表現。

- Mosaic
  - img_scale=(942, 1716)

- pad_val=114.0
- 機率=0.2

- MixUp
  - img_scale=(942, 1716)
  - ratio_range=(0.8, 1.6)
  - pad_val=114.0
  - 機率=0.2

【成功案例展示】

| | |
|---|---|
|  |  |
| Train data 122 | Train data 135 |
|  |  |
| Train data 219 | Train data 261 |

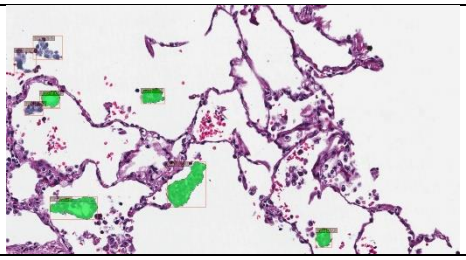
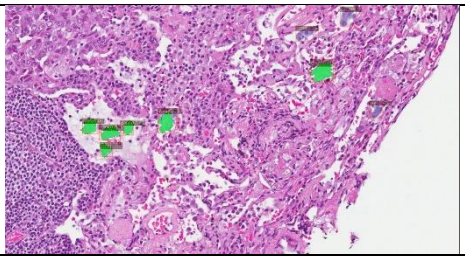
semantic segmentation的模型預測的結果以藍色表示，ground truth 以綠色表示，而bounding box皆為預測結果

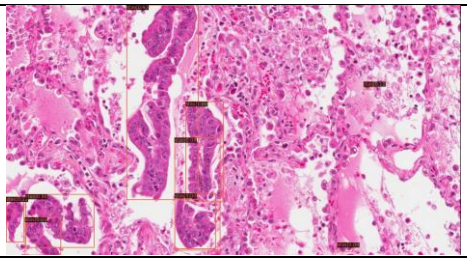
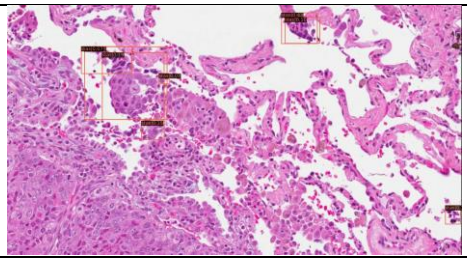
【透過案例分析潛在問題】

我們發現儘管我們嘗試了test time augmentation輔助邊緣物件的偵測，在某些的結果圖片當中，位於角落的腫瘤組織可能會獲得較低的信心，這暗示了模型可能若篩選標準再嚴格些，邊緣物件可能會被忽略。如圖5當中左下角的區域。

從圖6右上角的預測結果可以觀察出，許多預測框重疊比例高，未來或許能調整NMS threshold更細緻地篩除冗餘預測結果。

分割結果當中有一類腫瘤組織是較難分割出來的，例如圖6所示，腫瘤與附近的原有組織沾黏在一起、導致辨識效果下降。顏色上無明顯的色差也是導致分割效果大打折扣的其中一個因素。如若未來要提升辨識效果，我們認為可以針對腫瘤邊緣、沾黏嚴重的邊緣加以更詳細的標註與圖像處理。

| | |
|---|---|
|  |  |
| Train data 162 | Train data 304 |

| | |
|---|---|
|  |  |
| 圖5 | 圖6 |

semantic segmentation的模型預測的結果以藍色表示，ground truth 以綠色表示，而bounding box皆為預測結果

## 陸、雲端使用

未使用

## 柒、程式碼

https://github.com/Howard-Hsiao/AI-CUP_stas_semantic_segmentation.git，使用方式請詳閱 README.md

## 捌、使用的外部資源與參考文獻


- Solovyev, R., Wang, W., & Gabruseva, T. (2021). Weighted boxes fusion: Ensembling boxes from different object detection models. Image and Vision Computing, 107, 104117.

- Liu, Y., Wang, Y., Wang, S., Liang, T., Zhao, Q., Tang, Z., & Ling, H. (2020, April). Cbnet: A novel composite backbone network architecture for object detection. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 11653-11660).

- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 10012-10022).

- Liang, T., Chu, X., Liu, Y., Wang, Y., Tang, Z., Chu, W., ... & Ling, H. (2021). Cbnetv2: A composite backbone network architecture for object detection. arXiv preprint arXiv:2107.00420.

聯絡資料

**隊伍**

| 隊伍名稱 | Private leaderboard成績 | Private leaderboard名次 |
|---|---|---|
| TEAM_1821 | 0.88278 | 40/307 |

**隊員(隊長請填第一位)**

| 姓名<br>(中英皆需填寫) | 學校系所 | 電話 | E-mail |
|---|---|---|---|
| 蕭昀豪<br>(YUN-HAO, HSIAO) | 國立臺灣大學資訊網路與多媒體研究所 | 0963-098-836 | keepchangingtobe@gmail.com |
| 林揚昇<br>(YANG-SHENG, LIN) | 國立臺灣大學資訊工程研究所 | 0975-975-176 | jason27146913@gmail.com |
| 陳明心<br>(MING-HSIN, CHEN) | 國立成功大學人工智慧機器人碩士學位學程 | 0906-521-060 | travis.xin@gmail.com |

**指導教授**

若為「連結課程」的課堂作業或期末專題，請填授課教師，以利依連結課程彙整。

若非「連結課程」，但有教授實際參與指導，請填寫該位教授。

若以上兩者皆非，可不予填寫。

| 教授姓名 | 課程 | 課號 | 學校系所 | E-mail |
|---|---|---|---|---|
| | | | | |