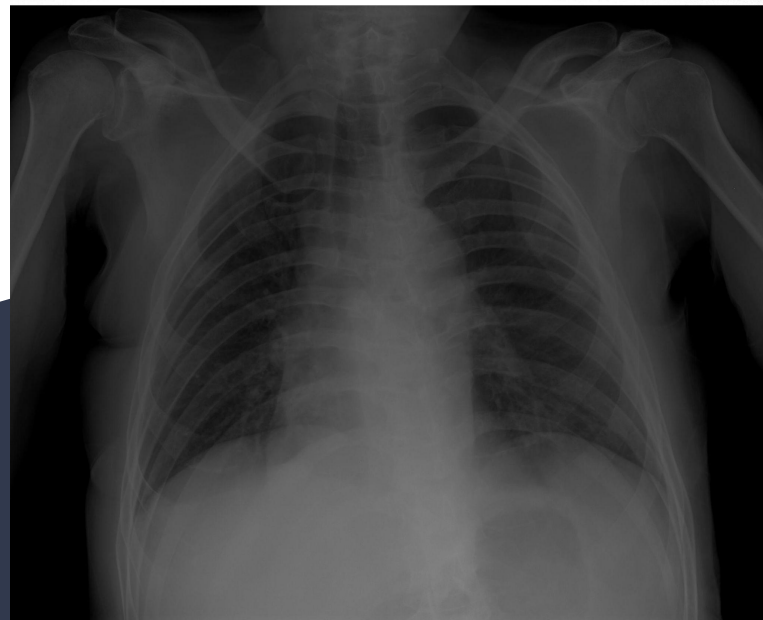


Detection of COVID-19

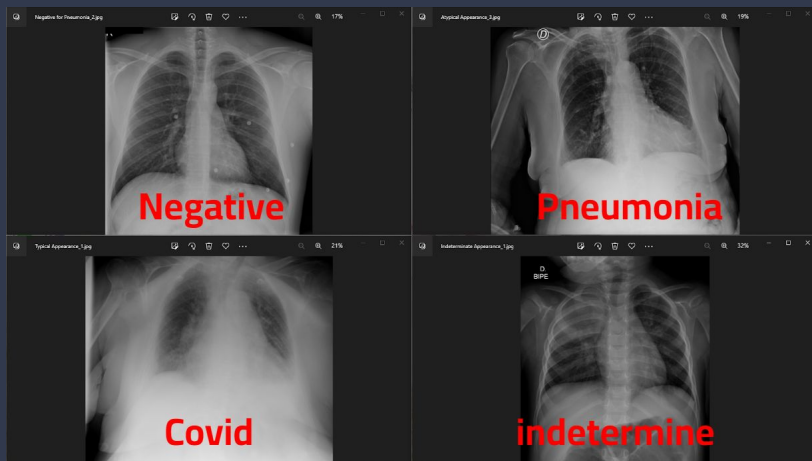
組員：李侑諭 涂崇仁
黃子軒 吳尚明



目錄

- 主題介紹
- 動機 & 目標
- 第一週 - 資料集下載 & 觀察
- 第二週 - 資料前處理
- 第三週 - 模型訓練 & 調整
- 第四週 - 模型優化 & 統整結果
- 最終成品展示
- 未來展望

主題介紹



Kaggle AI 競賽: Covid-19 Detection

The screenshot shows the Kaggle competition page for "SIIM-FISABIO-RSNA COVID-19 Detection". The page features a blue header with a chest radiograph background. The competition title is "SIIM-FISABIO-RSNA COVID-19 Detection" with a subtitle "Identify and localize COVID-19 abnormalities on chest radiographs". The prize money is listed as "\$100,000 Prize Money". The competition is organized by "SIIM Society for Imaging Informatics in Medicine (SIIM)" and has "1,305 teams" and "a year ago" of activity. The navigation menu includes "Overview", "Data", "Code", "Discussion", "Leaderboard", "Rules", and "Team". There are buttons for "Submissions" and "Late Submission".

Dataset Description

In this competition, we are identifying and localizing COVID-19 abnormalities on chest radiographs. This is an object detection and classification problem.

For each test image, you will be predicting a bounding box and class for all findings. If you predict that there are no findings, you should create a prediction of "none 1 0 0 1 1" ("none" is the class ID for no finding, and this provides a one-pixel bounding box with a confidence of 1.0).

Further, for each test study, you should make a determination within the following labels:

Files
7600 files

Size
128.51 GB

Type
dcm, csv

動機 & 目標

多做一些肺炎方面的相關研究，雖然說已目前的能力無法到在社會、醫療院所上有所作為，但多少能夠盡一份心力做貢獻 同時也能開拓視野、關心時事、投入在社會之中

目標是以我們四人的能力完成一份創新的專題出來

出發點：完全不去參考其他競賽者在這競賽方面的研究、心得 純然靠蒐集網路AI相關資料 & 隊友相互討論做出屬於我們自己的獨特作品。

從我們的角度開始出發，不會被拘束在別人成功的方法上

第一週

資料集下載 & 觀察

約有 89,700,000 項結果 (搜尋時間: 0.33 秒)

<https://www.internetdownloadmanager.com> 翻譯這個網頁

Internet Download Manager: The fastest download accelerator ✓

Internet Download Manager (IDM) is a tool to increase download speeds by up to 5 times, resume or schedule downloads and download streaming videos.

Download ✓

Registration. You may register 30-day free trial version of Internet ...

Buy Now ●

If you want to continue using IDM, you must register to get full ...

Features ✓

Explore all IDM features for 30 days for free: Download Speed ...

Language ✓

Download Internet Download Manager translation. Download ...

[internetdownloadmanager.com 的其他相關資訊](#) »

<https://zh.vpnclub.cc> 專用軟體

Internet Download Manager 下載軟體教學# IDM 中文免安裝版 ✓



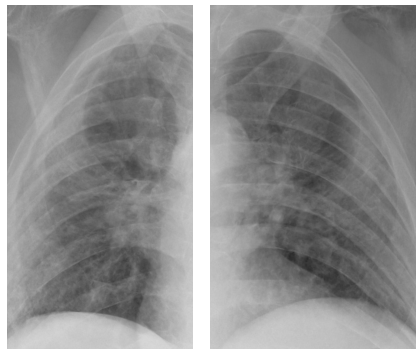
共 128.51 GB 的資料集下載, 3次失敗 平均 7~10 小時

最後成功靠 IDM 軟體 + 桌機

觀察 發現競賽中除了提供分類的csv 檔之外

還有另外提供資料細節的csv 檔, 如: 左肺右肺的 x y 座標(很精準) & X光肺部的透明度

方便我們對圖片做切割做多種的研究變化



第二週

資料前處理

Detail	Compact	Column	
A id	A boxes	A label	A StudyInstanceUID
6334 unique values	[null] 32% [{'x': 789.28836, 'y':... 0% Other (4293) 68%	none 1 0 0 1 1 32% opacity 1 789.2883... 0% Other (4293) 68%	6054 unique values
000a312787f2_image	[{'x': 789.28836, 'y': 582.43035, 'width': 1026.65662, 'height': 1917.30292}, {'x': 2245.91208, 'y':...}	opacity 1 789.28836 582.43035 1815.94498 2499.73327 opacity 1 2245.91208 591.20528 3340.5737 2352.75...	5776db0cec75
000c3a3f293f_image		none 1 0 0 1 1	ff0879eb20ed
0012ff7358bc_image	[{'x': 677.42216, 'y': 197.97662, 'width': 867.79767, 'height': 999.78214}, {'x': 1792.69064, 'y': 4...}	opacity 1 677.42216 197.97662 1545.21983 1197.75876 opacity 1 1792.69064 402.5525 2409.71798 1606.91...	9d514ce429a7
001398f4ff4f_image	[{'x': 2729, 'y': 2181.33331, 'width': 948.00012, 'height': 604}]	opacity 1 2729 2181.33331 3677.00012 2785.33331	28dddc8559b2
001bd15d1891_image	[{'x': 623.23328, 'y': 1050, 'width': 714, 'height':	opacity 1 623.23328 1050 1337.23328 2156 opacity 1 2578.56661	dfd9fdd85a3e

肺部切割:

分兩組分工, 其中一組做肺部切割& 合併

並非所有圖片都有提供x y 座標, 像是陰性不提供, 沒有異物(透明), 肺部有異物的才會給座標(有異物所以肺部不透明)



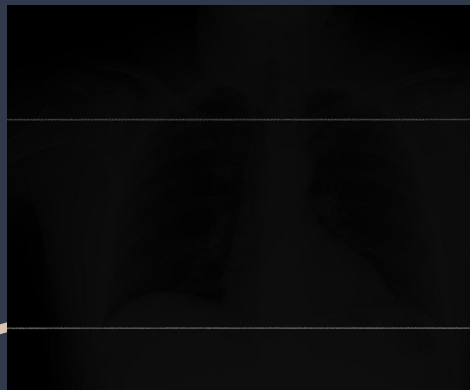
第二週

資料前處理

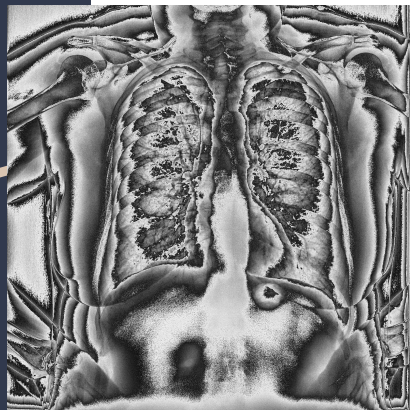
資料前處理:

轉檔 dcm to jpg (dcmtobjpg.py)

將 dcm 檔的資訊取出並將裡面的圖片轉成jpg 檔
code運作花了4 ~ 5小時 成功萃取訓練資料(3GB)



嘗試失敗



嘗試失敗



嘗試成功

第二週

資料前處理

資料前處理:

檢查損壞檔比例 (check_jpg.py)

轉檔成功:5545 份

轉檔失敗:532 份 損毀率:8.754%

資料分類 & 整合圖檔 (classify.py)

成功將類別區分於4個資料夾之中

Negative for Pneumonia/

Typical Appearance/

Indeterminate Appearance/

Atypical Appearance/

第二週

資料前處理

資料前處理:

計算 驗證資料量是否正確 (calculate.py)

將損會毀的資料移除 (remove_damage_file.py)

Negative for Pneumonia/

- 轉檔成功: 1538 份
- 轉檔失敗: 72 份

Typical Appearance/

- 轉檔成功: 2436 份
- 轉檔失敗: 307 份

Indeterminate Appearance/

- 轉檔成功: 897 份
- 轉檔失敗: 106 份

Atypical Appearance/

- 轉檔成功: 428 份
- 轉檔失敗: 32 份

第二週

資料前處理

```
transforms = T.Compose([
    #T.Resize((1024, 1024)),
    #T.RandomHorizontalFlip(p=0.5),
    #T.RandomVerticalFlip(0.5),
    #T.RandomRotation(10),
    #T.RandomEqualize(p=1),
    #Cutout(),
    #T.ToTensor(),
    #T.Normalize((0.485, 0.456, 0.406), (0.229, 0.224, 0.225)),
    #T.GaussianBlur(7, 3)
```

資料前處理:

對訓練集(dataset)做處理:

- Resize()
- RandomHorizontalFlip()
- RandomVerticalFlip()
- RandomRotation()
- RandomEqualize()
- Cutout()
- ToTensor()
- Normalize()
- GaussianBlur()

許多的方法對訓練集做前處理, 也都有相互結合使用,

至於我們成品最終會用到哪些處理方法後面會展示出來。

第三週

模型訓練 & 調整



用 kaggle 提供的編輯環境有免費的 GPU 30hr/1week

我們的程式碼需用 gpu 才跑得動

batch size, 透過batch從訓練目標中取樣, 來加快ML模型訓練的速度

設定 epochs 總共要用全部的訓練樣本重複跑幾回合

設定 learning rate 決定一次走多遠, 由此數值決定

從網路上中找到幾組對於影像辨識高評價的模型

Models: (5~6hr)

- resNet
- convnext small
- convnext large
- convnext base
- efficientNet
- mobileNet

第四週

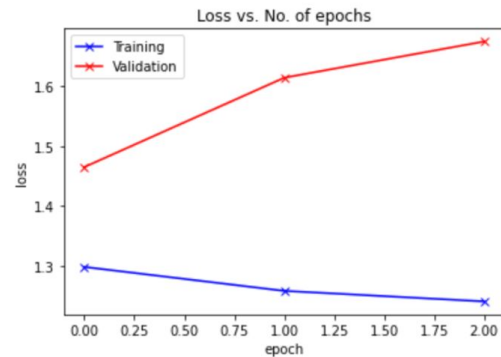
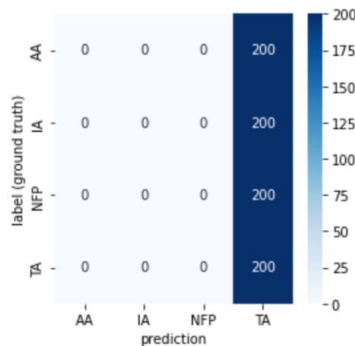
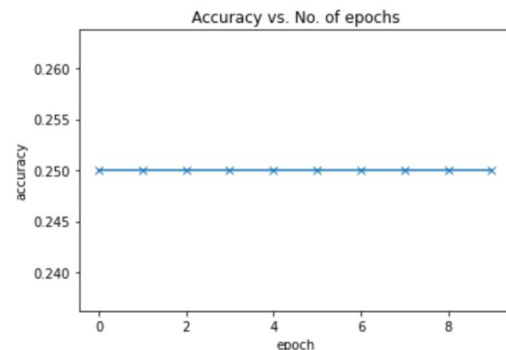
模型優化 & 統整結果

```
0%|          | 0/10 [00:00<?, ?it/s]
Epoch 1: train_loss: 1.2373, val_loss: 1.7520, val_acc: 0.2500
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:42: RuntimeWarning: invalid
['AA', 'IA', 'NFP', 'TA']
[nan nan nan 0.25]
Plot confusion matrix
10%|█        | 1/10 [22:48<3:25:19, 1368.79s/it]
Epoch 2: train_loss: 1.1777, val_loss: 1.5098, val_acc: 0.2500
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:42: RuntimeWarning: invalid
20%|██       | 2/10 [45:20<3:01:12, 1359.00s/it]
['AA', 'IA', 'NFP', 'TA']
[nan nan nan 0.25]
Plot confusion matrix
Epoch 3: train_loss: 1.1729, val_loss: 1.6162, val_acc: 0.2500
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:42: RuntimeWarning: invalid
30%|███      | 3/10 [1:07:51<2:38:06, 1355.23s/it]
Epoch 4: train_loss: nan, val_loss: nan, val_acc: 0.2500
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:42: RuntimeWarning: invalid
40%|████     | 4/10 [1:30:23<2:15:23, 1353.90s/it]
['AA', 'IA', 'NFP', 'TA']
[0.25 nan nan nan]
Plot confusion matrix
```

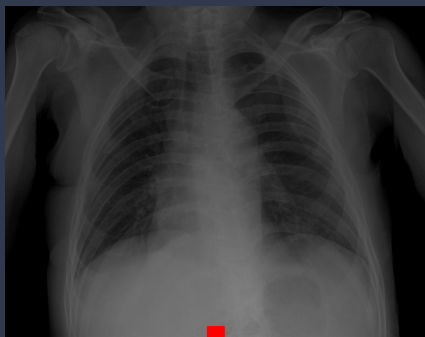
準確率:0.25% (失敗)

Convnext_small

尋找問題, 想辦法改進



最終成品展示



資料集:

- 選擇切出肺部

資料前處理:

- Resize(): 調整大小
- RandomVerticalFlip()
- Cutout(): 數據增強
- ToTensor(): 歸一化
- Normalize(): 歸一化

模型調整:

- model: efficientNet
- batch size: 16
- epochs: 10

最終成品展示

準確率:0.5887% (完成品)

Rank = 627 / 1305

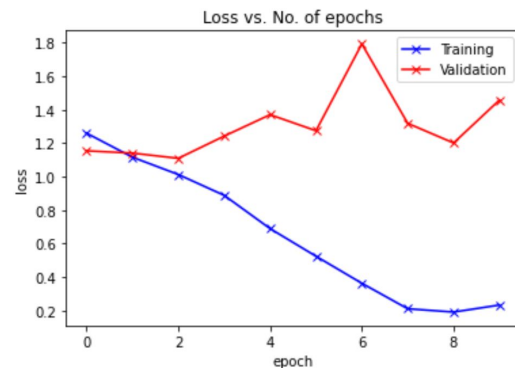
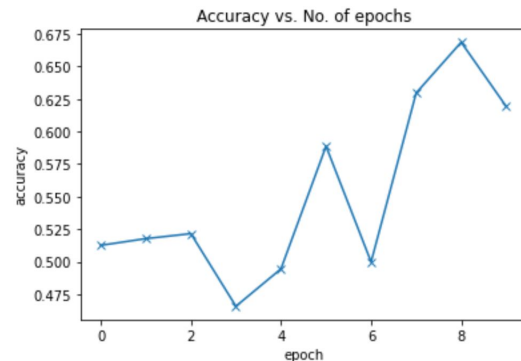
50%|███████ | 5/10 [1:58:50<1:59:12, 1430.53s/it]

```
['AA', 'IA', 'NFP', 'TA']
```

```
[0.36842105 0.26158382 0.67971759 0.84024896]
```

```
Plot confusion matrix
```

```
Epoch5: train_loss: 0.5244, val_loss: 1.2744, val_acc: 0.5887
```



未來展望

Prize Contenders

#	Team	Members	Score	Entries	Last	Solution
1	DungNB-VinBigdata-Vindr.ai		0.658	262	1y	
2	RTX 4090		0.654	239	1y	
3	Ayushman Nischay Shivam		0.653	249	1y	
4	[Aillis] Yuji & Ian		0.650	188	1y	
5	[dsmikz] School Zerde		0.648	151	1y	
6	A Team		0.648	170	1y	
7	LvLanLanLanLv		0.647	318	1y	
8	Guanshuo Xu		0.647	22	1y	
9	Watercooled		0.646	85	1y	
10	CGG		0.646	136	1y	
11	Train4Ever		0.645	353	1y	

開始去看看其他競賽者的研究方式，了解到雙方差異性

集成學習(Ensemble learning)

使用多種學習算法來獲得比單獨使用任何單獨的學習算法更好的預測性能。