

Chest X-ray image classification

Machine learning and biomedical application Group 6

林睿騰、吳定霖

1. 實驗數據集

使用 Kaggle 上的 Random Sample of NIH Chest X-ray Dataset.

Link: <https://www.kaggle.com/datasets/nih-chest-xrays/sample>

2. 實驗環境

在 Python 3.11 上建立環境，主要使用的函式庫包括：

Albumentations

Datasets

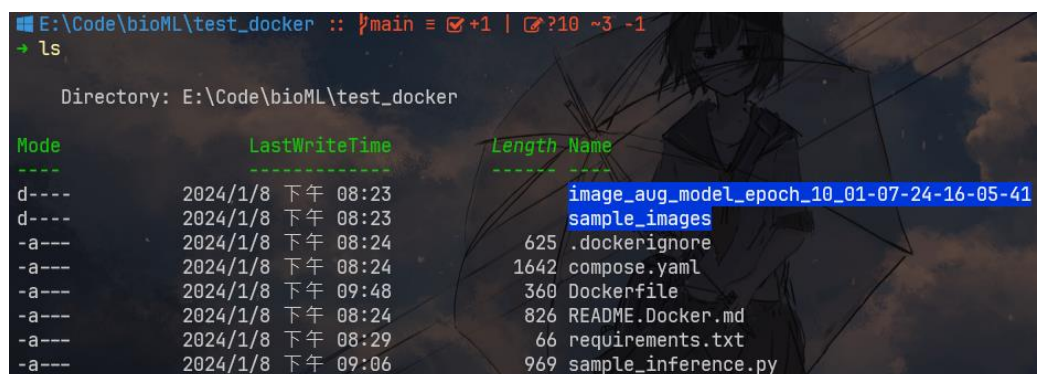
Transformers

PyTorch

剩餘的列在 requirements.txt 中。

3. 錄影與實驗檔案

- Google Drive: https://drive.google.com/drive/folders/1DPhjqKq8-S7bv4T7yffWJMfmcClez36?usp=drive_link
- 雲端上有報告、實驗錄影與一個建置 docker 和範例實驗檔案的壓縮檔（壓縮檔內的 requirements.txt 是運行 inference 的最低需求）。
- 建置與運行 Docker：
 - 解壓縮 test_docker.zip
 - 確認檔案無缺少並且在同一層



```
E:\Code\bioML\test_docker :: ?main = +1 | ?10 ~3 -1
→ ls

Directory: E:\Code\bioML\test_docker

Mode                LastWriteTime         Length Name
----                -
d-----          2024/1/8 下午 08:23             image_aug_model_epoch_10_01-07-24-16-05-41
d-----          2024/1/8 下午 08:23             sample_images
-a----          2024/1/8 下午 08:24              625 .dockerignore
-a----          2024/1/8 下午 08:24             1642 compose.yaml
-a----          2024/1/8 下午 09:48              360 Dockerfile
-a----          2024/1/8 下午 08:24              826 README.Docker.md
-a----          2024/1/8 下午 08:29              66 requirements.txt
-a----          2024/1/8 下午 09:06             969 sample_inference.py
```

- 在此資料夾運行 `docker build -t image .`，由於有 PyTorch 等 requirements，故要較長時間
- 運行 `docker run --rm -v ${pwd}/res:/tmp/res image`，完成後將會在 res 資料夾中看到結果圖。

4. 使用模型與預處理

模型使用 Huggingface.co 上預訓練過的 google/vit-base-patch16-224，資料預處理包括將圖片縮放成 224x224，轉換成 PyTorch Tensor，並將其的像素標準化。

5. 實驗參數

Learning rate: 5e-5,

Train batch size: 16,

Gradient accumulation steps: 4 (這相當於對每個 16x4 的 batch 做一次 gradient descent),

Eval batch size: 16,

Num train epochs: 3

Warmup ratio: 0.1

6. Image Augmentation

每張圖片分別經過以下的程序生成：

- 從原本的圖片中抽取一張
- 20%的機率加上隨機的亮度差
- 20%的機率隨機調整 gamma
- 20%的機率加上高斯噪音

7. 實驗結果

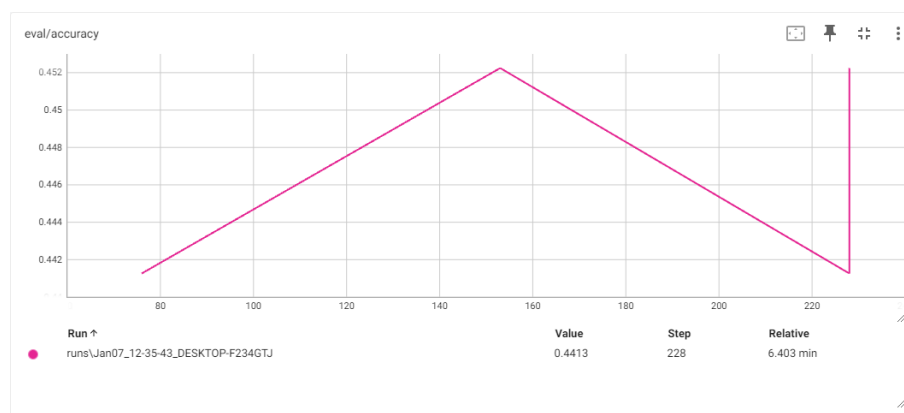
以下 accuracy 是以 top1/top3/top5 排序。

- 使用原始數據集，沒有使用過量取樣：

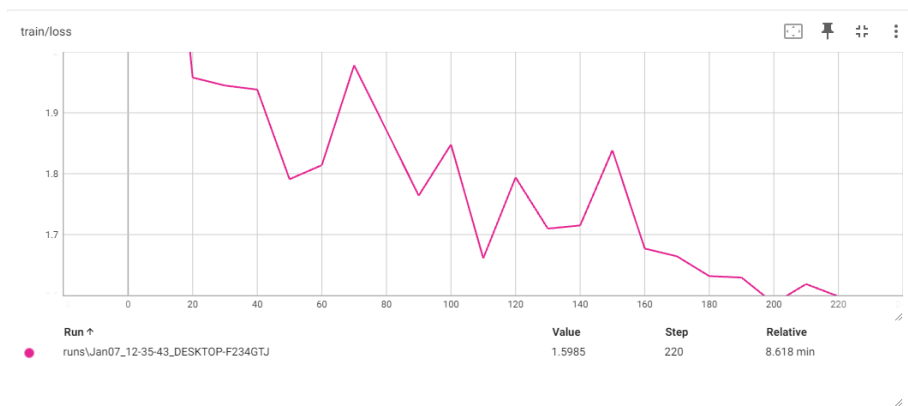
Evaluation accuracy: 45.2% / 68.4% / 80.7%

Total accuracy: 48.4% / 71.5% / 82.8%

Evaluation accuracy plot:



Training loss:



Class-wise accuracy:

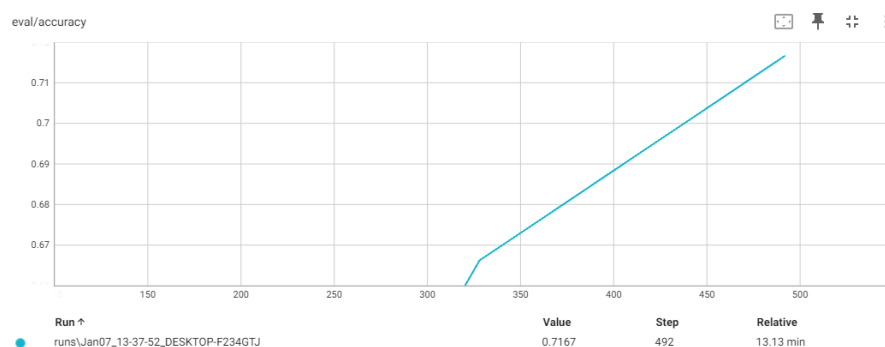
label	top_1	top_3	top_5
Healthy	0.93364	0.994415	0.999343
Atelectasis	0.137795	0.724409	0.937008
Consolidation	0.013274	0.128319	0.530973
Infiltration	0.227508	0.814891	0.989659
Pneumothorax	0.099631	0.420664	0.811808
Edema	0	0.050847	0.271186
Emphysema	0	0	0.03937
Fibrosis	0	0	0
Effusion	0.332298	0.849379	0.978261
Pneumonia	0	0	0
Pleural_Thickening	0	0.005682	0.028409
Cardiomegaly	0	0	0.021277
Nodule	0	0.185304	0.495208
Mass	0.010563	0.183099	0.489437
Hernia	0	0	0

- b. 使用原始數據集，並取樣每個分類的數據各 1000 張：

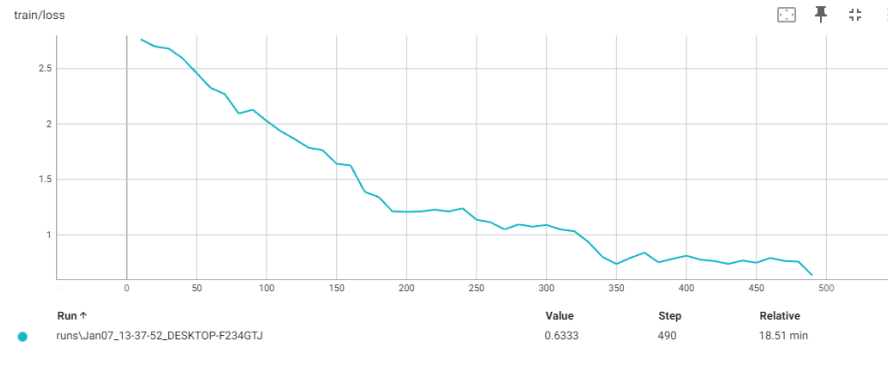
Eval accuracy: 71.6% / 91.8% / 96.4%

Total accuracy: 55.6% / 82.6% / 91.2%

Evaluation accuracy plot:



Training loss:



Class-wise accuracy:

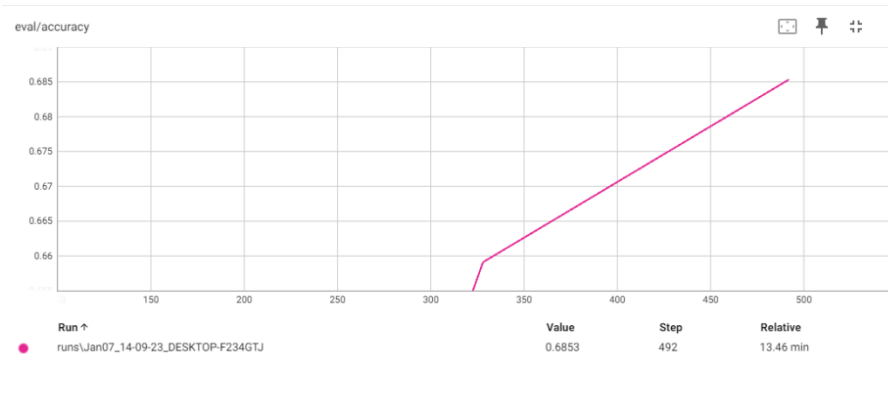
label	top_1	top_3	top_5
Healthy	0.586071	0.786794	0.874507
Atelectasis	0.496063	0.830709	0.931102
Consolidation	0.69469	0.924779	0.973451
Infiltration	0.302999	0.760083	0.914168
Pneumothorax	0.638376	0.926199	0.96679
Edema	0.855932	1	1
Emphysema	0.858268	0.96063	0.992126
Fibrosis	0.988095	1	1
Effusion	0.368012	0.804348	0.92236
Pneumonia	0.919355	1	1
Pleural_Thickening	0.75	0.971591	0.988636
Cardiomegaly	0.829787	0.985816	0.985816
Nodule	0.638978	0.881789	0.932907
Mass	0.612676	0.897887	0.940141
Hernia	1	1	1

- c. 將本來的圖片經過 Image augmentation，每個類別各生成 1000 張圖片：

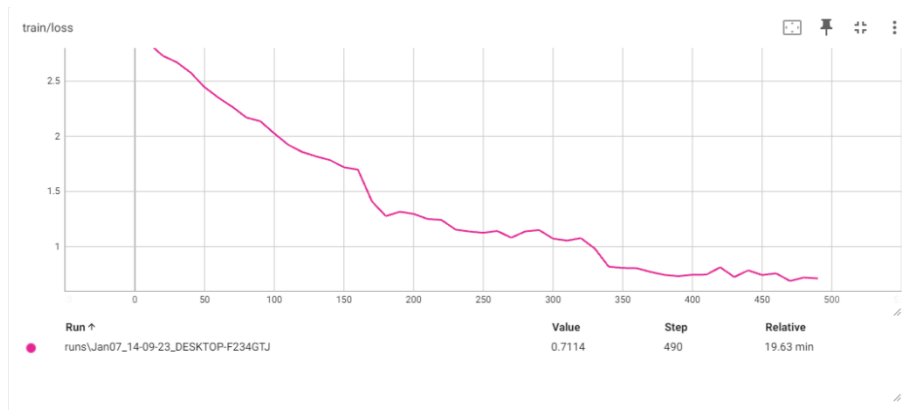
Evaluation accuracy: 68.5% / 90.0% / 94.5%

Total accuracy: 54.9% / 81.6% / 90.2%

Evaluation accuracy plot:



Training loss:



Class-wise accuracy:

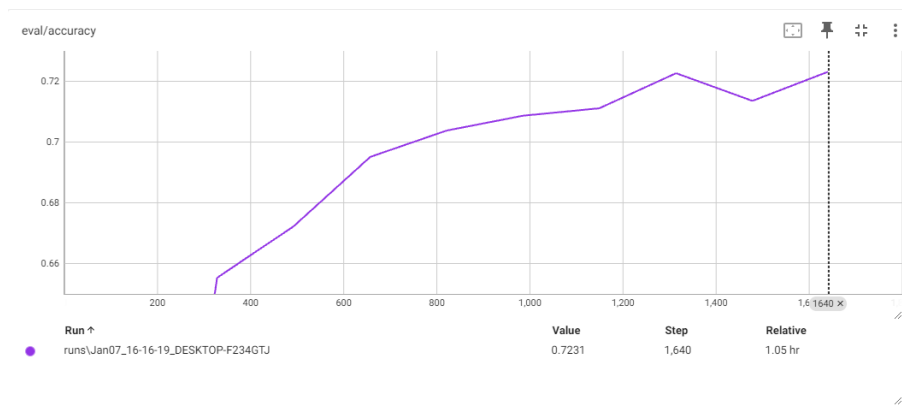
label	top_1	top_3	top_5
Healthy	0.5841	0.77431	0.861038
Atelectasis	0.438976	0.793307	0.911417
Consolidation	0.668142	0.90708	0.946903
Infiltration	0.322647	0.763185	0.908997
Pneumothorax	0.653137	0.918819	0.933579
Edema	0.805085	0.991525	0.991525
Emphysema	0.858268	0.992126	0.992126
Fibrosis	0.928571	1	1
Effusion	0.361801	0.77795	0.908385
Pneumonia	1	1	1
Pleural_Thickening	0.721591	0.954545	0.977273
Cardiomegaly	0.822695	0.978723	0.978723
Nodule	0.600639	0.878594	0.939297
Mass	0.605634	0.929577	0.971831
Hernia	1	1	1

d. 同 b.，但訓練 10 個 epoch：

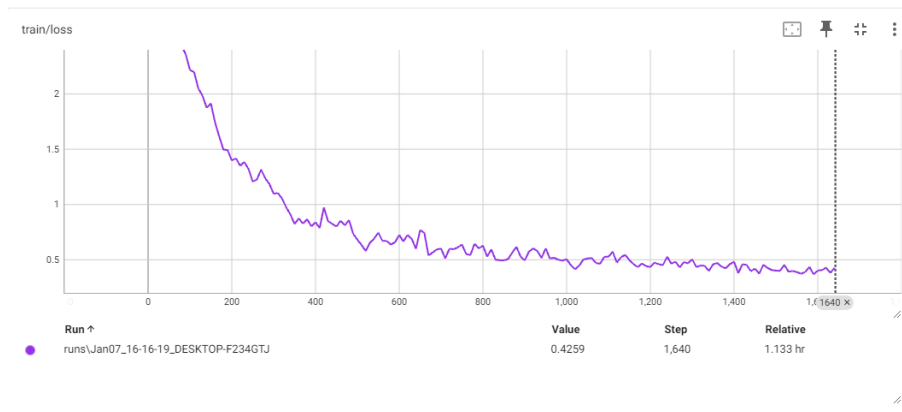
Eval accuracy: 72.3% / 93.8% / 96.8%

Total accuracy: 59.1% / 83.0% / 89.9%

Evaluation accuracy plot:



Training loss:



Class-wise accuracy:

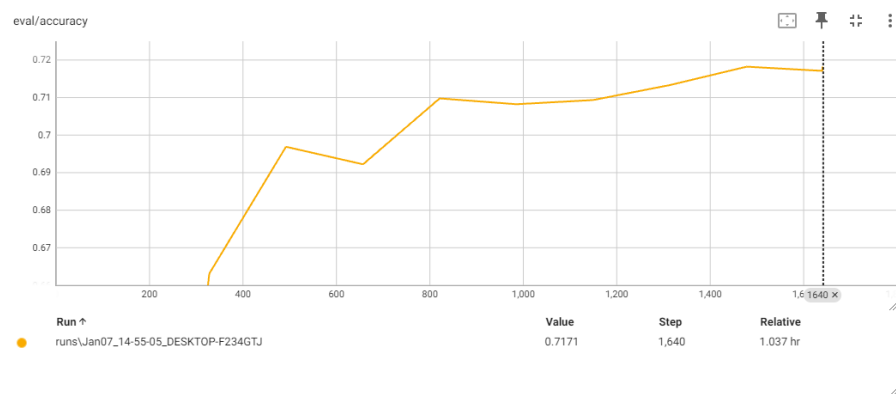
label	top_1	top_3	top_5
Healthy	0.609396	0.747043	0.818003
Atelectasis	0.529528	0.905512	0.954724
Consolidation	0.70354	0.951327	0.973451
Infiltration	0.405377	0.781799	0.936918
Pneumothorax	0.686347	0.9631	0.98155
Edema	0.822034	1	1
Emphysema	0.889764	1	1
Fibrosis	0.928571	1	1
Effusion	0.392857	0.852484	0.948758
Pneumonia	0.983871	1	1
Pleural_Thickening	0.789773	0.988636	1
Cardiomegaly	0.829787	1	1
Nodule	0.683706	0.952077	0.971246
Mass	0.630282	0.929577	0.971831
Hernia	1	1	1

e. 同 c.，但訓練 10 個 epoch：

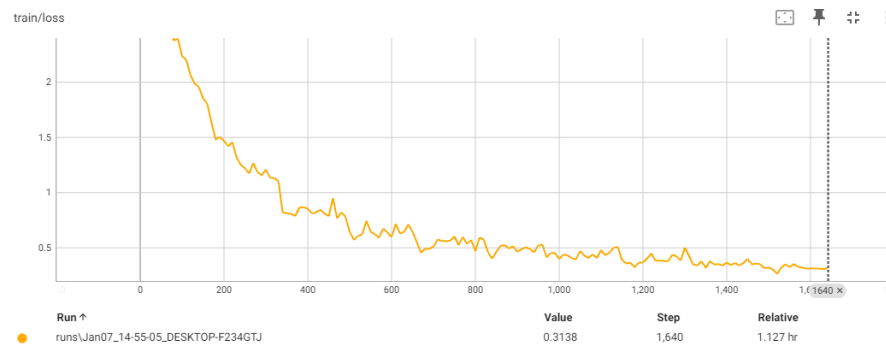
Eval accuracy: 71.8% / 92.7% / 96.2%

Total accuracy: 60.8% / 84.2% / 90.9%

Evaluation accuracy plot:



Training loss:



Class-wise accuracy:

label	top_1	top_3	top_5
Healthy	0.668857	0.787779	0.859067
Atelectasis	0.559055	0.852362	0.927165
Consolidation	0.650442	0.942478	0.982301
Infiltration	0.354705	0.800414	0.921406
Pneumothorax	0.638376	0.915129	0.95203
Edema	0.838983	1	1
Emphysema	0.84252	0.992126	1
Fibrosis	0.940476	1	1
Effusion	0.381988	0.833851	0.923913
Pneumonia	0.967742	1	1
Pleural_Thickening	0.795455	0.982955	0.988636
Cardiomegaly	0.822695	1	1
Nodule	0.686901	0.920128	0.948882
Mass	0.669014	0.961268	0.985915
Hernia	1	1	1

8. 討論

- 從 Class-wise accuracy 表格中可以看到，數據量較少的 label 如 Hernia, Edema 等，幾乎不會出現在前五的預測中，他們的 Top-5 accuracy 都接近 0。故處理不平衡資料集是個重要也幫助很大的步驟。
- 過量取樣和 Image Augmentation 的效果不相上下，因為擔心過度 Augmentation 會破壞某些 label 的特徵，故方法與參數的選擇較為保守，使他們有類似的效果。
- 因這次使用的 ViT 特徵擷取層是訓練在一般的圖片集上，或許無法完美擷取 X 光照片的特徵，使表現沒辦法更好。

9. 結語

這次使用簡單的 Vision Transformer 試圖訓練一個能夠透過 X 光照片判

斷出病症的模型，雖然只有約六成的機率能在十五個標籤中，輸出一個正確答案；但若採用前三個答案，則正確率可以到達 **85%**。故我認為 **Vision Transformer** 是能夠在 X 光判斷方面提供協助的，並且若有專門訓練過的特徵擷取，我預估會有更好的效果。