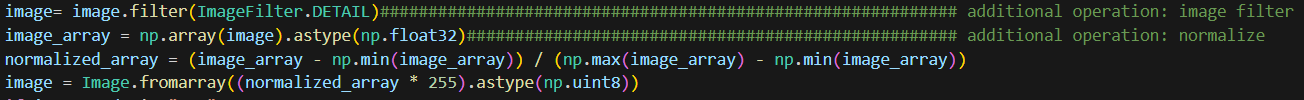
1. Data preprocessing analysis

What adjustment have you done to the Chest X-ray Images or the Reports? (20 pts)

Have you performed any additional operations to filter out noise or extract key information? (10 pts)

針對report進行tokenize、對image套用強化細節的filter後再進行normalize





1. Model & Training Method (20 pts)

What models do you choose? Why? (Please introduce the visual model and language model respectively.) (10 pts)

Encoder使用的是被肺炎X光照數據集finetune過的Vit(vision transformer)

使用這個模型的原因是因為考量到任務，使用類似資料訓練過的模型可以降低訓練成本。

Decoder 則是Bert模型(case sensitive)

會使用case sensitive版本是因為考量到report的格式，如果模型能考慮大小會比較好。

Your evaluation scores during training. (10 pts)



1. Analysis (30 pts)

Have you encountered any difficulties? How did you address them? Unlimited (10 pts)

問題一:

一開始本來打算使用LoRA來微調，但是hugging face 的 LoRA method不支援vision encoder decoder model。

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

問題二:

在將tokenizer換成bert專用的tokenizer時因為bos token的id設定不好導致生出來的text多了”##”

['##Chest PA view shows : Impression : - Suspicious pulmonary ed']

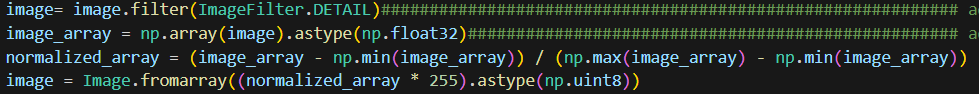
['##Chest PA view : Impression : - Increased both lung markings.']

['##Chest AP view showed : 1. s / p sternotomy and cardiac']

Solution: 將bos\_token\_id設定成400404

問題三:

在前處理的階段因為使用的API是pillow，所以最好透過pillow內建的工具完成想做的前處理

Solution:

問題四:

Vision encoder decoder model在generate時需要設定bos token id(sample code 沒有這個問題)

Solution: 研究過後發現vision encoder decoder model的config有兩種:

model.config 、 model.generation\_config。在這裡是因為generation config的部分沒有設定所以出問題。



備註:也許openai的gpt2有預先設置這個參數，bert沒有所以出現了這個問題

Sample code的rouge socre這麼慘也有可能是因為這樣，因為model.config跟model.generation\_config的參數不一樣，導致loss有在降，但是eval還是不好。