**深度學習期末報告**

1. **動機**

現在深度學習模型日益成熟的當下，日常生活中已經有許多應用，尤其是影像辨識的部分，更是表現得特別出色，因此將醫學影像和深度學習結合起來，可以有效的幫助醫生在臨床上做更快速的診斷，以減少醫師的負擔；此報告用的資料為義大醫院提供的胸部X光照，且有標記7種疾病，模型的目標是標記胸部X光照的病灶位置，且辨識出為何種疾病。

1. **資料介紹**

義大醫院提供的資料為胸部X光照(image)和標記病灶的位置(mark)，有標記的疾病分別為心臟肥大、主動脈硬鈣化、主動脈彎曲、肺尖肋膜增厚、肺野浸潤增加、胸椎退化性關節病變、脊椎側彎7種疾病，各疾病的胸部X光照和標記如圖一所示(包含正常人的資料normal)

一張含有 X 光影像, 醫學影像, 放射學, 醫學射線照相 的圖片

自動產生的描述

圖一:胸部X光照(image)和標記病灶的位置(mark)

1. **資料前處理**

第一步要處裡的是intensity log-transformation 跟 simplest color balance algorithm，可以看到我們影像資料都會有標記左右的L，而這兩個轉換可以將L的亮度調低，圖二即為轉換前後的差別(以ID為TDR02\_20161209\_161439的病患為例)。

一張含有 文字, 螢幕擷取畫面, 軟體, X 光影像 的圖片

自動產生的描述

圖二:影像資料處理前後的比較

再來我們希望將image和mark合併，使得一張圖就可以表達疾病位置，圖三為心臟肥大合併後的樣子

一張含有 文字, 螢幕擷取畫面, 軟體, X 光影像 的圖片

自動產生的描述

圖三:將心臟肥大的病灶框出來

1. **模型介紹-faster R-CNN**

接著我們將處理合併過後的影像資料丟入模型中，模型選用的是faster R-CNN，它的優點是準確率相比於大部分的模型表現是十分優異的，但取而代之的是他的架構十分複雜，因為有額外的RPN模塊以及需要兩階段的預測，因此訓練速度會比YOLO或SSD比起來還要慢，圖四是來自論文Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[1] 的faster R-CNN的模型架構:

一張含有 文字, 螢幕擷取畫面, 網站, 軟體 的圖片

自動產生的描述

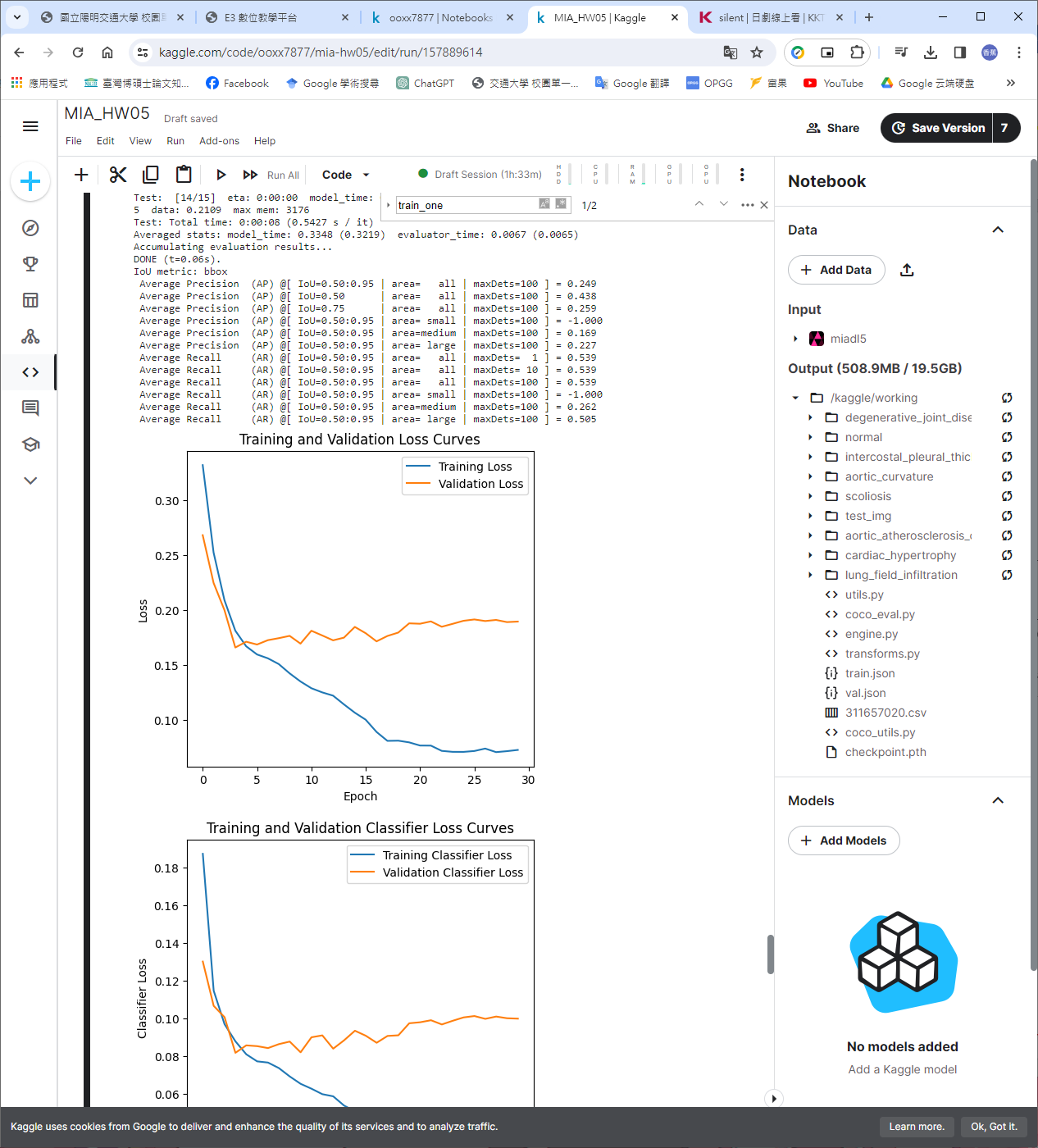
圖四: faster R-CNN的模型架構

其中conv layers是由卷積層(convolution)、relu、池化層(pooling)組成的，用於提取特徵；RPN(Region Proposal Network)的主要目的為生成潛在的物體候選區域，之後會用於物體檢測，會先由卷積層和relu提取特徵，再用Anchor為預定框，去判斷Anchor裡面有沒有物體，也就是去計算Anchor裡面有目標的概率，留下機率高的框框之後，即為候選區域，也就是圖四中的Proposals，；RoI(Region of Interest) pooling則是為了將剛剛通過RPN的候選區域整成同一個大小和形狀，也就是用一層pooling去提取RoI的feature maps；最後Classifier則是要做出最終預測了，但與前面的RPN二分類不同，這裡需要判斷前面的positive Anchor具體屬於哪一類，也就是說，Classifier就是一個多類別分類器，通過全連接層和softmax將proposals的各個類別機率算出來，再對proposal進行bounding box regression，以獲得更準確predicted box。

1. **訓練結果**

**Loss function:**

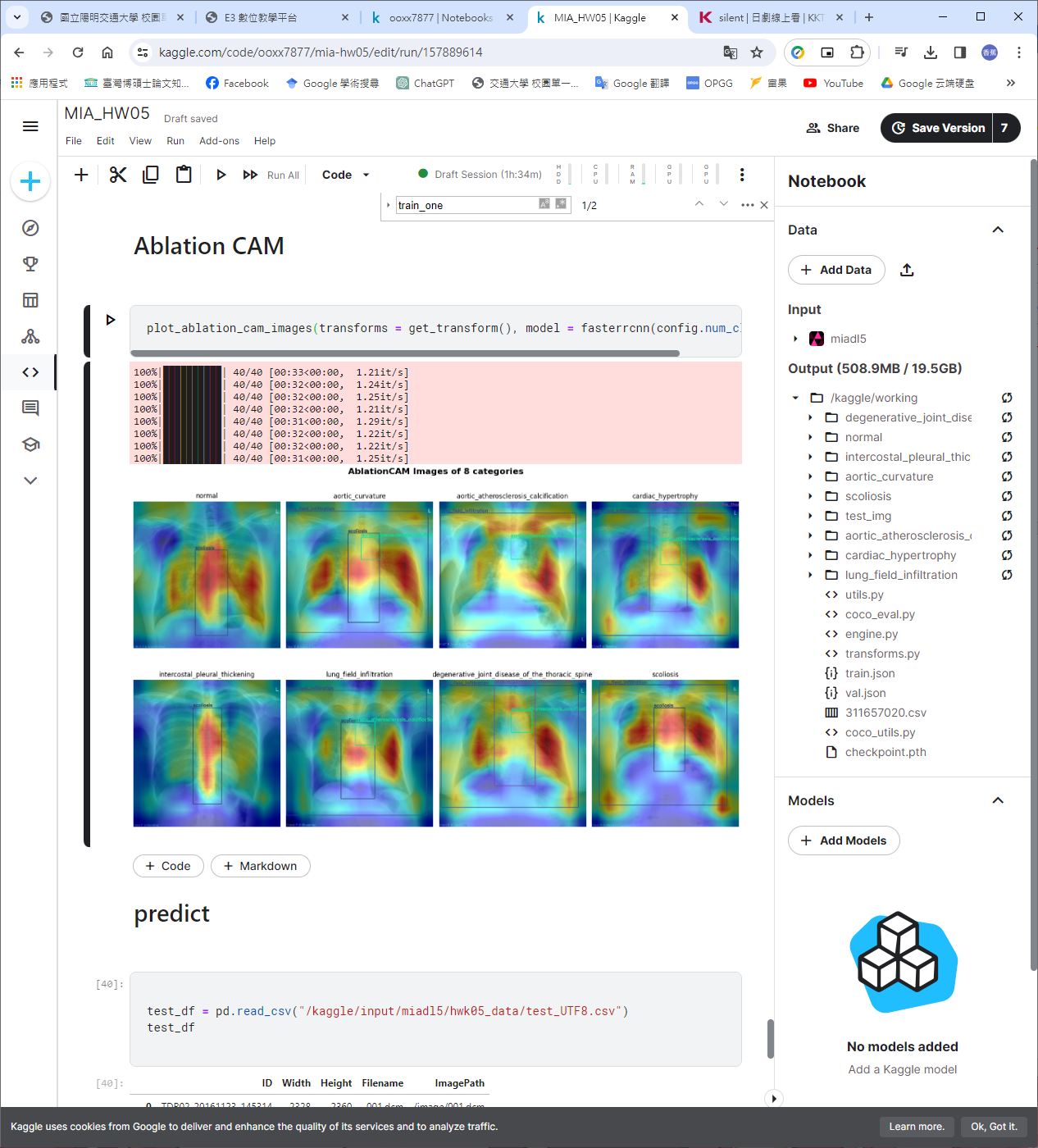
我們首先要確認的是模型該如何訓練以及到底有沒有成功訓練，在PyTorch的Faster R-CNN模型中，它會自動幫我們回傳訓練的loss，其中包含四種loss，分別是:loss\_classifier(分類器損失)、loss\_box\_reg(bounding box regressor 之損失)、 loss\_rpn\_box\_reg(RPN 中 bounding box regressor 之損失)、loss\_objectness(RPN 中分類器之損失，此分類器用以判斷 bounding box 中是否包含物體)，總損失為這四種loss的總和，我們將訓練目標設為找出最小的Validation Loss，為了檢查模型是否有在收斂，我們將模型的訓練結果如圖五所示。



圖五:模型的Loss Curves

**Ablation - CAM(Class Activation Map)**:

深度學習最令人詬病的一點是，模型就如同黑盒子一般，缺乏解釋力，有時候也會有預測能力很高，但其實根本是用一些奇怪的地方在做預測，為了避免這種情況，Ablation-CAM: Visual Explanations for Deep Convolutional Network via Gradient-free Localization[2]這篇論文中提出了Ablation-CAM這個方法，他的想法是消融 activation 並測量輸出的變化，如果輸出大幅下降，表示該 activation的重要性較高，由此去視覺化深度學習模型的重要程度，我們模型的視覺化結果如圖六所示。



圖六: Ablation-CAM

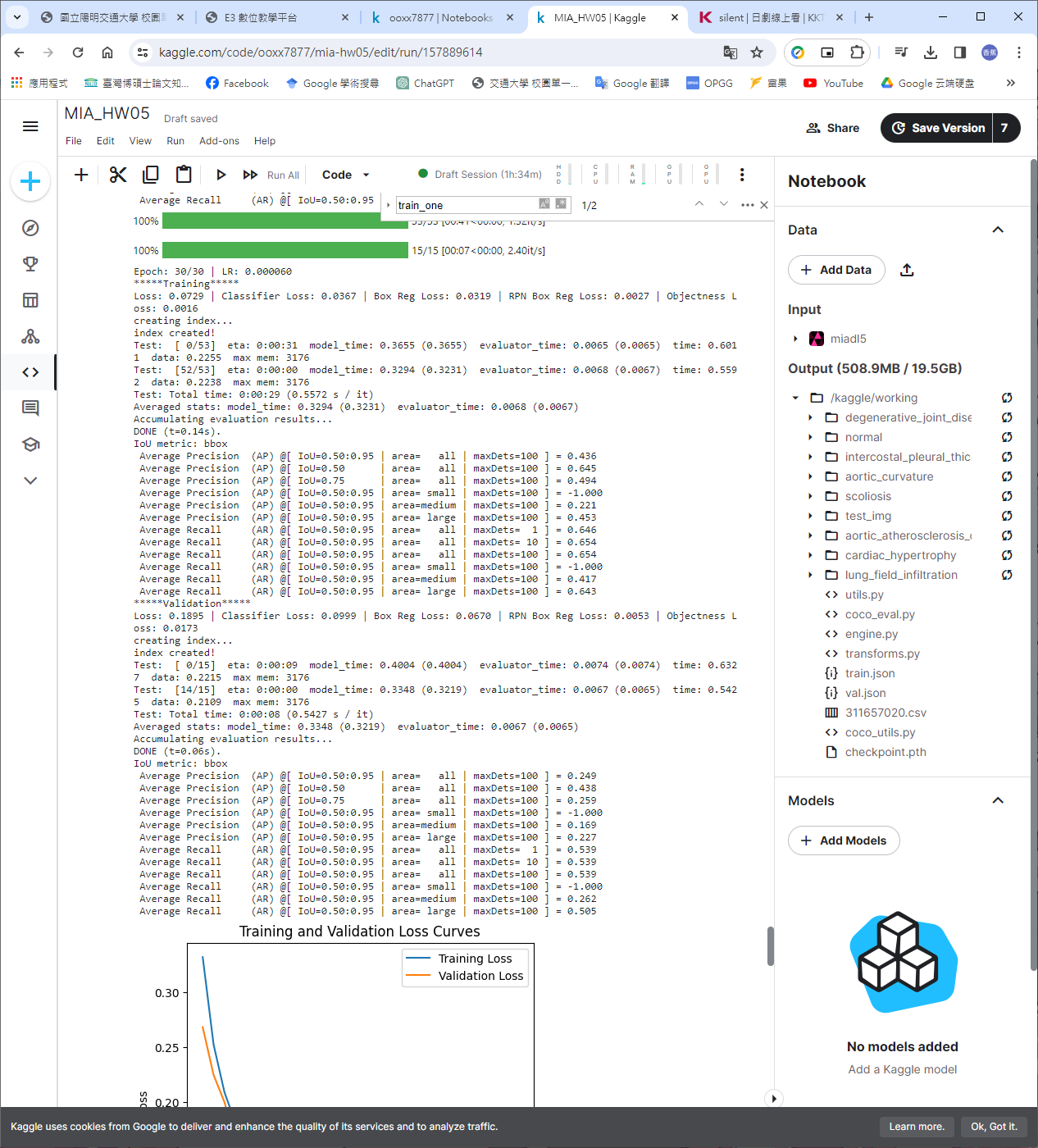
**IoU(Intersection over Union)及mAP(Mean Average Precision)**:

確認模型有正確訓練後，我們想知道的問題依舊是這個模型能不能正確預測出病灶位置，首先我們會先確認有沒有框到正確的病灶，也就是用IoU去衡量，數學式如圖七所示，其中 和 為預測區域和目標區域，之後在看框對幾個目標除上模型共有幾個預測目標即為Precision，在訓練過程中，我們也可以實時監控各IoU下的mAP，如圖八所示，我們的目標是希望能在IoU=0.5的情況下，達到0.2以上的準確率。

一張含有 文字, 螢幕擷取畫面, 軟體, 電腦圖示 的圖片

自動產生的描述

圖七:IoU的算法



圖八:各個條件下的mAP

1. **結論**

此次研究主要的目的是希望能借助深度學習中優秀的物件偵測能力，來幫助醫師做出診斷，而準確率較高的faster-R-CNN自然就是我們的首選模型，可以看到我們的loss function有穩定的在收斂，以及CAM圖也可以看出來模型是對正確地方進行訓練，最後的預測結果有達到標準，但還是有許多可以改進的地方，像是希望模型訓練的部位再準確一些、提高訓練速度、或是能否再提高mAP，都是未來可以研究的方向。

**參考文獻**

1. Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems 28 (2015).
2. Ramaswamy, Harish Guruprasad. "Ablation-cam: Visual explanations for deep convolutional network via gradient-free localization." proceedings of the IEEE/CVF winter conference on applications of computer vision. 2020.