

ML final project proposal

隊名: NTU_r07942091_test

隊員: 1.許博閔 r07942091 2.李利元 b04501073

3.陳鈞廷 b04201040 4.白佳灝 r07942092

題目:RSNA Pneumonia Detection Challenge (DeepQ challenge)

Problem study: 在 object detection 的方法中，大致可以分為 two stage detectors 和 one stage detectors 這兩類。two stage detectors 會預先選出許多可能的 bounding box 位置，再一一檢測這些 box 是否含有 object。而 one stage detectors 直接將一張圖片分成許多網格，用回歸的方法預測 box 的位置，以下是兩者的比較：

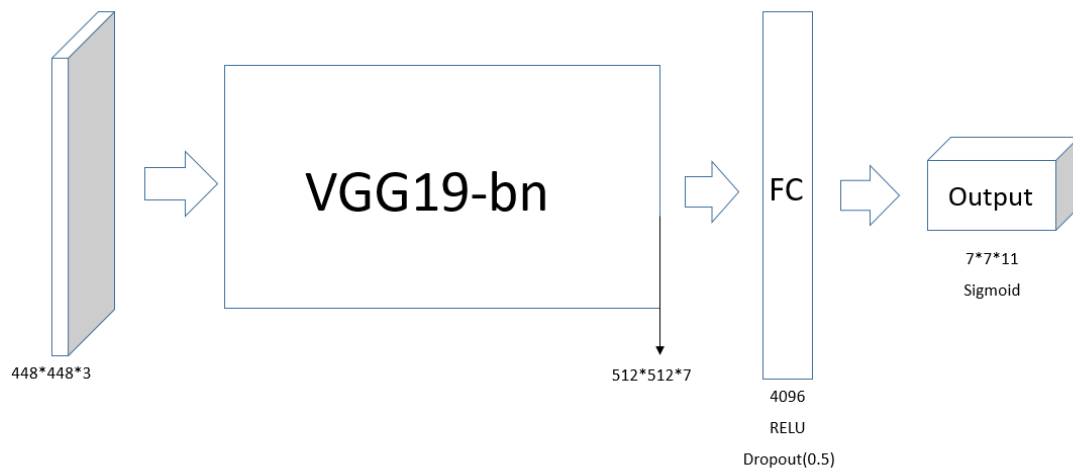
	two stage detectors	one stage detectors
常見方法	RCNN/faster RCNN 等	YOLO/SSD 等
精準度	較高	較低
速度	慢	快

目前使用 YOLOv1 的實作，因此簡要介紹 YOLOv1 的方法:

1. 將輸入圖片分成 $7*7$ 的網格，每個網格只負責預測一個 object，每個網格預測兩個 bounding box，因此總共會有 $7*7*2=98$ 個 bound box。
2. 輸出形式，預測每個 bound box 的中心位置座標，長、寬和 confidence score，和 one hot encoding 的 class label，因此 YOLOv1 的輸出維度是 $7*7*(10+N)$ ，N 代表有幾種 class，以本次 final 來說， $N=1$ 。
3. Loss function : location loss + confidence score loss + classification loss，其中 location loss 會乘上較大的權重，而沒有 object 的 confidence loss 會乘上一個較小的權重。
4. Non-maximum suppression(NMS): 用來決定最後那些 bounding box 該留下來的的方法，首先會保留 confidence score 最高的 box，並計算剩餘的 box 和分數最高 box 的 IOU(兩個 box 重疊的比例)，IOU 超過閾值的 box 就會被放棄，NMS 最一直重複這個動作直到沒有 bounding box 為止。

Proposal method: 目前用 yolov1 並配合不同的 backbone，如 VGG19_bn、Resnet50、Densenet121 等等，較令人好奇的地方是用 VGG 的分數是最好的。另外兩種 model 的表現反而比較差。

Model 架構：



訓練方法:

將圖片 resize 到 448*448

Epoch : 20

Batch size = 16

Optimizer : SGD, momentum = 0.9, weight decay = 5e-4

Learning rate : 前 10 個 epoch = 0.001 , 後 10 個 epoch = 0.0001

Data augmentation : 水平翻轉、縮放、平移

Output confidence threshold = **0.1**

NMS IOU threshold = 0.5

預計改進的發法

1. 使用精準度更高的 object detection 方法，如 yolov3、SSD、retinanet 等方法
2. 因為有許多圖片並沒有 bounding box，是否能先訓練一個 binary 的分類器，先分出那些圖片是有生病的，再將這些圖片用 object detection 的方法找出 bounding box。

Reference

1. yolov1 paper
<https://arxiv.org/abs/1506.02640>
2. https://blog.csdn.net/v_july_v/article/details/80170182
3. <https://github.com/xiongzihua/pytorch-YOLO-v1>