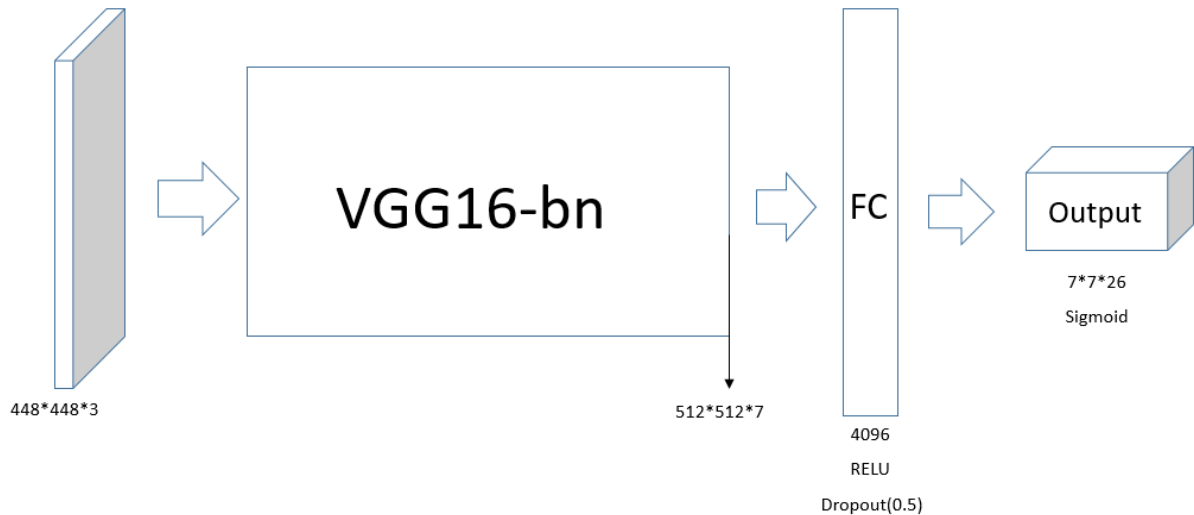1. (　5%) Print the network architecture of your YoloV1-vgg16bn model and describe your training config. (optimizer,batch size….and so on)

上圖是我的 network architecture ，將 resize 後的圖片(448*448*3)放入 VGG16-bn，再接到 4096 的 fully connected 層，經 RELU 和 Dropout 後得到 7*7*26 的 output

Training config :
Epoch : 50
Batch size = 16
Optimizer : SGD, momentum = 0.9, weight decay = 5e-4
Learning rate : 前 40 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

2. (10%) Show the predicted bbox image of "val1500/0076.jpg", "val1500/0086.jpg", "val1500/0907.jpg" during the early, middle, and the final stage during the training stage. (For example, results of 1st, 10th, 20th epoch)

Output confidence threshold = **0.1** (雖然有可能沒有 bounding box，但調太小會有很多 boinding box，因此用 0.1) ，NMS IOU threshold = 0.5
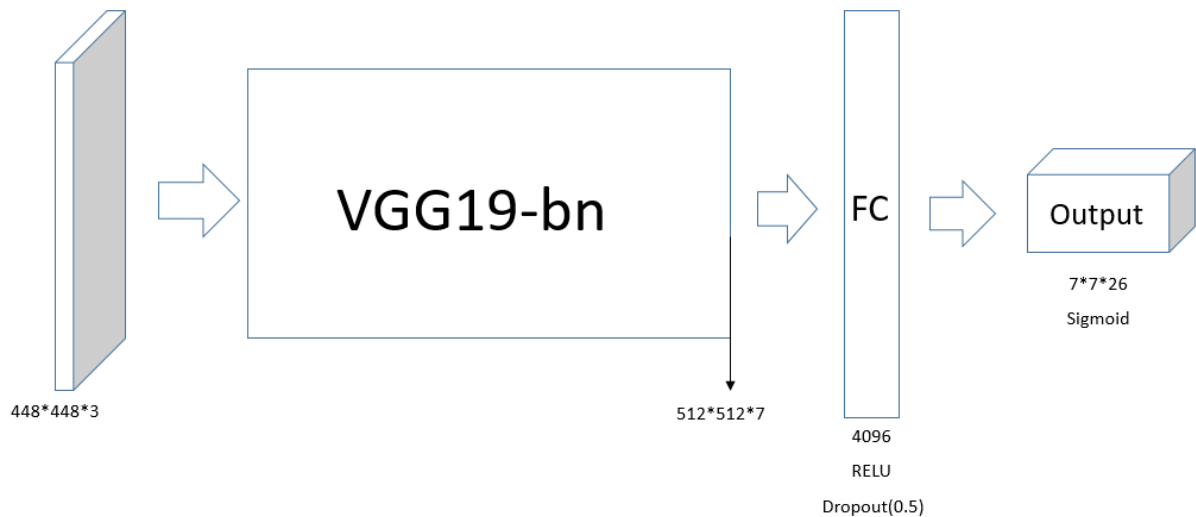
Early : 第 5 epoch
Middle : 第 20 epoch
Final : 第 40 epoch

Early epoch 的 image0907 就沒有 bounding box

id	0076	0086	0907
E a r l y			
M i d d l e			
F i n a l			

3. (10%) Implement an improved model which performs better than your baseline model. Print the network architecture of this model and describe it.

上圖是我的 improved model network architecture ，將 resize 後的圖片(448*448*3)放入 VGG19-bn，再接到 4096 的 fully connected 層，經 RELU 和 Dropout 後得到 7*7*26 的 output，和第一題相比，只差在將 backbone 改成 VGG19-bn

Training config :
Epoch : 50
Batch size = 16
Optimizer : SGD, momentum = 0.9, weight decay = 5e-4
Learning rate : 前 40 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

4. (10%) Show the predicted bbox image of "val1500/0076.jpg", "val1500/0086.jpg", "val1500/0907.jpg" during the early, middle, and the final stage during the training process of this improved model.
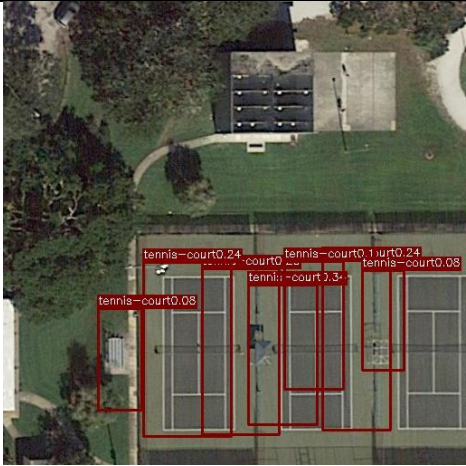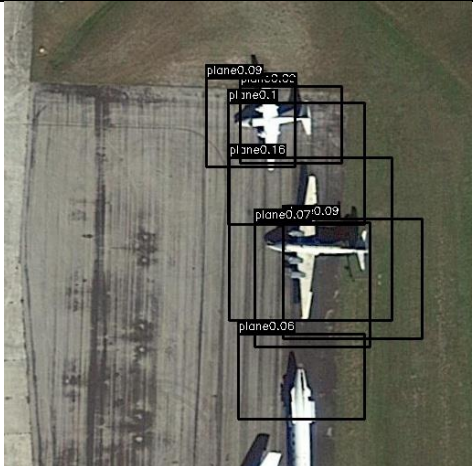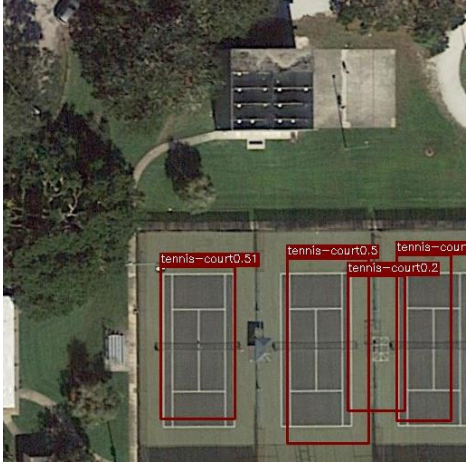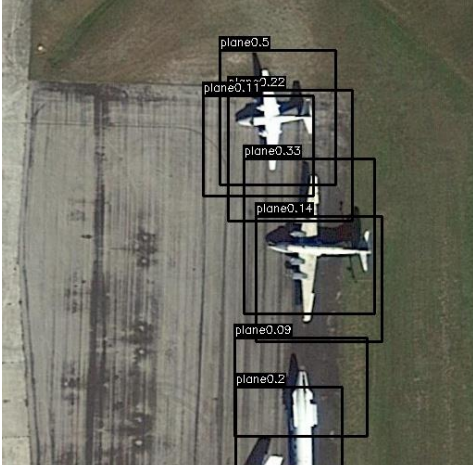
Output confidence threshold = **0.1** ，NMS IOU threshold = 0.5，和第二題相同

Early : 第 6 epoch
Middle : 第 25 epoch
Final : 第 50 epoch

Early epoch 的 image 0907 仍是沒有 bounding box

id	0076	0086	0907
E a r l y			
M i d d l e			
F i n a l			

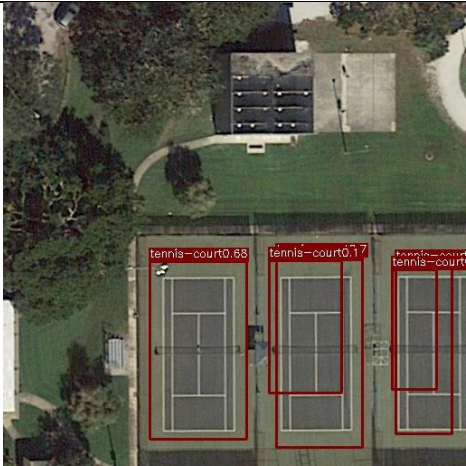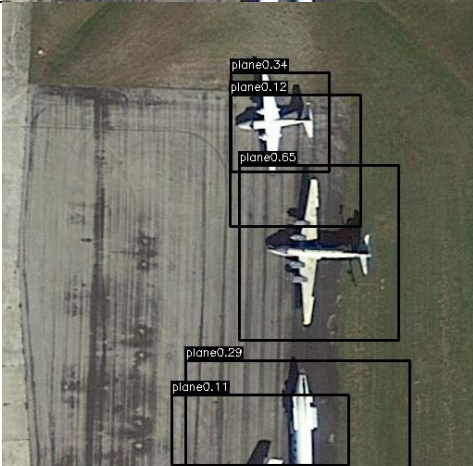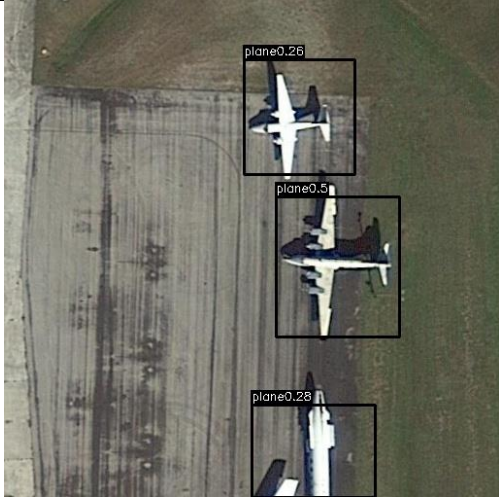5. (15%) Report mAP score of both models on the validation set. Discuss the reason why the improved model performs better than the baseline one. You may conduct some experiments and show some evidences to support your reasoning.

	Yolov1 vgg-16bn	Improved model
mAP	classname: plane ap: 0.3320707891827251 classname: baseball-diamond ap: 0.0 classname: bridge ap: 0.09090909090909091 classname: ground-track-field ap: 0.045454545454545456 classname: small-vehicle ap: 0.09090909090909091 classname: large-vehicle ap: 0.12083333333333333 classname: ship ap: 0.09090909090909091 classname: tennis-court ap: 0.5808344484956914 classname: basketball-court ap: 0.0 classname: storage-tank ap: 0.09090909090909091 classname: soccer-ball-field ap: 0.2272727272727273 classname: roundabout ap: 0.0 classname: harbor ap: 0.17337946637432006 classname: swimming-pool ap: 0.1018981018981019 classname: helicopter ap: 0.0 classname: container-crane ap: 0.0 map: 0.12158623597798801	classname: plane ap: 0.4379879191519468 classname: baseball-diamond ap: 0.0 classname: bridge ap: 0.13216991341991344 classname: ground-track-field ap: 0.045454545454545456 classname: small-vehicle ap: 0.09090909090909091 classname: large-vehicle ap: 0.15957243283432926 classname: ship ap: 0.10643626346105685 classname: tennis-court ap: 0.6563721736701885 classname: basketball-court ap: 0.09090909090909091 classname: storage-tank ap: 0.06238498343761502 classname: soccer-ball-field ap: 0.10822510822510822 classname: roundabout ap: 0.0 classname: harbor ap: 0.17167409704723138 classname: swimming-pool ap: 0.0909090909090909 classname: helicopter ap: 0.0 classname: container-crane ap: 0.0 map: 0.13456279433932547

Improved model 把 yolov1 vgg16-bn 的 Map 高了大約 1.3%
1.觀察上面表格，可以發現 vgg16-bn 的 basketball-court 的 ap = 0.0，但是 improved model 的 ap = 0.0909，代表 improved model 可以辨識出 basketball-court 而 vgg16-bn 無法，推測是因為 improved model 的 backbone 用了 vgg19-bn，model 架構較深，參數比較多，可以多處理一個 class，因此表現的比較好。

2.plane, tennis-court 的 AP 也有明顯的上升，因這兩個 class 原本在 vgg16-bn 的結果就不錯，因此進步的原因應該是 bounding box 畫的更準，或是 class specific confidence score 變高，所以結果更好。若觀察前幾題 0076.jpg 的圖片，是 improved model 的結果好一點，若把 confidence threshold 調高使重疊的 bounding box 減少會更清楚。

	Improved model	Vgg16-bn
0076		
0086		

上表是將 confidence threshold 調成 0.15 後畫出的圖，我認為比較明顯的差異就是每個 Bounding box 的 class specific confidence score，幾乎都是 improved model 的分數比較高，因此最終的 AP 的分數比較高。至於 bounding box 畫的好不好可能參雜一些個人的主觀判斷，例如助教提供的 0086.jpg 的 True label txt 畫出來的結果，並不包含圖片中最下面只有一半的飛機，但我覺得那應該也是一台飛機。

綜合上述兩點，我的 improved model 會得到較好的結果主要是因為：能夠多辨識出一種 class(basketball-court)，以及在表現好的 class 上有更高的 confidence score。

6. **bonus (5%)** Which classes prediction perform worse than others? Why? You should describe and analyze it.

從第五題的 map 的來看，ap = 0.0 的 class 如 baseball-diamond, roundabout, helicopter, container-crane 的結果比其他 class 差，推測是 training data 就已經有 class unbalanced 的問題存在，而且我在 train 的時候，一張圖最多只會有 49 個 bounding box，有些 bounding box 會被濾掉使 class unbalanced 的問題更嚴重。

	Training label 數量	Validation ap
baseball-diamond	498	0.0
roundabout	524	0.0
helicopter	250	0.0
container-crane	94	0.0
harbor	6940	0.1733
plane	7522	0.3320
tennis court	2936	0.5808

上表是計算 15000 張 training image bounding box 的數量，ap = 0.0 的 bound box 數量都很少，因此 model 無法辨識出這些 class

7. Reference
   1. pytorch yolo v1
   https://github.com/xiongzihua/pytorch-YOLO-v1
   2. IOU code
   https://blog.csdn.net/leviopku/article/details/81629492
   3.助教投影片 80 頁的 4 個網址都有參考