Video streaming and tracking - HW2

311551096 錢承

1. Experiment Setup

在訓練前，將label的格式轉換成COCO格式，也就是將class、x\_center、y\_center、width、height轉換成class、left、top、right、bottom，如下圖。



訓練時候，我包含SE於與不包含SE的模型，分別訓練了600個epochs。

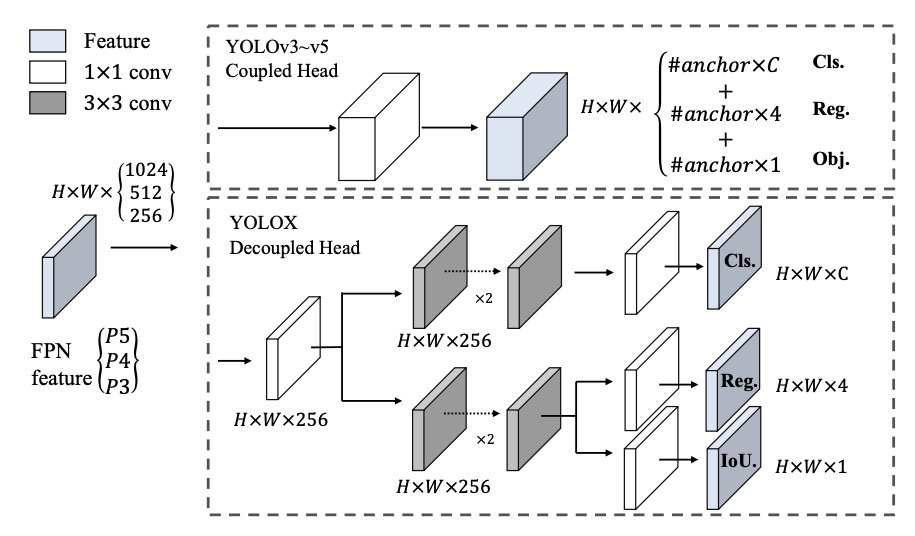
在計算mAP時候，會將label由class、x\_center、y\_center、width、height格式，轉換成class、left、top、right、bottom格式，如下圖。



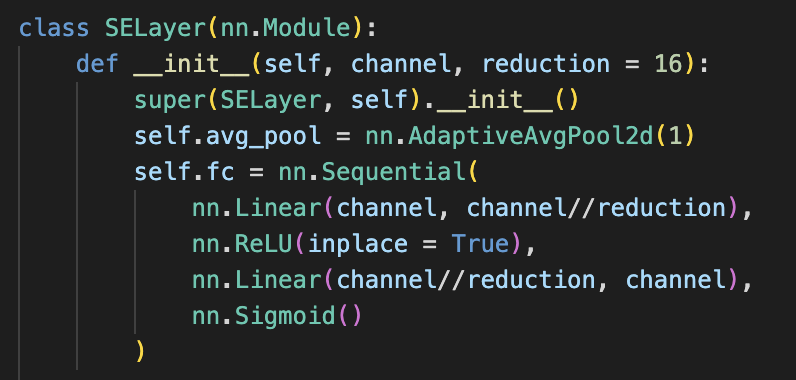
1. Brief explain your code

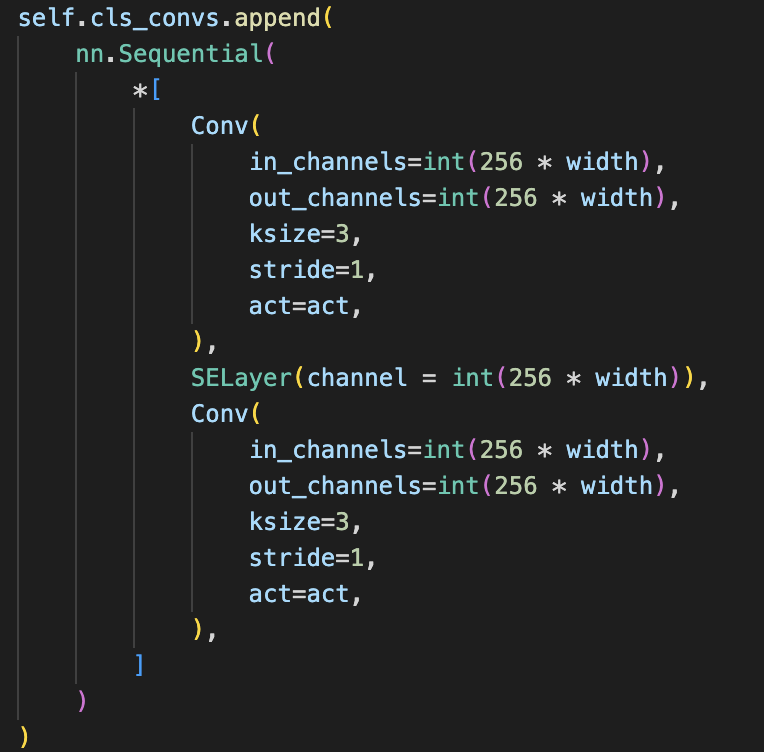
不包含SE的模型，於原本的YOLOX無異。

而包含SE的模型，我在YOLOX中Decoupleed Head加入兩個SE layer，於下圖紅圈處。



實作細節如下圖：







利用weight產生result流程如下，

（1）系統資訊：

System: Ubuntu 20.04.5 LTS

Kernel: 5.15.0-52-generic

Cuda: 11.6

GPU: NVIDIA GeForce RTX 3060 Ti

（2）環境建置：

$ sudo apt install cmake

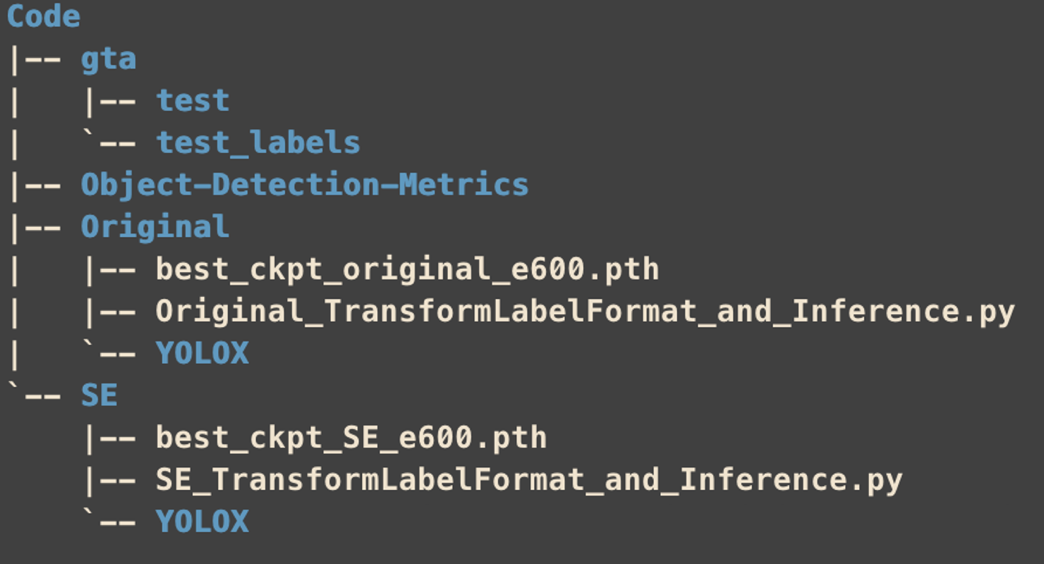
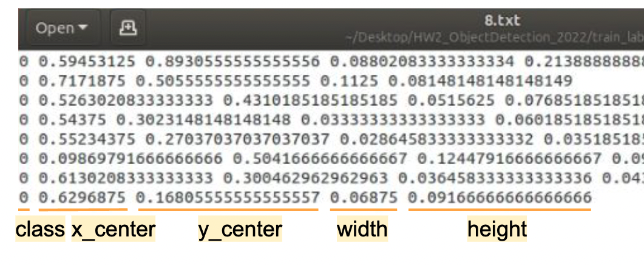
$ conda env create -n lab2 python==3.9.13

$ conda activate lab2

$ pip3 install torch torchvision torchaudio --extra-index-url <https://download.pytorch.org/whl/cu116>

\*其中重要套件版本為torch==1.13.0+cu116、torchaudio==0.13.0+cu116、torchvision==0.14.0+cu116

（3）Testing data防止位置及命名：

 將testing data放置於Code資料夾下，如下右圖。再來將testing data的資料解命名為gta，且圖片部分的資料夾命名為test，label部分的資料夾命名為test\_labels。其中label格式依序為class、x\_center、y\_center、width、height如下左圖。

（4）執行沒有加SE（Original），產生預測結果:

$ cd Code/Original/YOLOX

$ pip3 install -v -e .

$ cd Code/Original

$ python Original\_TransformLabelFormat\_and\_Inference.py

\*此結果將轉換label(groundtruths)的格式，格式依序為class、left、top、right、bottom。並且利用訓練完的weight產生預測結果，然後將結果輸出到gta/Original/detections中。

（5）執行有加SE，產生預測結果:

$ cd Code/SE/YOLOX

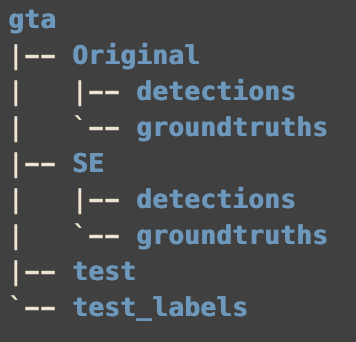
$ pip3 install -v -e .

$ cd Code/SE

$ python SE\_TransformLabelFormat\_and\_Inference.py

\*此結果將轉換label(groundtruths)的格式，格式依序為class、left、top、right、bottom。並且利用訓練完的weight產生預測結果，然後將結果輸出到gta/SE/detections中。

\*執行完（4）（5）點後，資料夾結構如下圖，預測結果放置於gta/Original/detections和gta/SE/detections。



（6）計算mAP(沒有加SE):

$ cd Code/Object-Detection-Metrics

$ python pascalvoc.py -gt ../gta/Original/groundtruths -gtformat xyrb -det ../gta/Original/detections -detformat xyrb -t 0.85

（7）計算mAP(有加SE):

$ cd Code/Object-Detection-Metrics

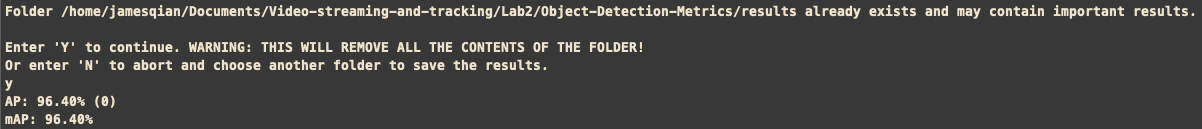
$ python pascalvoc.py -gt ../gta/SE/groundtruths -gtformat xyrb -det ../gta/SE/detections -detformat xyrb -t 0.85

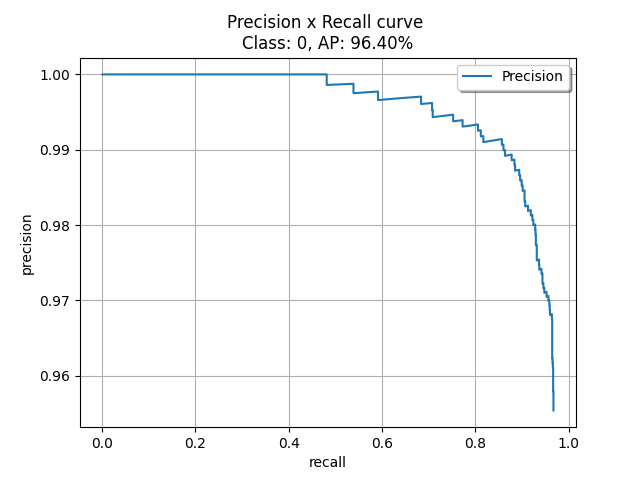
Reference: <https://github.com/Megvii-BaseDetection/YOLOX>

1. Screenshot your validation results on your two models (with/without SE module)

without SE：

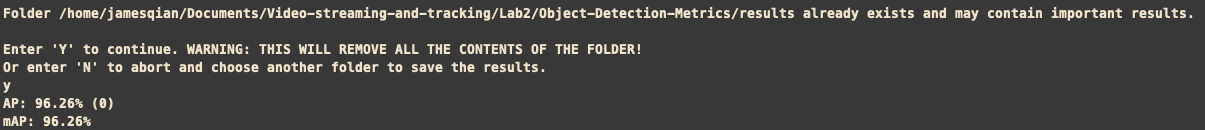
mAP: 96.40%

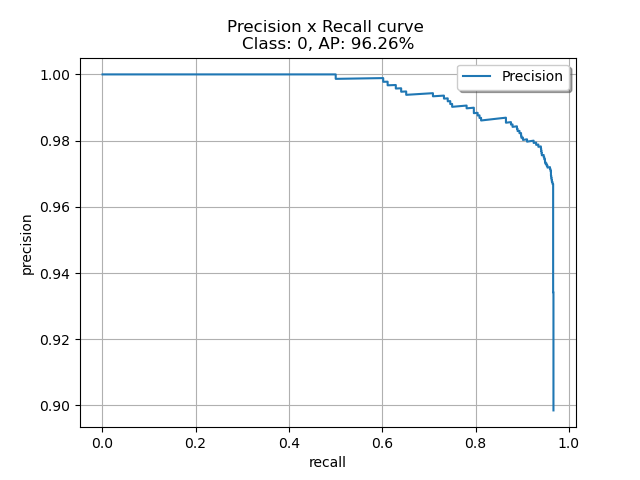




with SE：

mAP: 96.26%





1. Discussion

在實作過程中，主要遇到的困難是label的格式問題，我採用YOLOX官方的code，而官方的dataloader主要是讀取COCO格式，我因此我撰寫程式轉換作業提供的格式到COCO格式。另外一個格式問題，主要是要去Object-Detection-Metrics（https://github.com/rafaelpadilla/Object-Detection-Metrics）計算mAP的時候，groundtruths的格式是需要轉換成class、left、top、right、bottom。

 增加SE layer後，每一個bbox的準確率有提升，可是在conf. 85卻有出現錯誤的bbox，下左圖為無SE layer，下右圖為有SE layer。