定義：

TP：ground truth 為 true，predict 為 true

FP：ground truth 為 false，predict 為 true(誤報)

FN：ground truth 為 true，predict 為 false(漏報)

IoU：Intersection of Union，預測框和真實框的重疊

IoU threshold：重疊度高於此域值，則為 TP，反之為 FP，一般設為 0.5

Precision：TP/(TP+FP) 1-誤報度

Recall：TP/(TP+FN) 1-漏報度

F1-score：2 * (precision * recall) / (precision + recall)，衡量 precision 和 recall 是否平衡

Average Precision：IoU threshold 為不同值時，Precision 的平均

Mean Average Precision：每個 class 的 AP 的平均值

由於 testing data 無 ground truth boundary，因此框到人臉且類別正確就視為 TP

若框中但類別標錯，則算入 FP

若一張人臉被框中兩次，則第二次算入 FP

若連框都沒框，則不計入

實驗一：

模型：pretrained darknet53 on ImageNet dataset

https://github.com/VictorLin000/YOLOv3_mask_detect

訓練集：90% of VictorLin000(610 images)

驗證集：10% of VictorLin000(68 images)

測試集：額外於網路上抓取每類各 25 張

實驗二：

模型：pretrained darknet53 on ImageNet dataset

訓練集：使用 Mask-Face-Net 和 FFHQ 的 data，每類各取 1000 張，boundary 為全圖 (2700 images)

驗證集：10% of dataset described above(300 images)

測試集：同上

實驗一集實驗三的 train/valid 資料集(多人多 class/張)和我們抓的 test 資料集(單人單 class/張)差異大，但實驗二中差異較小，皆是單人單 class/張

Analysis：

1. 使用 VictorLin000 資料集比 MFN+FFHQ 還要好
2. 實驗二完全的 overfit 在 train/valid dataset 上，使得在 test dataset 上表現極差，且由於生成 boundary 時設定為全圖，模型傾向將全圖框為 boundary

3. 可以發現實驗一和 two-stage 的 model 相比，於 test dataset 上的表現極佳，且使用了更少的資料量，反而更有效率。可以看到除了是 dataset 本身的差異外，還有 darknet53 架構上的不同，採用 Residual 和 Feature Pyramid Network，使得 model 能根據不同 scale 來進行 detection。

短評：