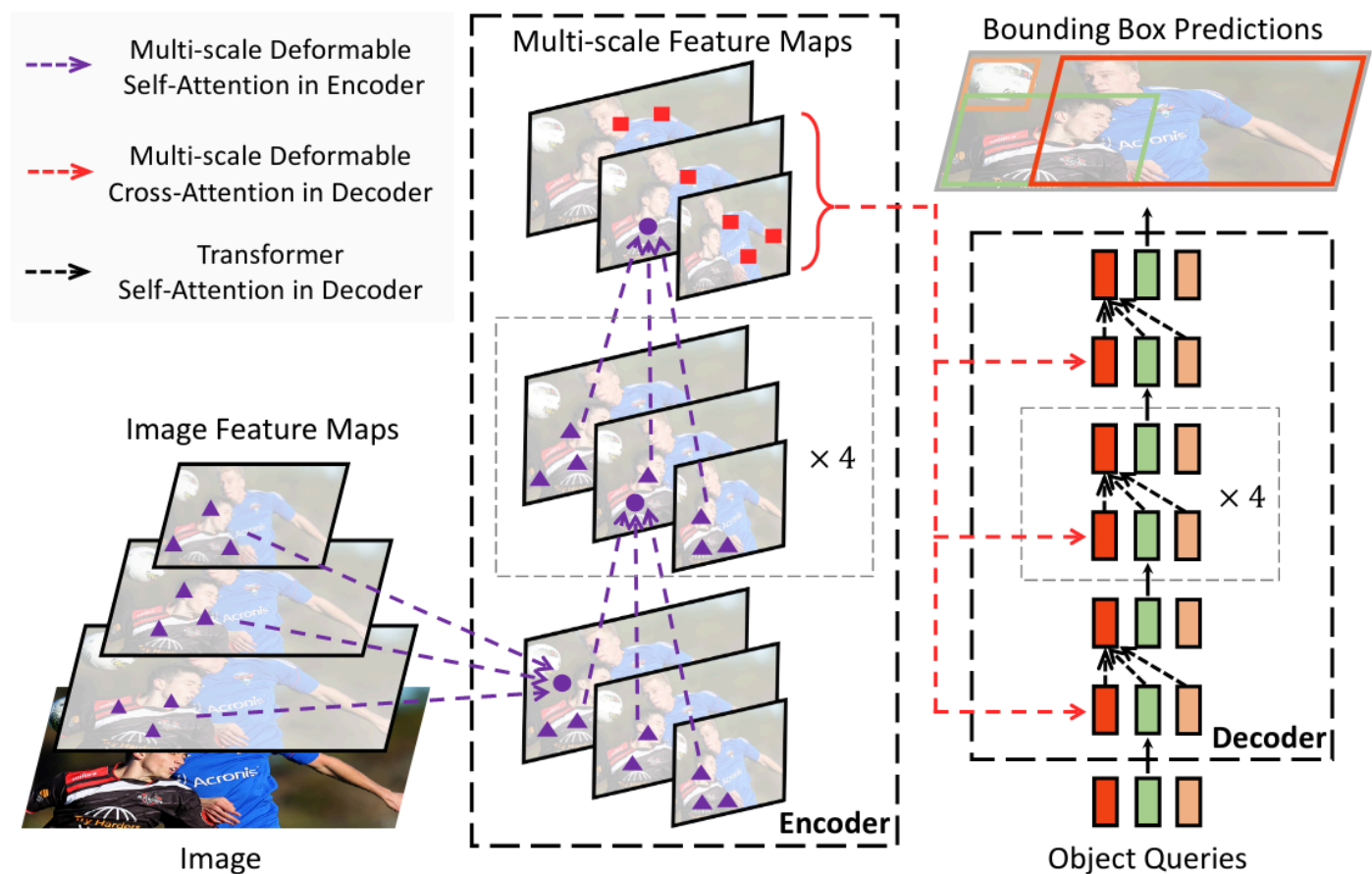


CVPDL HW1

資工碩二 R12922016 葉丞勛

Q1 : Architecture of my model

在本次作業，我使用的 model 是 two-stage 的 **Deformable DETR**，以 ResNet-50 作為 backbone。此模型透過 ResNet-50 來提取圖片中的特徵，並使用兩階段的方式，先在第一階段生成 object 的候選位置 (proposals)，然後第二階段再對這些 proposals 進行分類和 bounding box 的預測。



- reference : https://github.com/open-mmlab/mmdetection/tree/main/configs/deformable_detr (https://github.com/open-mmlab/mmdetection/tree/main/configs/deformable_detr).

Q2 : Implement details

Augmentation

- RandomFlip : 有 50% 的機率圖片會以隨機的方式(水平或垂直)進行翻轉。
- RandomChoiceResize : 在給定的多個不同的尺寸中隨機選擇一個進行縮放。
- RandomCrop : 隨機對圖片進行裁剪。

Loss function

- loss_bbox : 是一種 L1Loss，用於計算 predicted bboxes 跟 ground truth bboxes 之間的位置差異。
- loss_cls : 是一種 FocalLoss，用於處理 training 時 class 不平衡的情況，通過減少容易分辨的 class 的 weights，專注在比較難分辨的 class 上。
- loss_iou : 是一種 GloULoss，用於計算 predicted bboxes 跟 ground truth bboxes 之間重疊程度，並基於 Generalized IoU 的指標進行優化，提升 predicted bboxes 的準確率。

Parameter settings

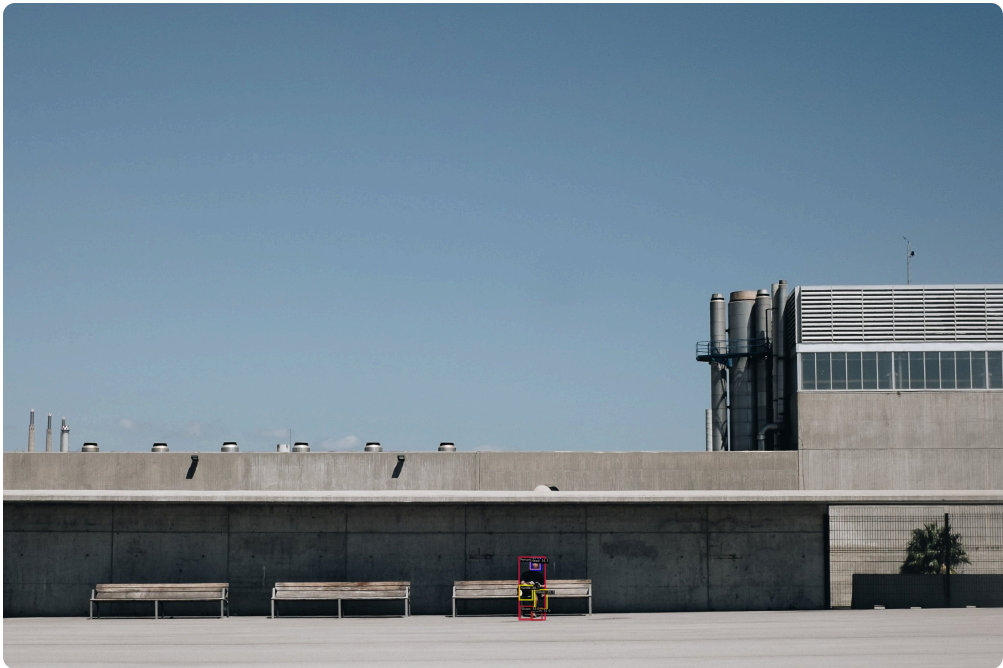
- epoch : 50
- optimizer : AdamW
- learning rate : 0.0002
- weight_decay : 0.0001
- batch_size : 4

Q3 : Performance for validation set

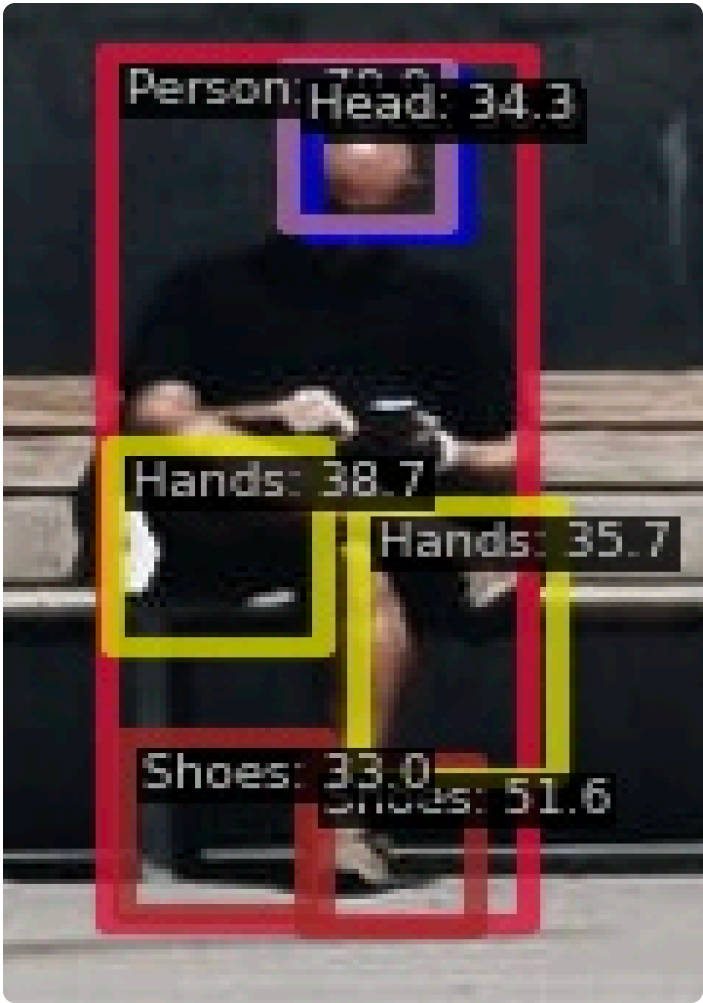
- mAP(50-95): 0.6405
- AP_50: 0.8605
- AP_75: 0.7218

Q4 : Visualization and discussion

result :



↓ (zoom in)



long tail effect

Long tail effects是指在 training data 中，某些類別的出現頻率較低(tail classes)，另一些類別出現頻率比較高 (head classes)，導致 model 難以對 tail classes 進行準確的學習與預測。而目前主要包括以下幾種方法，來解決此問題：

1. Resampling：透過增加 tail classes samples，或是減少 head classes samples，來平衡不同 class 的分佈。
2. Loss Function Adjustment：透過設計不同的 loss function，讓 model 對於 tail classes 更加敏感，進而提升 model 辨識 tail classes 的準確率。

以下為這次作業 training data 的分佈，可以發現 id：2(Earmuffs), 4(Face-guard), 13(Medical-suit), 15(Safety-suit) 異常偏低，導致他們的 AP 很低，model 很難精準的預測出他們。

