**Training setting:**

The YOLOv5 is based on PyTorch 1.7.0 framework. The experiment was performed using google colab consist (equipped with) of Intel(R) Xeon(R) @ 2.30 GHz processor, 13 GB RAM, NVIDIA Tesla P40, 21.01 GB disk space.

As the dataset is small, transfer learning is used to train the model. The model initialize the initial weights from the pretrained model which was trained on coco dataset over 80 classes. This is important to reduce overfitting of the model. Different data augmentation technique like HSV-Hue augmentation, HSV-Saturation augmentation, image rotation, translation, scale, shear, perspective, flip up-down, flip left-right, mosaic and mix-up are used to overcome less data over different classes.

It ran for 300 epochs and batch size was 16. The initial learning rate is 0.01, the momentum and weight decay are 0.937 and 0.0005 respectively. And the warmup momentum and warmup bias learning rate are 0.8 and 0.1 respectively.

**Result analysis:**

Model evaluation metrics:

The precision and recall are the most crucial topic for object detection model. Precision is the measurement predicted positive class which is actually positive class. The equation for precision is:

p =

And recall means the number of predicted positive class out of all Positive objects in the dataset. The equation for recall is:

R =

Here True Positive (TP) indicates the number of object correctly predicted to that specific class. True Negative (TN) indicates the number of non-object object correctly. False Positive (FP) indicates the number of object class detected but that doesn’t belongs to that specific class. False Negative (FN) indicates the number of object is not detected.

mAP and F1 score:

The mean average precision (mAP) how correctly bounding box are predicted over the object. It is calculated by taking mean of Average Precision (AP) of the all classes and their overall IOU (Intersection Over Union) threshold. Average Precision (AP) can be represented as:

AP =

The F1 score is used to conclude overall model performance of the model considering Precision and recall value. The equation for F1 score can be represented as:

F1 =

Frame Per Second and inference time:

FPS or Frame per second