…after explaining YOLO…

Training

Datasets

4 models of YOLOv5 are publicly available, each having its own pre-trained weights on the COCO dataset. The said dataset is not inclusive of objects/animals found underwater, which necessitates the need for transfer learning. Images of [INSERT CLASSES] were taken from publicly available datasets from sources like [INSERT SOURCES]. Training of YOLO models require the annotations of each image, that is the coordinates of the rectangle that encompasses the required object in the said image. While some datasets came along with their annotations, others required manual annotation via software like labelImg that allows the user to manually select the coordinates of the object.

[INSERT DETAILS ABOUT YOLO MODELS IF REQUIRED]

[INSERT PHOTOS IF REQUIRED]

Object Detection Metrics

Mean Average Precision or mAP is a very commonly used evaluation metric in computer vision for object detection models. An understanding of multiple other metrics is vital before we derive the mAP of a model.

{PULUTHIFY BELOW PARAGRAPH}

In general Data Science, Precision and Recall are two metrics used to evaluate the accuracy of the predictions in classification problems. Precision is commensurate [epuudii] to the percentage of predictions that are correctly classified, while Recall corresponds to how well the Positives are classified.

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

TP - True Positives (Positives that were classified as positive) (Object detected)

FP - False Positives (Negatives that were classified as positive) (Object absent but detected)

FN - False Negatives (Positives that were classified as negative) (Object present but not detected)

{PULUTHIFY BELOW PARAGRAPH}

As discussed before, YOLO predictions are made in the form of a bounding box, around the object that we aim to detect. While evaluating, the overlap between the predicted box and the ground truth is measured. This value is knows as IoU (intersection-over-union) and it denotes how much our prediction coincides with the truth box. Basically higher the IoU, higher the accuracy.

Since there is no default Yes/No values for classification with respect to IoU, a threshold ‘T’ is defined. If the IoU of our prediction is greater than T, it is classified as True Positive. If not, it is classified as False Positive.

The generic method to calculate the value of Average Precision (AP) is to estimate the area under the Precision-Recall curve. mAP can be determined as the average of AP.

Talking in particular to object detection, the mAP score is calculated by computing the mean AP over all IoU thresholds, depending upon the specific parameters of the model.

Training

Since the availability of a CUDA compatible GPU is highly beneficial, a Google Colab environment was used for training. A total of [N, original n=764] images [per class] were trained on the model for [N, original=50 epochs] epochs. Figure [N] shows the training and validation loss of the model, which also threw a mAP accuracy score of [N, original=0.95].