

YoloV5 基础篇

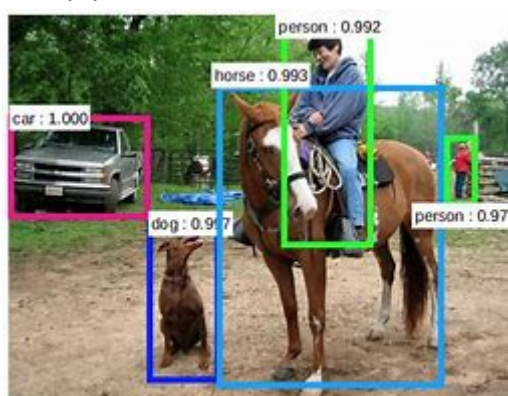
目标检测的任务说明

目标检测 (Object Detection) = What and Where

目标检测任务是计算机视觉中比较基础的一类任务，在上述给出的简单公式中，What即**识别**，也就是识别一个物体。

在识别到物体之后，在图像上标注出物体的位置，也就**定位**，在图像上标注出物体的位置，通常还会评估物体属于某个标签，以及属于这个标签的置信度。

如下图：



这个经过目标检测标注的图片就是一张典型的输出，其中包含What（识别）、Where（定位）、Label（标签）、Score（置信度得分）。

例如其中的dog是一个标签，通过物体检测识别到其中有dog，并且定位到dog画出其最小外接矩形，然后评估置信度得分。

除了目标检测之外，计算机识别还有其他几个基础任务：

1. 分类：将多张图片分为不同类别
2. 分类+定位：将图像分类后，在图像中标注出分类依据，也就是给出主体定位
3. 实体分割：在图像中检测到不同的实体，并绘制出轮廓将其分割出来。

总结一下：

定位：寻找图像中给定标签的单个目标。

检测：寻找图像中给定标签的所有目标。

目标检测常用数据集

目标检测最常用的数据集：

- PASCAL VOC
- MS COCO

PASCAL VOC数据集来自PASCAL VOC挑战赛，其在2005至2012年间展开，在PASCAL VOC 2007数据集中，包含9963张图像、24640个标注，在PASCAL VOC2012数据集中，包含11530张图像、27450个标注。

该数据集有20个分类：

- Person: person
- Animal: bird, cat, cow, dog, horse, sheep
- Vehicle: aeroplane, bicycle, boat, bus, car, motorbike, train
- Indoor: bottle, chair, dining table, potted plant, sofa, tv/monitor

数据集地址：

[The PASCAL Visual Object Classes Challenge 2012 \(VOC2012\) \(ox.ac.uk\)](http://www.robots.ox.ac.uk/visipedia/PASCAL_VOC/)

另一个MS COCO数据集，MS COCO的全称是Microsoft Common Objects in Context，起源于是微软于2014年出资标注的Microsoft COCO数据集，与ImageNet竞赛一样，被视为是计算机视觉领域最受关注和最权威的比赛之一。

在ImageNet竞赛停办后，COCO竞赛就成为是当前目标识别、检测等领域的一个最权威、最重要的标杆，也是目前该领域在国际上唯一能汇集Google、微软、Facebook以及国内外众多顶尖院校和优秀创新企业共同参与的大赛。

数据集地址：<http://cocodataset.org/>

目标检测性能指标与计算方法

性能指标计算

检测性能指标计算包括两方面；**检测精度和检测速度**。

在检测精度中常用的指标计算方法：

1. **Precision**: 精确率，指所有预测为正类的样本中，实际为正类的比例，用于评价预测精准度。
2. **Recall**: 召回率，指实际正类中，预测为正类的比例，用于评价预测完整度。
3. **F1 Score**: 精确率和召回率的调和平均数，提供一个综合考虑两者的指标。
4. **IoU (Intersection over Union, 交并比)**: 预测框与真实框的交集和并集面积之比，用于评价目标检测定位精度。IoU 越高表示模型对目标位置的预测越准确。
5. **P-R Curve(Precision-Recall Curve, 精确率-召回率曲线)**: 通过不同阈值计算出一系列精确率、召回率拟合的曲线。用于分析模型在不同阈值下的表现。
6. **AP (Average Precision 平均精度)**: P-R 曲线下的面积，即在各个阈值下精确率的平均值。P 常用来衡量目标检测任务中一个类别的检测精度，是 P-R 曲线的一个量化指标。

7. **mAP (mean average precision, 平均平均精度)**：对于多类别检测任务，将每个类别的 AP 取平均得到的值。mAP 是衡量多类别目标检测模型整体表现的重要指标，越高表示模型整体检测精度越高。

混淆矩阵 (confusion matrix)

对于分类问题，最常用的评价指标集计算方法是混淆矩阵法。

		预测 Prediction	
		Positive	Negative
实际 Actual	True	TP	FN
	False	FP	TN

第一位T/F: 表示预测的对错

第二位P/N: 表示预测的结果

如上图所示，混淆矩阵的行表示样本的实际值，列表示样本的预测值，交叉得出实际值与预测值的组合，例如TP即为实际为True且预测为Positive表示正确的预测，而FN为实际为True，预测为Negative表示错误的预测。

FP表示错误的预测为了Positive，TN表示正确的预测为Negative。

根据混淆矩阵可以计算出Precision、Recall、Accuracy、F1-score。

他们的计算公式为：

Precision (精确率) :

$$\frac{TP}{TP + FP}$$

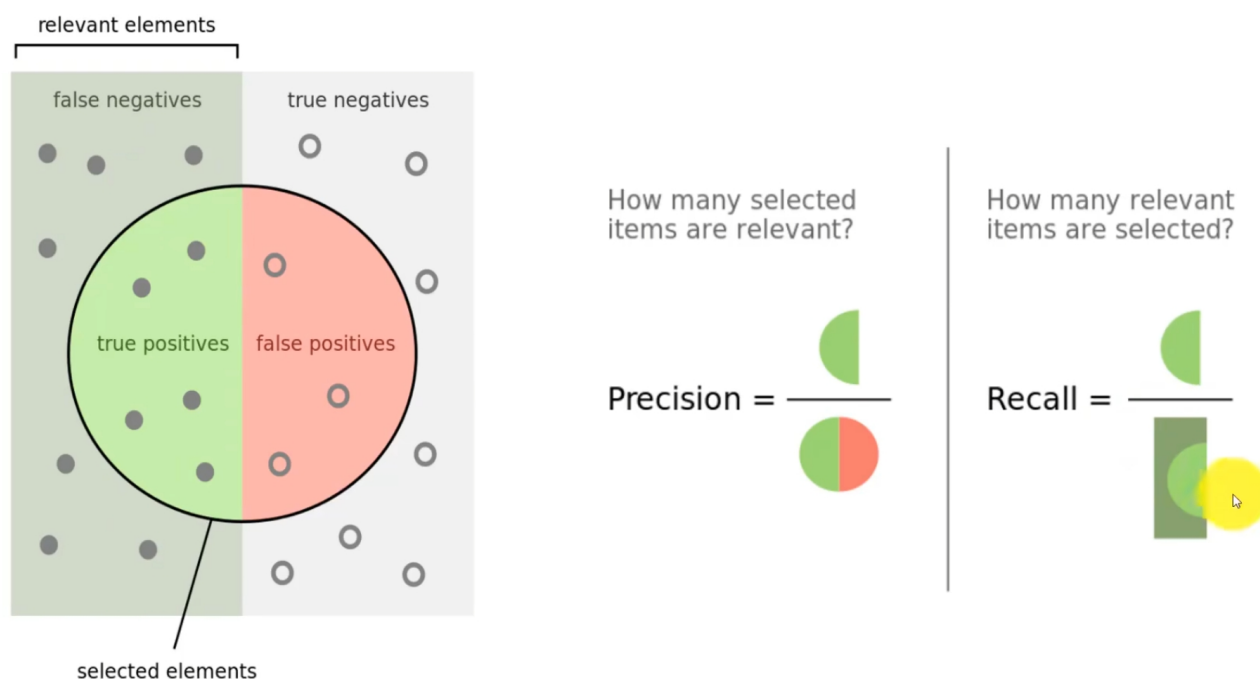
表示所有预测为正确的值中，实际正确值的比例，用于衡量模型的预测精准度。

Recall (召回率) :

$$\frac{TP}{TP + FN}$$

表示正确预测为True的样本在所有True的样本的比例，用于评估预测的完整度，Recall越大，表示模型的预测完整度越高。

下图可以直观的展示Precision与Recall的计算过程：



Accuracy (准确率) :

$$\frac{TP + TN}{TP + TN + FP + FN}$$

与Precision不同的是，准确率衡量的是所有的预测正确的样本占总体样本的比例，其加入了实际样本中为False的样本。而Precision仅评价标签为True的样本。

F1 Score:

$$\frac{2 * Precision * Recall}{Precision + Recall}$$

F1 Score是精确率和召回率的调和平均数，用于评价模型在精确率和召回率的均衡表现。

Extend: 关于调和平均数

在数理中，调和平均数指一组数字的倒数的算数平均数的倒数。

调和平均数的公式：

$$\frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

与算数平均数相比，调和平均数更加注重较小的值，由于他对数值取倒数进行计算，因此针对较小的值对结果的影响会更加显著。

假设有四个数值：5,3,9,2

他们的算数平均数为 $5+3+9+2=19/4=4.75$

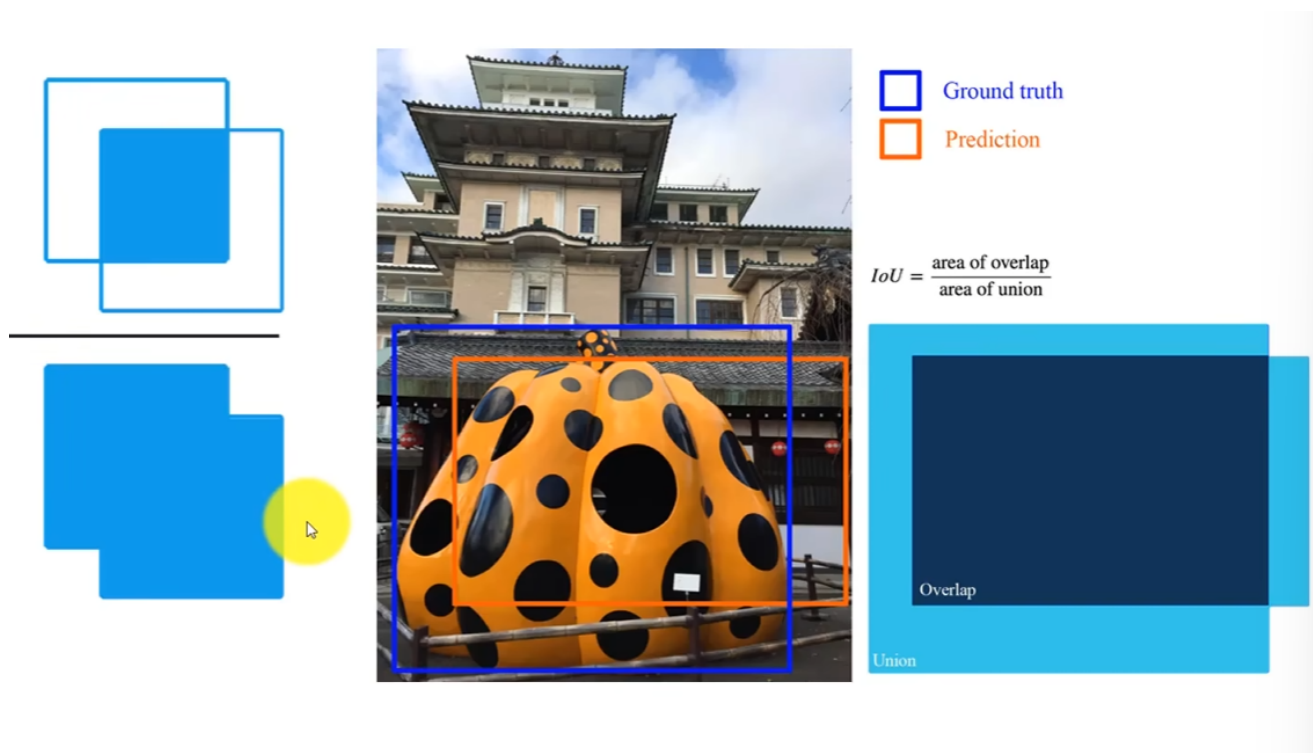
调和平均数为：

$$\frac{4}{\frac{1}{5} + \frac{1}{3} + \frac{1}{9} + \frac{1}{2}} = 0.57$$

IOU (intersection over union 交并比)

交并比即在图像目标检测任务中，识别到的区域 (prediction) 与实际的物体区域 (Ground truth) 两者重叠部分与组合部分的比。

交并比越接近1，说明模型预测的精度越高。反之说明预测精度越低。



如上图所示。

当IOU等于1时 说明预测的区域和实际的区域完全重合，这时候预测的效果最好。

实际的模型训练中，可以设置一个IOU阈值来对预测进行分类。

例如：

当IOU>0.5时，将目标检测的结果分类为True Positive (TP)

当IOU<0.5时，将目标检测的结果分类为False Positive (FP)

当图像中含有Ground Truth，但是模型并没有预测到任何区域时，将结果分类为False Negative (FN)

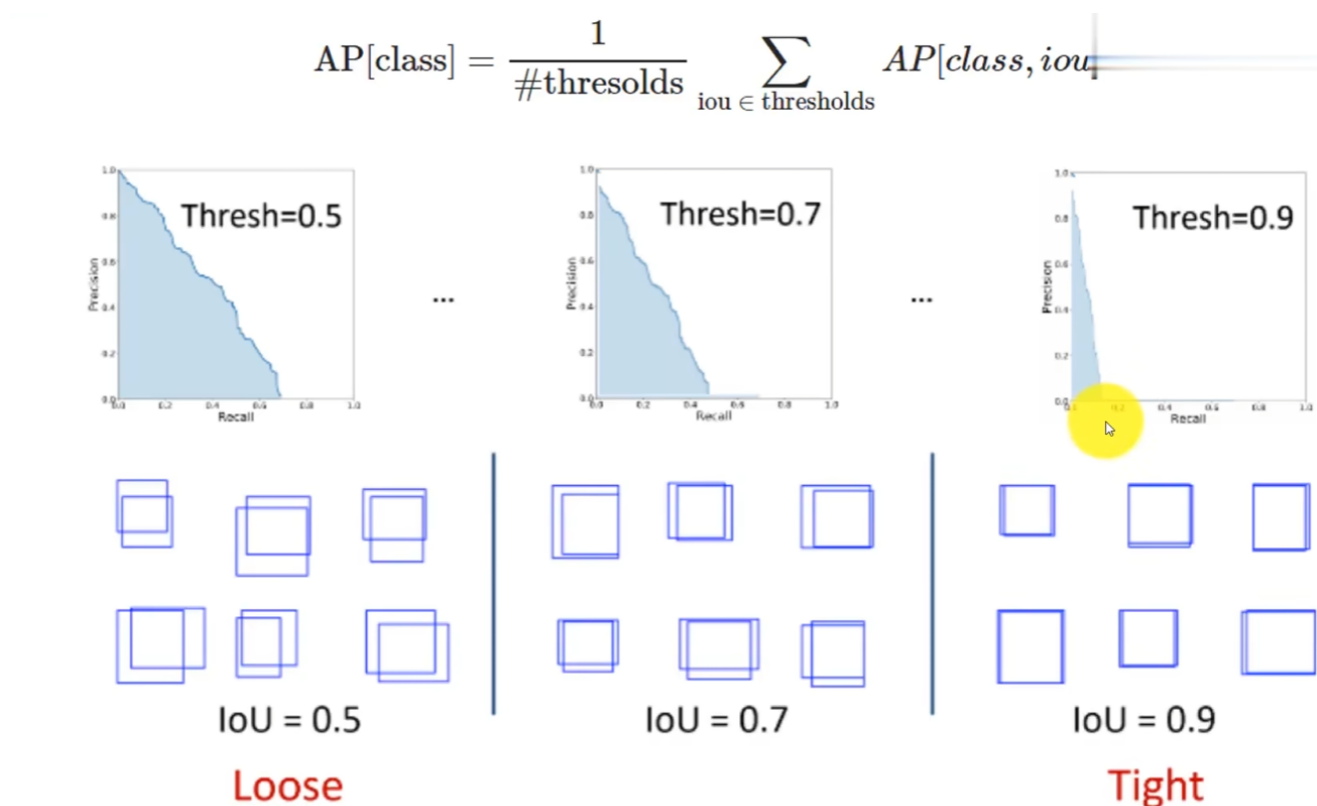
当图像中没有Ground Truth，并且模型没有预测到任何区域，将结果分类为True Negative

同时有一个AP指标，即average Precision 平均精确率。用于评价模型在在一个类别上的好坏。

而mAP即mean of average precision over classes 取每个类别的平均精确率累加再计算平均值，用于评价模型在所有类别上的好坏。

针对不同的IOU阈值，在模型评价上的表现即为不同的评价要求。

一般来说，当IOU阈值设定越高，则评价越严格，此时对应的FP值则越大，通过混淆矩阵计算的召回率也就越小，在以下图示中可以直观的体现，其中图表部分的数据即为不同的IOU阈值对模型评价的Recall指标的影响：



不同数据集的AP与mAP的定义

对PASCAL VOC Challenge 数据集 我们通常认为当IOU>0.5为正样本，但是如果检测对同一目标的多个检测，则认为第一个检测为正样本，其他的均为负样本。

对于MS COCO数据集则对AP有自己的定义。

[AP@.50](#)表示IOU=0.5的结果的均值

[AP@.75](#)表示IOU=0.75的结果的均值

也就是，相当于PASCAL VOC Challenge 数据集中的mAP。

MS COCO数据集对AP的定义是多个IOU值相同的结果的精确率的均值。

对于PASCAL VOC数据集，AP@[.5:.95]表示 IOU为0.5到0.95间隔0.05步长的平均平均精确率

对于MS COCO数据集，其表示从0.5到0.95的0.05步长的十个IOU等级，在80个类别的平均平均精确率的平均值。

计算公式为：

$$mAP_{coco} = \frac{mAP_{0.50} + mAP_{0.55} + \dots + mAP_{0.95}}{10}$$

AP的计算

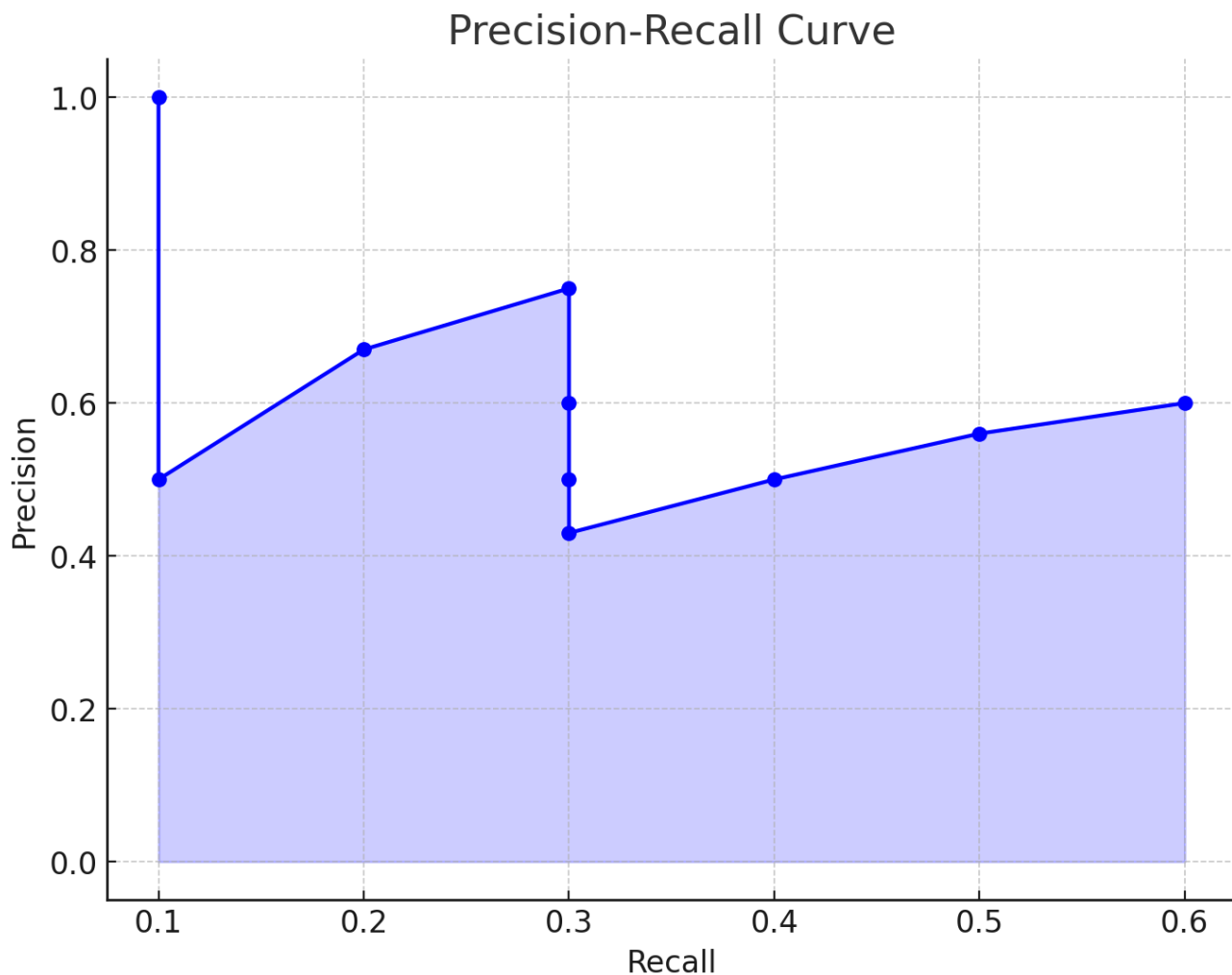
1. 梯形面积法

首先举一个例子，我们有十张含有小狗的图像作为目标检测任务的测试集，然后通过混淆矩阵对模型进行评价，通过模型对十张图片分别处理，结果如下表格所示：

序号	是否正确	Precision	Recall
1	True	1.0	0.10
2	False	0.5	0.10
3	True	0.67	0.2
4	True	0.75	0.3
5	False	0.6	0.3
6	False	0.5	0.3
7	False	0.43	0.3
8	True	0.5	0.4
9	True	0.56	0.5
10	True	0.6	0.6

由上表趋势可以看出，整个预测过程中，Precision的值是上下波动的，而Recall的值是随着预测的次数不断增加的。

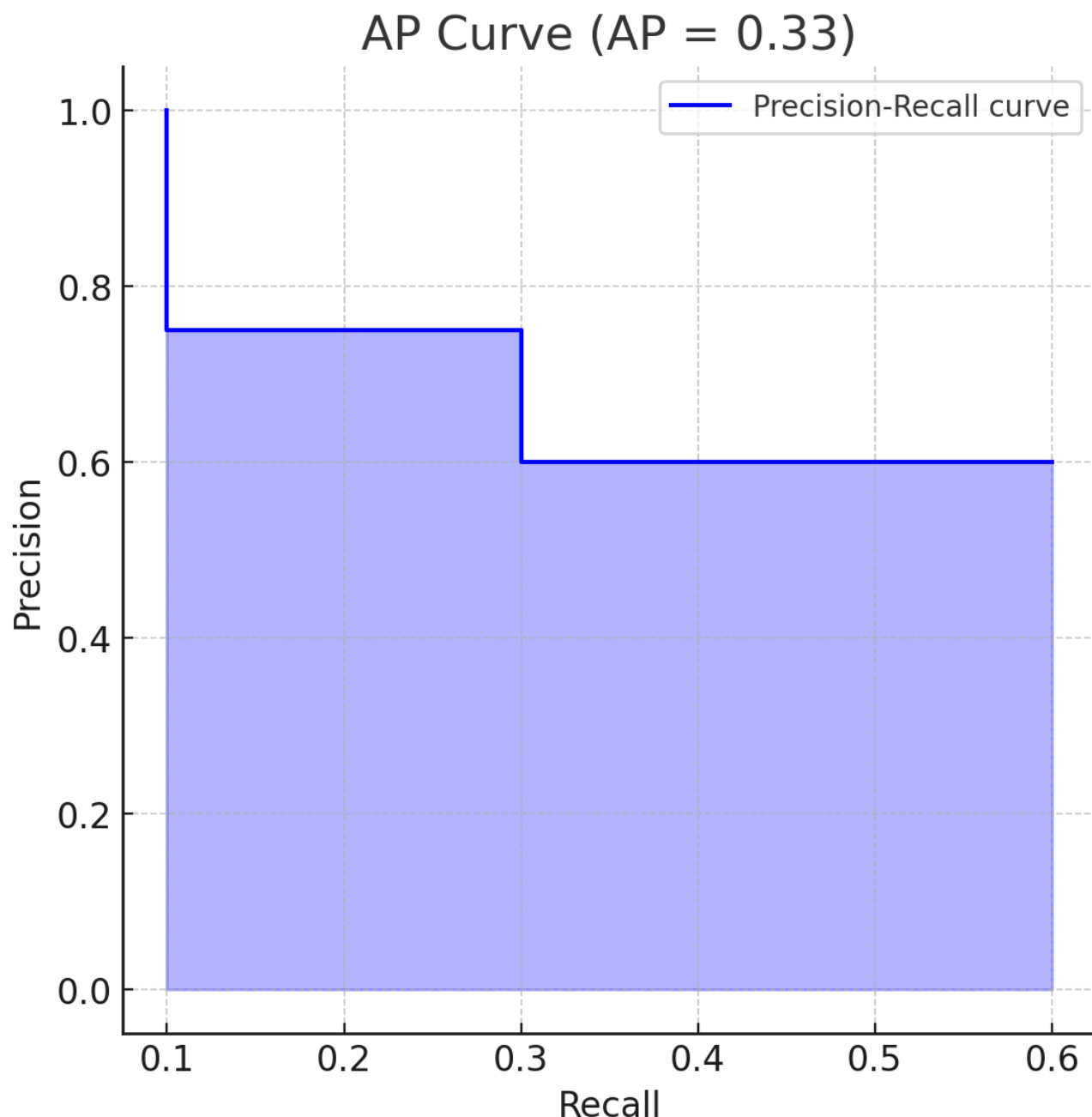
将上边绘制为图：



从图中可以直观的看出，随着Recall的不断增加，Precision呈现上下波动的趋势。

AP在概念上可以被视为Precision-Recall curve（精确率-召回率曲线）下方的曲线。

通常我们可以将PR曲线在顶点处的Precision值拉直，然后**计算每个Recall段的面积**来计时计算AP值。



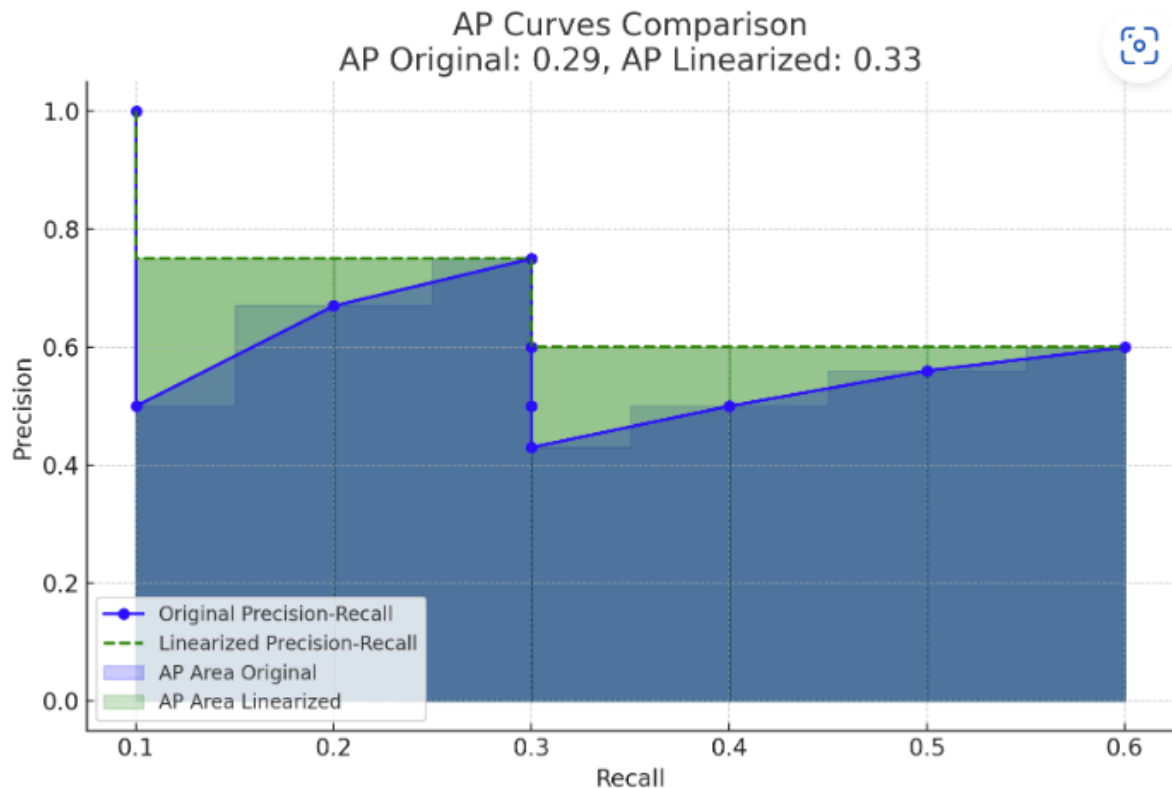
如上图所示，将Precision值在顶点处（0.67、0.6）处拉直，然后根据公式：

$$AP = \sum_{i=1}^{n-1} P(R_i) \times (R_{i+1} - R_i)$$

计算每一个P-R点的组合的面积即近似为AP值。

2. 积分法

积分法是在梯形面积法之后提出的AP计算方式，他的计算方式也很简单，通过划分几个大的积分区域，计算这几个区域的面积的和即可得出近似的AP值。



如上图绿色部分所示。

计算 $0.2 \times 0.75 + 0.3 \times 0.6 = 0.33$ 即为大致的A值 通过原始的数据进行计算，得出的原始值是0.29，与积分法计算的AP差距不大。

检测速度的指标

检测速度的指标包括以下几个部分：

1. **前传耗时**：以毫秒作为单位，表示从输入一张图像到输出最终结果所消耗的时间，同时包含前处理耗时（例如图像归一化）、网络前传耗时、后处理耗时（例如非极大值抑制等）
2. **每秒帧数FPS**：每秒能够处理的图像数量，与硬件配置相关
3. **浮点运算量 (FLOPS)**：指处理一张图像所需要的浮点运算量，与具体的软硬件环境没有关系，可以公平的比较不同的算法之间的检测速度。