

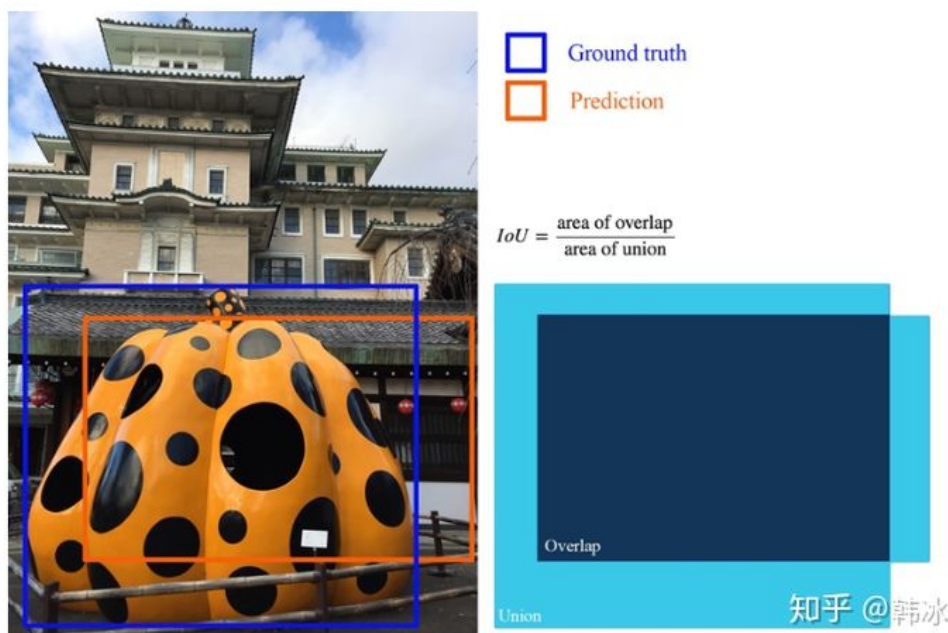
韩昊

[博客园](#) | [首页](#) | [新随笔](#) | [联系](#) | [管理](#) | [订阅](#) | [RSS](#)

随笔- 42 文章- 0 评论- 33 阅读- 64788

目标检测中的precision, recall, AP, mAP计算详解

交并比IoU衡量的是两个区域的重叠程度，是两个区域重叠部分面积占二者总面积（重叠部分只计算一次）的比例。如下图，两个矩形框的IoU是交叉面积（中间图片红色部分）与合并面积（右图红色部分）面积之比。



IoU的定义

在目标检测任务中，如果我们模型输出的矩形框与我们人工标注的矩形框的IoU值大于某个阈值时（通常为0.5）即认为我们的模型输出了正确的

精准率与召回率 (Precision & Recall)

大雁与飞机

假设现在有一个测试集，测试集中的图片只由大雁和飞机两种图片组成，如下图所示：



假设你的分类系统最终的目的是：能取出测试集中所有飞机的图片，而不是大雁的图片。

现在做如下的定义：

True positives：飞机的图片被正确的识别成了飞机。

 昵称：韩昊
 园龄：2年6个月
 粉丝：4
 关注：3
[+加关注](#)

2022年3月						
日	一	二	三	四	五	六
27	28	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2
3	4	5	6	7	8	9

搜索

常用链接

[我的随笔](#)
[我的评论](#)
[我的参与](#)
[最新评论](#)
[我的标签](#)
[更多链接](#)

随笔档案

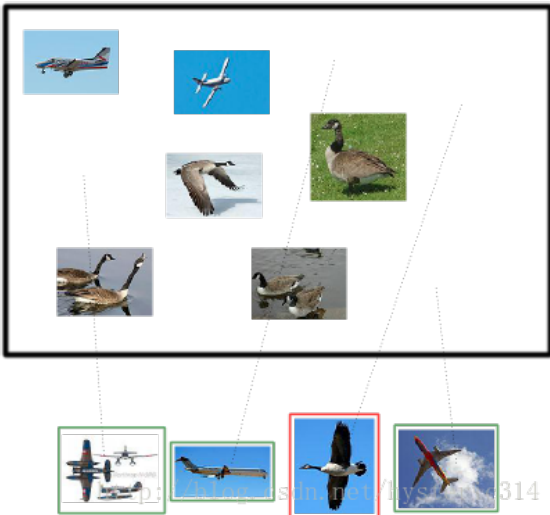
[2021年1月\(2\)](#)
[2020年10月\(3\)](#)
[2020年8月\(1\)](#)
[2020年7月\(1\)](#)
[2020年6月\(3\)](#)
[2020年4月\(5\)](#)
[2020年3月\(2\)](#)
[2020年2月\(3\)](#)
[2019年12月\(1\)](#)
[2019年11月\(3\)](#)
[2019年10月\(4\)](#)
[2019年9月\(14\)](#)

阅读排行榜

1. pycharm中导入模块错误时提示Try to run this command from the system terminal. Make sure that you use the...(18143)

True negatives: 大雁的图片没有被识别出来，系统正确地认为它们是大雁。
False positives: 大雁的图片被错误地识别成了飞机。
False negatives: 飞机的图片没有被识别出来，系统错误地认为它们是大雁。

假设你的分类系统使用了上述假设识别出了四个结果，如下图所示：



那么在识别出的这四张照片中：
True positives: 有三个，画绿色框的飞机。
False positives: 有一个，画红色框的大雁。

没被识别出来的六张图片中：
True negatives: 有四个，这四个大雁的图片，系统正确地没有把它们识别成飞机。
False negatives: 有两个，两个飞机没有被识别出来，系统错误地认为它们是大雁。

Precision 与 Recall

Precision其实就是在识别出来的图片中，True positives所占的比率：

$$\text{precision} = \frac{tp}{tp + fp} = \frac{tp}{n}$$

其中的n代表的是(True positives + False positives)，也就是系统一共识别出来多少照片。
在这一例子中，True positives为3，False positives为1，所以Precision值是 3/ (3+1) =0.75。
意味着在识别出的结果中，飞机的图片占75%。

Recall 是被正确识别出来的飞机个数与测试集中所有飞机的个数的比值：

$$\text{recall} = \frac{tp}{tp + fn}$$

通常tp+fn+fn在目标检测中指groundTruth中的真实目标数量，

Recall的分母是(True positives + False negatives)，这两个值的和，可以理解为一共有多少张飞机的照片。
在这一例子中，True positives为3，False negatives为2，那么Recall值是 3/ (3+2) =0.6。
意味着在所有的飞机图片中，60%的飞机被正确的识别成飞机。

Precision 和 Recall最早是信息检索中的概念，用来评价一个信息检索系统的优劣。Precision 就是检索出来的条目中（比如：文档、网页等）有多大比例是我们需要的，Recall就是所有我们需要的网页的条目有多大比例被检索出来了。用到目标检测领域，假设我们有一组图片，里面有若干待检测的目标，Precision就代表我们模型检测出来的目标有多打比例是真正的目标物体，Recall就代表所有真实的目标有多大比例被我们的模型检测出来了。

我们如何来计算Precision和Recall的值呢。

	检测出的矩形框大于置信度阈值的 (Positive)	检测出的矩形框小于置信度阈值的 (Negative)
与数据集中某个目标框iou值大于0.5的 (True)	TP	TN
与数据集中所有目标框iou值均大于0.5的、重复检测的 (False)	FP	FN

目标检测TP, FP, TN, FN含义

2. 目标检测中的precision, recall, AP, mAP计算详解(7674)
3. 将py文件打包为exe文件方法(7298)
4. 目标检测中的AP计算(5199)
5. Windows系统下Pytorch与python版本不匹配导致模块包导入错误(4413)

评论排行榜

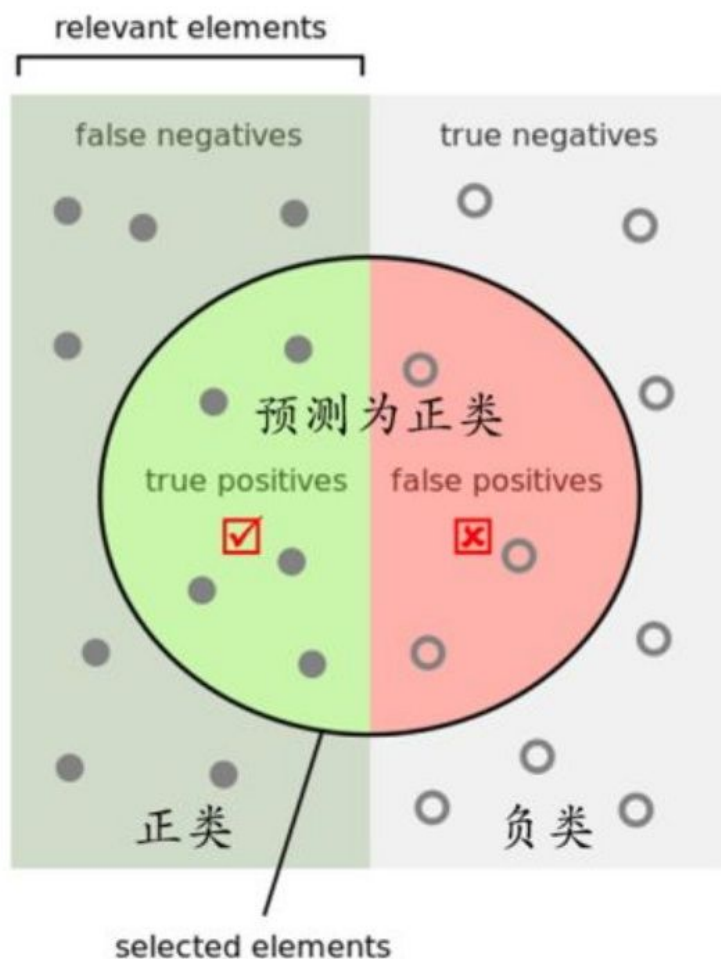
1. 将py文件打包为exe文件方法(6)
2. 韩昊 20190919-2 功能测试(4)
3. 目标检测中的precision, recall, AP, mAP计算详解(3)
4. 总结报告(3)
5. 关于四个定义的理解(3)

推荐排行榜

1. 目标检测中的precision, recall, AP, mAP计算详解(2)
2. 韩昊 20190905-2 博客作业(1)

最新评论

1. Re:pycharm python3.7成功安装d lib库
好人一生平安
--盈握之璧
2. Re:pycharm python3.7成功安装d lib库
感谢博主！！
--爱错了
3. Re:目标检测中的precision, recall, AP, mAP计算详解
请问TP,FP那个表格左下角那一栏，应该是“与数据集中所有目标框iou值均小于0.5”吧？
--HLXYH
4. Re:目标检测中的precision, recall, AP, mAP计算详解
@仕瑞_japina 为了提高计算准确度，防止过拟合...
--韩昊
5. Re:目标检测中的precision, recall, AP, mAP计算详解
谢谢博主好文章，有一处不明白的地方，麻烦您请教。“在实际应用中，我们并不直接对该PR曲线进行计算，而是对PR曲线进行平滑处理。”这里的原因是什么呀？为什么不直接对PR曲线进行计算？貌似可以计算得出来...
--仕瑞_japina



知乎 @韩冰

图

释TP, FP, TN, FN

首先我们根据上图的规则计算出TP, FP, TN的值, 则Precision, Recall可表示为

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

PR曲线

我们当然希望检测的结果P越高越好, R也越高越好, 但事实上这两者在某些情况下是矛盾的。比如极端情况下, 我们只检测出了一个结果, 且是准确的, 那么Precision就是100%, 但是Recall就很低; 而如果我们把所有结果都返回, 那么必然Recall必然很大, 但是Precision很低。

因此在不同的场合中需要自己判断希望P比较高还是R比较高。如果是做实验研究, 可以绘制Precision-Recall曲线来帮助分析。

这里我们举一个简单的例子, 假设我们的数据集中共有五个待检测的物体, 我们的模型给出了10个候选框, 我们按照模型给出的置信度由高到低对候选框进行排序。

Rank	Correct?	Precision	Recall
1	True	1.0	0.2
2	True	1.0	0.4
3	False	0.67	0.4
4	False	0.5	0.4
5	False	0.4	0.4
6	True	0.5	0.6
7	True	0.57	0.8
8	False	0.5	0.8
9	False	0.44	0.8
10	True	0.5	1.0

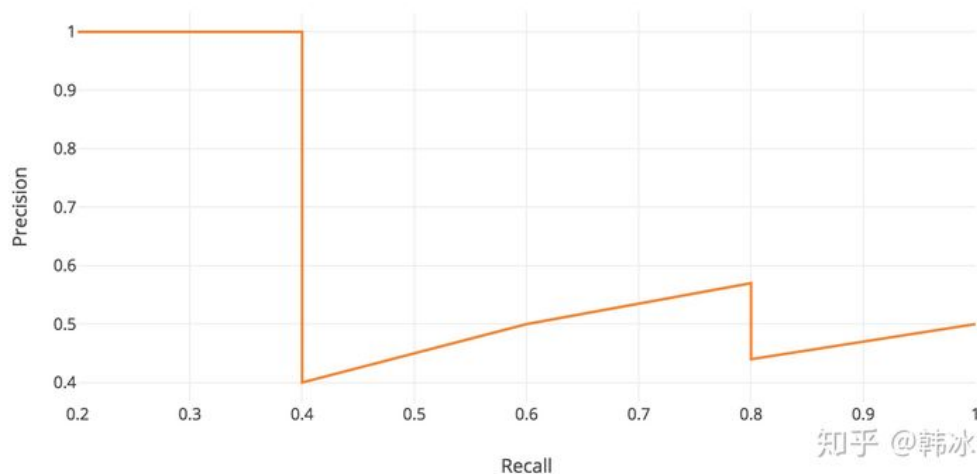
表格第二列表示该候选框是否预测正确（即是否存在某个待检测的物体与该候选框的iou值大于0.5）第三列和第四列表示以该行所在候选框置信度为阈值时，Precision和Recall的值。我们以表格的第三行为例进行计算：

$$TP = 2 \quad FP = 1 \quad TN = 3$$

$$Precision = \frac{2}{2+1} = 0.67$$

$$Recall = \frac{2}{2+3} = 0.4$$

由上表以Recall值为横轴，Precision值为纵轴，我们就可以得到PR曲线。我们会发现，Precision与Recall的值呈现负相关，在局部区域会上下波动。



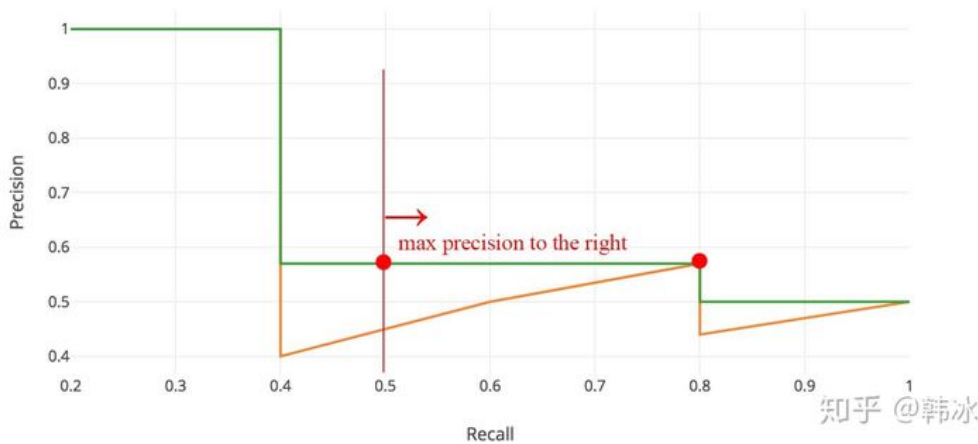
PR曲线

AP(Average Precision)

顾名思义AP就是平均精准度，简单来说就是对PR曲线上的Precision值求均值。对于pr曲线来说，我们使用积分来进行计算。

$$AP = \int_0^1 p(r) dr$$

在实际应用中，我们并不直接对该PR曲线进行计算，而是对PR曲线进行平滑处理。即对PR曲线上的每个点，Precision的值取该点右侧最大的Precision的值。



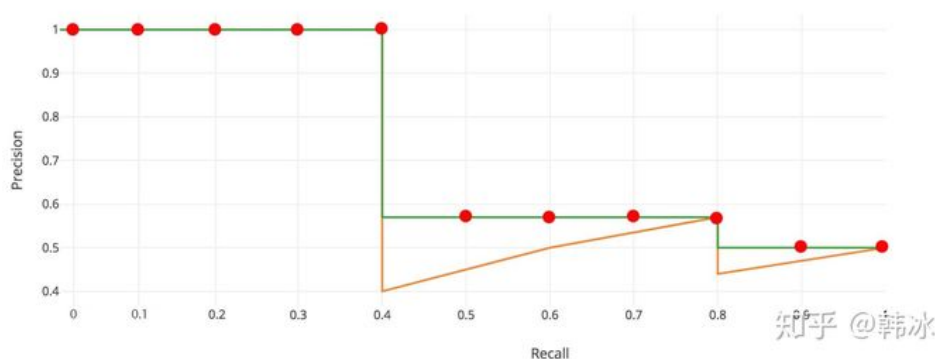
PR曲线的平滑处理

用公式来描述就是 $P_{smooth}(r) = \max_{r' \geq r} P(r')$ 。用该公式进行平滑后再用上述公式计算AP的值。

Interpolated AP (Pascal Voc 2008 的AP计算方式)

Pascal VOC 2008中设置IoU的阈值为0.5, 如果一个目标被重复检测, 则置信度最高的为正样本, 另一个为负样本。在平滑处理的PR曲线上, 取横轴0-1的10等分点 (包括断点共11个点) 的Precision的值, 计算其平均值为最终AP的值。

$$AP = \frac{1}{11} \sum_{0,0.1...1.0} P_{smooth}(i)$$



Pascal Voc 2008 AP计算方式

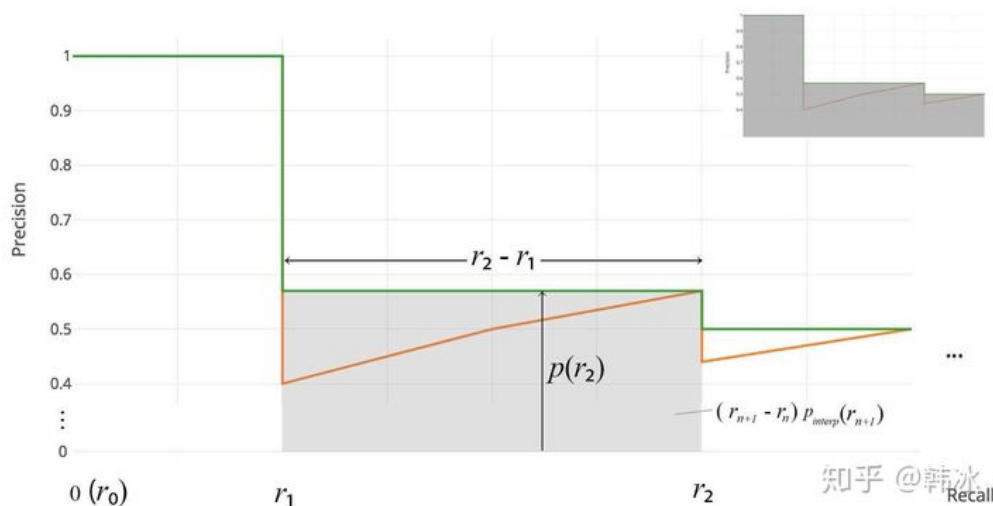
在我们的例子里

$$AP = \frac{1}{11} (5 \times 0.1 + 4 \times 0.57 + 2 \times 0.5) = 0.753$$

Area under curve

上述方法有两个缺陷, 第一个是使用11个采样点在精度方面会有损失。第二个是, 在比较两个AP值较小的模型时, 很难体现出两者的差别。所以这种方法在2009年的Pascalvoc之后便不再采用了。在Pascal voc 2010之后, 便开始采用这种精度更高的方式。绘制出平滑后的PR曲线后, 用积分的方式计算平滑曲线下方的面积作为最终的AP值。

$$AP = \int_0^1 p_{smooth}(r) dr$$



Pascal voc 2010-2012 AP 计算方式

COCO mAP

最新的目标检测相关论文都使用coco数据集来展示自己模型的效果。对于coco数据集来说，使用的也是Interpolated AP的计算方式。与Voc 2008不同的是，为了提高精度，在PR曲线上采样了100个点进行计算。而且iou的阈值从固定的0.5调整为在 0.5 - 0.95 的区间上每隔0.5计算一次AP的值，取所有结果的平均值作为最终的结果。

比如我们看一下YOLOv3的作者在论文中展示的在coco数据集上的实验结果

	backbone	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L
<i>Two-stage methods</i>							
Faster R-CNN+++ [3]	ResNet-101-C4	34.9	55.7	37.4	15.6	38.7	50.9
Faster R-CNN w FPN [6]	ResNet-101-FPN	36.2	59.1	39.0	18.2	39.0	48.2
Faster R-CNN by G-RMI [4]	Inception-ResNet-v2 [19]	34.7	55.5	36.7	13.5	38.1	52.0
Faster R-CNN w TDM [18]	Inception-ResNet-v2-TDM	36.8	57.7	39.2	16.2	39.8	52.1
<i>One-stage methods</i>							
YOLOv2 [13]	DarkNet-19 [13]	21.6	44.0	19.2	5.0	22.4	35.5
SSD513 [9, 2]	ResNet-101-SSD	31.2	50.4	33.3	10.2	34.5	49.8
DSSD513 [2]	ResNet-101-DSSD	33.2	53.3	35.2	13.0	35.4	51.1
RetinaNet [7]	ResNet-101-FPN	39.1	59.1	42.3	21.8	42.7	50.2
RetinaNet [7]	ResNeXt-101-FPN	40.8	61.1	44.1	24.1	44.2	51.2
YOLOv3 608 × 608	Darknet-53	33.0	57.9	34.4	18.3	35.4	41.9

COCO for YOLOv3

我们发现除了AP，还有 AP_{50} , AP_{75} 等值，这些事代表什么意思呢？

AP_{50} : IoU阈值为0.5时的AP测量值

AP_{75} : IoU阈值为0.75时的测量值

AP_S : 像素面积小于 32^2 的目标框的AP测量值

AP_M : 像素面积在 32^2 - 96^2 之间目标框的测量值

AP_L : 像素面积大于 96^2 的目标框的AP测量值

注：通常来说AP是在单个类别下的，mAP是AP值在所有类别下的均值。在这里，在coco的语境下AP便是mAP，这里的AP已经计算了所有类别下的平均值，这里的AP便是mAP。

好文要顶

关注我

收藏该文



韩昊

关注 - 3

粉丝 - 4

+加关注

« 上一篇: 转: 图像分类、物体检测、物体分割、实例分割、语义分割

» 下一篇: AlexNet、VGG、NIN、GoogLeNet、ResNet

posted @ 2020-03-05 20:28 韩昊 阅读(7674) 评论(3) 编辑 收藏 举报
刷新评论 刷新页面 返回顶部

登录后才能查看或发表评论，立即 登录 或者 逛逛 博客园首页

【推荐】百度智能云2022开年大促0.4折起，企业新用户享高配优惠

【推荐】开发者必读：华为 2022 年移动应用趋势洞察白皮书

编辑推荐：

- 记我第一次做线下技术分享的那些事
- 换个数据结构，一不小心节约了 591 台机器！
- [ASP.NET Core] MVC 模型绑定：非规范正文内容的处理
- 全面认识数据指标体系
- 项目开发中对成长的一些思考



最新新闻：

- 理想L9官方剧透：全自研旗舰级智能驾驶系统，4月16日正式发布
 - 晶圆代工厂争相“砸钱”扩产 中芯国际预计资本支出50亿美元
 - 美媒：波音会不会失去中国市场？
 - 2.7亿人每天听2.5小时，喜马拉雅竟还在亏钱
 - 在俄罗斯，Google为什么没有赢过Yandex？
- » 更多新闻...