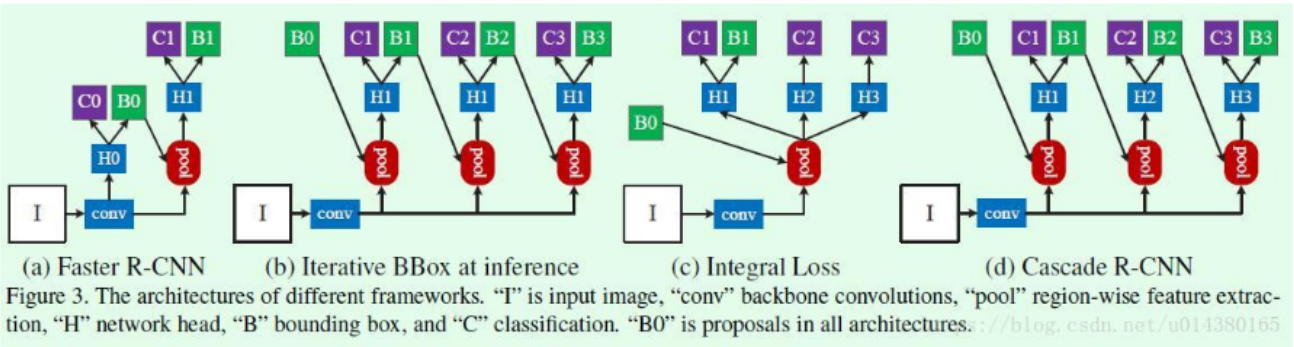


Cascade RCNN

- 问题：
 - (a) $u=0.5$ 也是常用的正负样本界定的阈值，但是当阈值取0.5时会有较多的误检，因为0.5的阈值会使得正样本中有较多的背景，这是较多误检的原因。
 - (b) 用0.7的IOU阈值可以减少误检，但检测效果不一定最好，主要原因在于IOU阈值越高，正样本的数量就越少，因此过拟合的风险就越大。
- 算法亮点：几个检测网络是基于不同IOU（交并比）阈值确定的正负样本上训练得到的。
- 解决问题
 - 较大的IOU阈值会导致阳性训练样本以指数方式减小，使得高 u 训练策略容易被过度适用。
 - 检测器与输入假设之间的推断质量不匹配。
- 算法思想：Cascade R-CNN是由一系列的检测模型组成，每个检测模型都基于不同IOU阈值的正负样本训练得到，前一个检测模型的输出作为后一个检测模型的输入，因此是stage by stage的训练方式，而且越往后的检测模型，其界定正负样本的IOU阈值是不断上升的。
 - 每一个阶段的检测器都专注于检测IOU在某一范围内的proposal，因为输出IOU普遍大于输入IOU，检测效果会越来越好。
 - **使用不同的IOU阈值，训练了多个级联的检测器。**
- 特点：使用一个阶段的输出来训练下一个阶段
- 扩展了两级架构Faster R-CNN。第一阶段是一个提议子网络，应用于整个图像，产生初步的检测假设。第二阶段，这些假设被一个感兴趣区域检测子网络处理，表示为检测头。最后一个分类评分和一个边界框被分配到每个假设。



边界框回归

- 一个边界框 $b = (b_x, b_y, b_w, b_h)$ 包含一个图像块 x 的四个坐标，边界框回归的任务是使用一个回归变量 $f(x, b)$ 将一个候选边界框 b 回归到一个目标边界框 g 中。从一个训练样本 (g_i, b_i) 中学习，以最小化边界框 L_1 损失函数。

检测质量

- 分类器 $h(x)$ 将一个图像块 x 赋值给 $M+1$ 个类中的一个，其中类 0 包含了背景和剩余的待检测目标。给定训练集，通过最小化分类交叉熵损失进行学习。由于边界框通常包含一个目标和一些背景，所以很难确定检测是正的还是负的。这通常用 IOU 标准来表示。如果 IOU 高于一个阈值 u ，那么这个块就是这个类的一个例子。

Cascade R-CNN

级联边界框回归

- 回归函数

$$f(x, b) = f^T \circ f^{(T-1)} \circ \dots \circ f^1(x, b)$$

T为级联阶段的总数。

- 1.级联回归是一个重新采样的过程，改变了不同阶段处理的假设的分布。
- 2.同时用于训练和推断，因此训练和推断分布没有差异。
- 3.多元回归器进行了优化，得到了不同阶段的重采样分布。

级联检测

- 在每个阶段，R-CNN包括一个分类器h和一个为IoU阈值优化的回归器f。

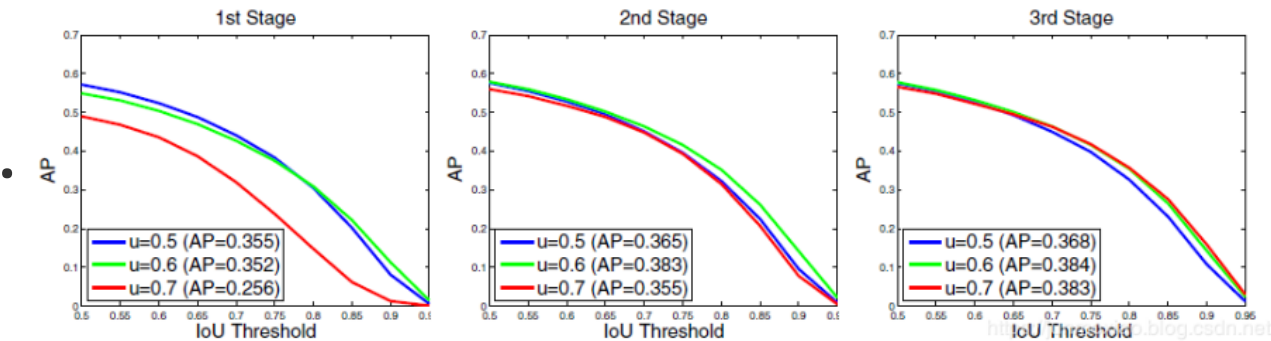
实现细节

- 四个阶段
 - 一个 RPN 和三个 $U=\{0.5, 0.6, 0.7\}$ 的检测阶段。
 - 第一检测阶段的采样。下面的阶段，只需使用前一阶段的回归输出即可实现重采样。

基线网络

- 为了测试Cascade R-CNN的通用性，使用三种常用的基线检测器：主干网为 VGG-Net 的 Faster R-CNN，R-FCN 和 ResNet 的 FPN。采样端到端训练。
- Faster R-CNN：网络头有两个全连接层。为了减少参数，修建不太重要的连接。每一个全连接层保留 2048 个单元，并删除 dropout 层。训练以 0.002 的学习率开始，在 60k 和 90k 迭代时降低 10 倍，在 100k 迭代时停止，在 2个同步GPU 上，每个GPU每次迭代持有 4 张图片。每个图像使用 128 个感兴趣区域。
- R-FCN：R-FCN向 ResNet 添加了卷积、边界框回归和分类层。Cascade R-CNN的所有头部都有这种结构。训练以 0.003 的学习率开始，在 160k 和 240k 迭代时学习率降低了 10 倍，在 280k 迭代时停止，在 4 个同步GPU上，每个GPU每次迭代持有一张 图像。每个图像使用 256 个感兴趣区域。
- FPN：使用 RoIAlign 做为更强的基线。使用 ResNet-50 进行模型简化测试，使用 ResNet-101 进行最终测试。在 8 个同步gpu上，120k 迭代的学习率为 0.005，60k 迭代的学习率 0.0005。每个图像使用 256 个感兴趣区域。
- 损失函数：和Faster R-CNN一致，分类采用 softmax，回归采用 smooth L1 loss，而且会对box的坐标进行归一化操作。

质量失配

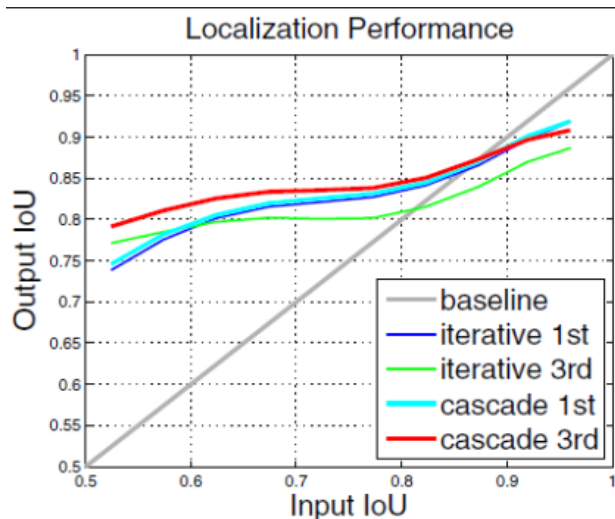


所有级联R-CNN检测器在各个级联阶段的检测性能

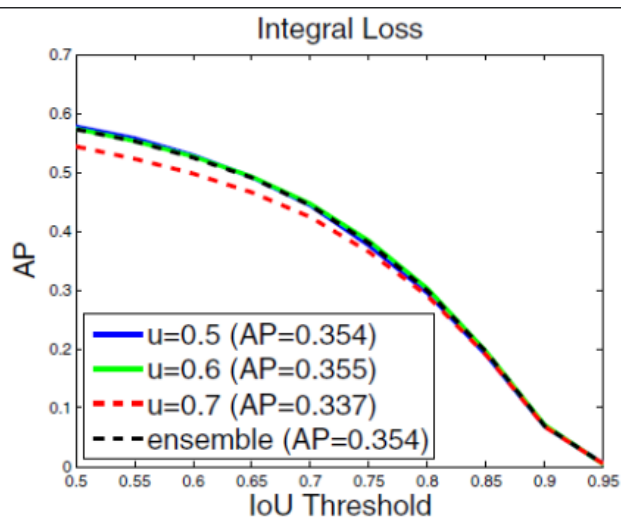
- 图a: $u = 0.7$ 比其他两个都差，通过多级训练， $u = 0.7$ 在几乎所有IoU级别上都达到了最佳性能。

与迭代BBox和积分损耗的比较

	AP	AP ₅₀	AP ₆₀	AP ₇₀	AP ₈₀	AP ₉₀
FPN+ baseline	34.9	57.0	51.9	43.6	29.7	7.1
<i>Iterative BBox</i>	35.4	57.2	52.1	44.2	30.4	8.1
<i>Integral Loss</i>	35.4	57.3	52.5	44.4	29.9	6.9
Cascade R-CNN	38.9	57.8	53.4	46.9	35.8	15.8



(a)



(b)

- 级联回归器在几乎所有的IoU级别上都优于迭代BBox。