**ATSS（CVPR2020）**

anchor-free的检测出了很多优秀的论文，不少大佬都有对anchor-based和anchor-free进行分析，其中指出过说，只铺1个anchor的RetinaNet和类似FCOS的anchor-free方法非常像，这也是我们这个工作的出发点。既然两者非常像，为什么性能存在不少的差距？

AP-50.7%

自适应训练样本选择

该方法的两大特性：

1.保证了所有的正样本anchor都是在groundtruth的周围。

2.最主要是根据不同层的特性对不同层的正样本的阈值进行了微调。

Anchor-free和anchor-based的区别是如何选择正样本和负样本

如果采用相同的正负样本定义，则它们的表现无明显差异，表明如何选择正负样本对当前的目标检测很重要

Anchor-free的兴起是因为FPN和Focal Loss

Anchor-free有两种方法：

1.关键点的anchor-free方法：首先定位几个预定义或自学习的关键点，然后绑定对象的空间范围，如ConerNet，ExtremeNet

2.基于中心的方法：用对象的中心点或区域来定义正样本，然后预测从正样本到对象边界的四个距离,如FCOS

Anchor-free方法的优势：

这些无锚检测器能够消除那些与锚相关的超参数，并取得了与基于锚的检测器相似的性能，使其在泛化能力方面更具潜力。

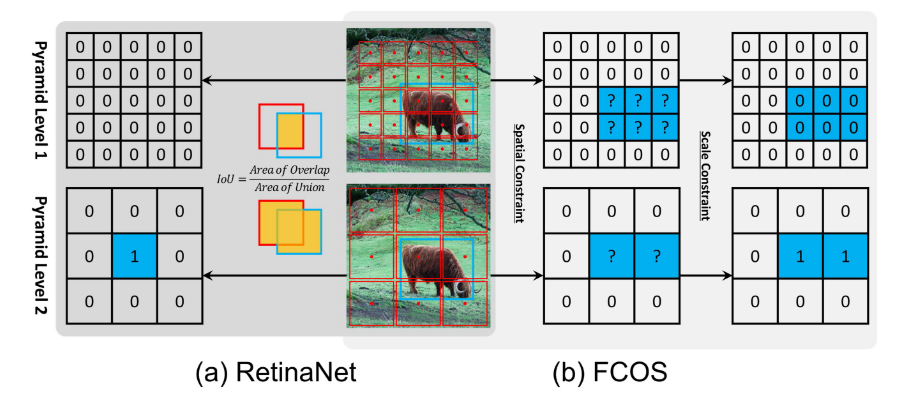
以一级anchor-based检测器RetinaNet[33]和center-basedanchor-free检测器FCOS[56]为例，它们之间主要有3个区别：

(1) 每个位置平铺的锚点。 RetinaNet在每个位置平铺多个锚框，而FCOS在每个位置平铺一个锚点1。

(2)正负样本的定义。RetinaNet根据IoU区分正负样本，FCOS利用空间和尺度约束来确定正负样本

(3) 回归开始状态。RetinaNet从预设的锚框回归对象边界框，而 FCOS 从锚点定位对象。

第一阶段：选正样本



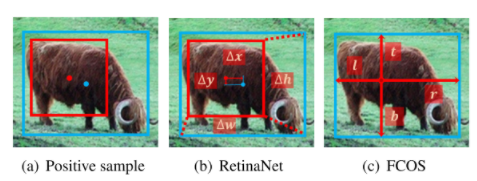


FCOS现在每个特征层上选样本，然后再选定特征层上的做正样本。

RetinaNet基于anchor和GT之间的IoU和设定的阈值来确定正负样本，而FCOS通过GT中心点和铺设点之间的距离和尺寸来确定正负样本。这1点可以从下图的对比中看到，牛这张图像中蓝色框和点表示GT，红色框表示RetinaNet铺设的anchor，红色点表示FCOS铺设的点，左右两边类似表格上的数值表示最终确定的正负样本，0表示负样本，1表示正样本。

。

第二阶段：预测过程



自适应的选取正样本的方法

1.对于每个输出的检测层，选计算每个anchor的中心点和目标的中心点的L2距离，选取K个anchor中心点离目标中心点最近的anchor为候选正样本（candidate positive samples）

2.计算每个候选正样本和groundtruth之间的IOU，计算这组IOU的均值和方差

3.根据方差和均值，设置选取正样本的阈值：t=m+g ；m为均值，g为方差

4.根据每一层的t从其候选正样本中选出真正需要加入训练的正样本

5.训练

**Probabilistic Anchor Assignment with IoU Prediction for Object Detection(ECCV2020)**

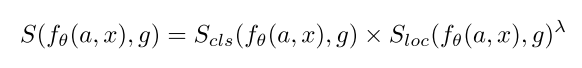
**锚点自适应分为GT的正负样本**

划分正负样本你的方法，设置anchor与GT的IoU阈值，大于该阈值为正样本，小于它为负样本

缺陷：它有一个明显的局限性，因为它忽略了实例的实际内容，这些内容可能包含嘈杂的背景、附近的对象或要检测的目标对象的很少有意义的部分（不能一刀切）

Probabilistic Anchor Assignment Algorithm

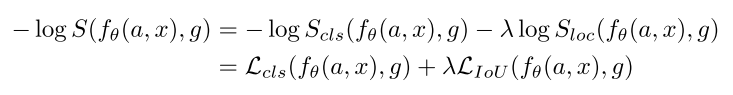
反映锚点质量的分数



其中Scls、Sloc和λ分别是给定g的anchor的分类和定位分数和控制两个分数相对权重的标量

x是输入图像和 fθ是参数为 θ 的模型





Lcls为二元交叉熵损失1

Liou为Iou损失

多模态



w1,m1,p1 和 w2,m2,p2 分别表示两个高斯分布的权重、均值和精度。 给定一组锚点分数，可以使用期望最大化 (EM) 算法优化此 GMM 的可能性。

用 EM 估计的 GMM 参数，每个

可以确定锚是正样本还是负样本。 有了这些概率值，可以使用各种技术将锚点分成两组。 图 3 说明了基于锚点概率的分离边界的不同示例。 使用这些边界方案之一的提议算法在程序 1 中描述。为了计算锚点分数，锚点首先分配给最高 IoU 的 GT（第 3 行）。 为了使 EM 高效，我们从每个金字塔级别（第 5-11 行）收集前 K 个锚点并执行 EM（第 12 行）。 Non-top K anchors 被指定为负样本（第 16 行）

论文解读

https://zhuanlan.zhihu.com/p/254683633

**FSAF(CVPR2019)**

**由于特征金字塔中有多个层次，为每个实例选择合适的特征层次是一个关键问题。基于锚的方法通过锚匹配机制进行隐式选择，该机制基于尺度和纵横比等特殊启发式算法**

1.概述

在训练期间，我们将每个实例动态分配到最合适的特征级别。在推理时，FSAF 模块可以独立工作，也可以与基于锚的分支共同工作。

如何将无锚框模块应用于传统的anchor-based目标检测模块？

1）如何在网络中创建无锚分支

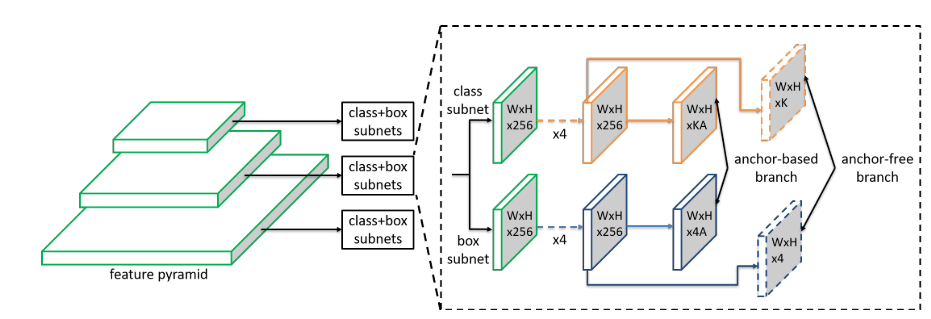
2）如何为无锚分支生成监督信号；

3）如何为每个实例动态选择特征级别；

4）如何联合训练和测试无锚和基于锚的分支。

2.网络结构

特征金字塔由具有从 P3 到 P7 的级别的骨干网络构建，其中 l 是金字塔级别，Pl 具有输入图像的 1/2l 分辨率。为简单起见，仅显示了三个级别。 金字塔的每一层都用于检测不同尺度的对象。 为此，将分类子网和回归子网附加到 Pl。 它们都是小型的全卷积网络。 分类子网为每个 A 锚点和 K 个对象类预测每个空间位置的对象概率。 回归子网预测从每个 A 锚点到附近实例的 4 维类无关偏移量（如果存在）。



**FSAF模型细节**：

在 RetinaNet 之上，我们的 FSAF 模块引入了

每个金字塔级别只有两个额外的卷积层，如图中的虚线特征图所示。这两层分别负责无锚分支中的分类和回归预测。

更具体地说，一个带有 **K 个过滤器**的 3 × 3 conv 层被**附加**到分类子网中的特征图上，然后是 sigmoid 函数，与基于锚点的分支并行。 它为 K 个对象类预测每个空间位置的对象概率。（anchor-free的每个位置相当于只有一个anchor，有K个类别，那么该分类子网的卷积核只有K个）

（而anchor-based网络每个位置有A个anchor，对于每个anchor要预测K个类别，因此卷积核数是K\*A）

类似地，一个带有四个过滤器的 3×3 conv 层被附加到回归子网中的特征图，然后是 ReLU [26] 函数。 它负责预测以无锚方式编码的框偏移量。 为此，anchor-free 和anchor-based 分支以多任务方式共同工作，共享每个金字塔级别的特征。

3.损失函数

对于一个物体实例，如果有k个类别，并且bbox的坐标为b=[x,y,w,h]这里的b是groundturth

(x,y)是bbox的中心坐标，w和h分别是bbox的宽度和高度

The instance can be assigned to arbitrary fea- ture level Pl during training.

每个实例在训练时可以被分配到任意的特征层PI中

下面的是GT在每个特征层上的投影

这是b的在特征层Pl上的映射

**Pl has 1/2l resolution of the input image**





经过比例缩放后变成以下

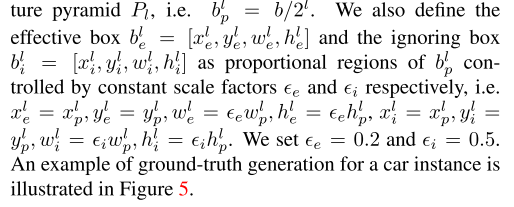
effective box bl（可能是GT投影后的框周边的有效区域）



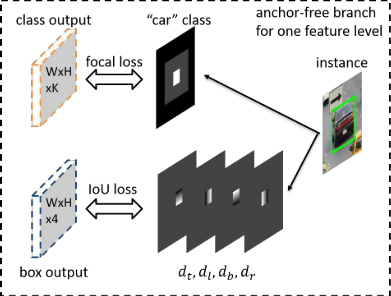
ignoring box



有效框和可忽略框的细节



4.Classification Output分类输出



分类输出要对照的GTbox是K个map，每个map对应一个类别

1.有效框是图中白色部分

作为正样本区域，表明实例的存在

2.灰色区域



gradients in this area are not propagated back to the net- work.

这个区域没法梯度反向传播

3.the ignoring boxes in adjacent feature levels (bl−1, bl+1 ) are also ignoring regions if exists.

在GT map的黑色区域代表没有物体

The rest region of the ground-truth map is the negative (black) area filled by zeros, indicating the absence of objects.

anchor-free的总分类损失

所有非忽略区域的focalloss之和 ，由所有有效框区域内的像素总数归一化。

5.边界框回归输出

idea：

自适应的为不同的物体选不同特征层进行融合（nasFPN已经很多人做过了）

不考特征的融合，考虑训练的时候用几个尺度，也就是使用几个特征层，把loss大的筛掉

效果不一定好

**Soft Anchor-Point Object Detection**

**思想：为每个实例选择多个特征层**

**与FSAF不同的是FSAF limited to one level per instance**

**（一个实例只在一个特征层上进行训练）**

anchor-free方法分为两大类,anchor point detection（FCOS这种）key point detection（CornerNet这种）

anchorpointdetection具有 速度上的优势

anchor-free的优势

无锚检测器通常有一些优点：1）无需手动调整 锚配置的超参数； 2) 通常检测头的架构更简单； 3）更少的训练记忆成本

对象边界框编码和解码为具有相应点到边界距离的锚点，其中锚点是 金字塔特征图，它们与它们所在位置的特征相关联，就像锚框一样。

anchor point detection的优点

1）更简单的网络架构；

2）更快的训练和推理速度；

3）从特征金字塔的增强中受益的潜力[33,21,26]；

4) 灵活的特征级别选择

keypoint detection优点

准确性更强

问题：什么阻碍了anchorpoint检测器的精度？

提出了Soft Anchor Point Detector

我们从anchor point的角度制定了检测问题，并将无效训练确定为阻碍锚点检测器在特征金字塔层内和跨层级探索更多网络能力潜力的主要障碍

To achieve this, we formulate the detection problem from the anchor point’s per- spective and identify ineffective training as the major obstacle impeding anchor- point detector from exploring more potentials of network power both within and across the feature pyramid levels.

传统的训练策略有两个被忽视的问题

1.false attention within each pyramid level每个金字塔级别内的错误注意力

对于同一金字塔层级的anchorpoint（就是小块块），收到错误注意力的会在推理过程中产生具有不必要的高置信度但定位较差的检测，从而抑制一些定位准确但分数较低的锚点，混淆后处理步骤

非极大值抑制中高分检测通常优先于低分检测，从而导致在严格的 IoU 阈值下 AP 分数较低

问题：错误注意力是什么？

2. feature selection across all pyramid levels所有金字塔级别的特征选择–有点像FSAF

对于不同金字塔层级上相同位置的锚点，他们的相关联的特征是相似的

但是它们对网络损失的贡献程度是在没有仔细考虑的情况下决定的

当前的方法基于特殊的启发式方法（例如实例规模）进行选择，并且通常每个实例仅限于一个级别。 这会导致未选择的特征的浪费。（浪费指的是在这一特征层上效果不好，然而还进行了训练，浪费计算资源）

解决方法：

提出新的训练策略：

soft-weighted anchor points和soft-selected pyramid levels

对于同一金字塔级别的锚点，我们根据它们与实例框的几何关系（距离）重新加权它们对网络损失的贡献，从而减少虚假注意力（这里的注意力指的是广义的注意力，不是注意力机制）。 我们认为，越靠近实例边界，由于特征未对齐，锚点就越难精确定位对象，它们对网络损失的贡献就越小。

我们通过其金字塔级别的依赖于实例的“参与”程度进一步重新加权锚点。 我们实现了一个轻量级的特征选择网络来学习给定对象实例的每级“参与”度。 特征选择网络与检测器联合优化，不参与检测器推理。（像FSAF）

how to make the anchor-free detection head better?

how to utilize the power of feature pyramid better?

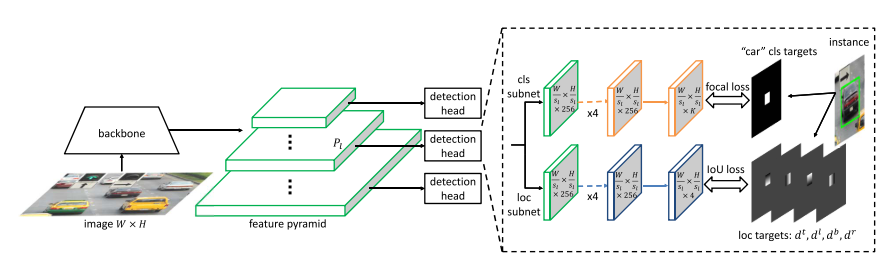
分别提出两种训练策略：soften optimization techniques，

1）soft-weighted anchor points

2）soft-selected pyramid levels

基线里用到了TridentNet（好像也是anchor-free的）

1.模型细节：



金字塔级别表示为 Pl，其中 l 表示级别编号，它具有输入图像大小 W × H 的 1/sl 分辨率。sl 是特征步长，sl = 2l。 l 的典型范围是 3 到 7

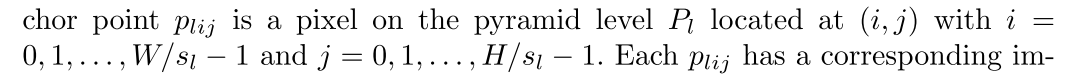
一个检测头有两个特定任务的子网，即分类子网和定位子网。它们都有五个 3 × 3 的卷积层。分类子网为 K 个对象类中的每一个预测对象在每个锚点位置的概率。如果锚点为正（接下来定义），则定位子网预测从每个锚点到附近实例边界的 4 维类无关距离。

2.监督目标

是金字塔Pl层的像素（位置）

W是原图宽度，H是原图高度

具体i，j的范围如下所示（原图对步距的放缩）



这里定义了图像空间位置（将特征图上的点映射为原图上的点）

**Xlij = sl(i + 0.5) 和 Ylij = sl(j + 0.5)**

GT的边界框

B=（c,x,y,w,h）



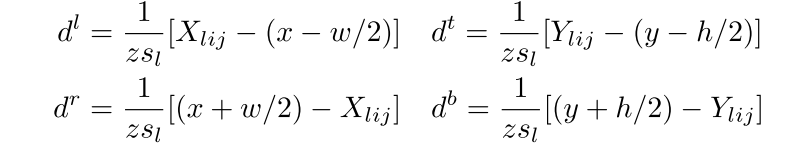
Bv是B的中心点固定，宽高伸缩



**定义正负样本的方法**

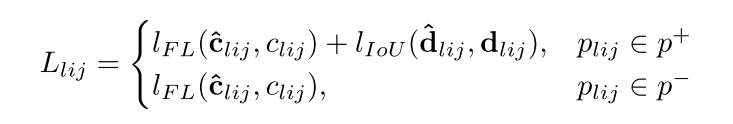
当且仅当某个实例 B 分配给 Pl 并且 plij 的**图像空间位置**（Xlij，Ylij）在 Bv 内部时，锚点 plij 为正，否则为负锚点。

对于正锚点，其分类目标为c，定位目标分别计算为从锚点到B的左、上、右、下边界的归一化距离d = (dl, dt, dr, db)

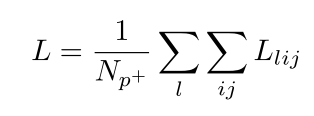


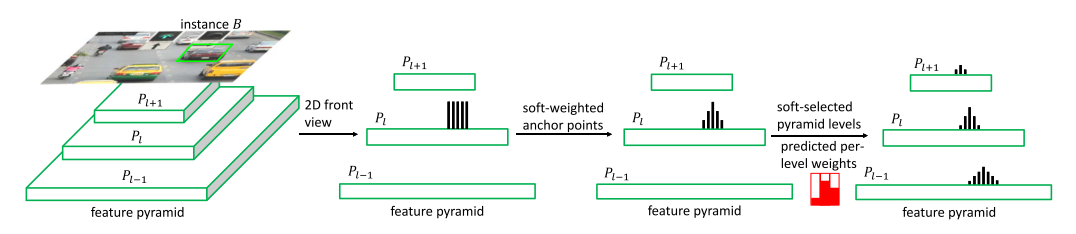
这个公式其实就是坐标变换

z 是归一化标量。 对于**负锚点，它们的分类目标是背景（c = 0），定位目标设置为空**，因为它们不需要学习定位目标，只学习分类



其中 p+ 和 p− 分别是正锚点和负锚点的集合。 整个网络的损失是所有锚点损失的总和除以正锚点的数量





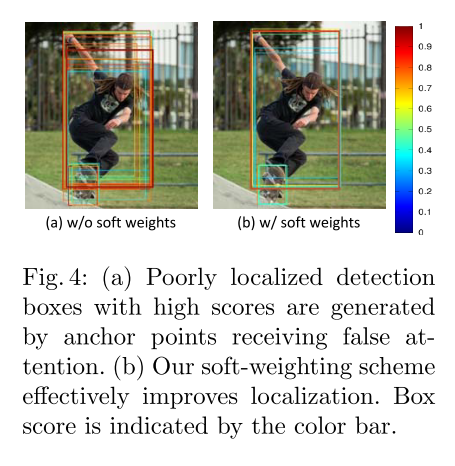
图中黑条是正样本对网络损失贡献的权重

该过程的亮点是将anchorpoint作为一组在特征金字塔级别内和跨级别进行联合优化

细节（详细阐述，最后会和上面内容整合）

**Soft-Weighted Anchor Points**

**下面图片的粗体是GTbox**



传统训练方法存在问题：

1.一些高置信度但是定位分数低的会抑制定位准确但是置信度略低的（因为优先考虑置信度，后考虑边界框回归参数，通过NMS抑制）

2. 即他们得到同等的关注。对于 Bv 内部的一组锚点，它们的空间位置和关联特征是不同的。所以他们定位B的能力也不同。我们认为靠近实例边界的锚点没有与实例很好地对齐的特征。它们的特征往往会受到实例之外的内容的影响，因为它们的感受野包含来自背景的太多信息，导致精确定位的表示能力较低。因此，强迫这些锚点与具有强大特征表示的锚点一样表现会误导网络。与训练中围绕中心的锚点相比，对靠近实例边界的锚点的关注较少。换句话说，网络应该更多地关注优化具有强大特征表示的锚点，并减少对他人的错误关注。

解决方法：

**Soft-Selected Pyramid Levels**

将实例分配给多个特征级别可以在一定程度上提高性能。但是分配太多级别可能会严重损害性能

所以要分配适当的特征级别

Therefore, the solution lies in reweighting the pyramid levels for each in-

stance.

因此，解决方案在于为每个实例分配金字塔级的权重。

换句话说，根据特征响应为每个金字塔级别分配一个权重，使选择变得柔和。 这也可以看作是将一部分实例分配给一个级别。

那么如何决定每个实例每个金字塔级别的权重呢？ 我们建议训练一个**特征选择网络**来**预测软特征选择的权重**。 网络的输入是从所有金字塔级别中提取的**与实例相关的特征响应**。 这是通过将 RoIAlign 层 [7] 应用于每个金字塔特征然后进行连接来实现的，其中 RoI 是实例真实值框然后提取的特征通过特征选择网络输出概率分布的向量，如图6所示。我们使用概率作为软特征选择的权重。

特征选择网络

输入：每个特征金字塔层提取出来的实例特征

怎么把每个实例的特征提取出来？

这是通过将 RoIAlign 层 [7] 应用于每个金字塔特征然后进行连接来实现的，其中 RoI 是实例GT框。

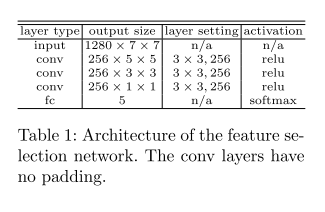
输出：一个关于概率分布的向量

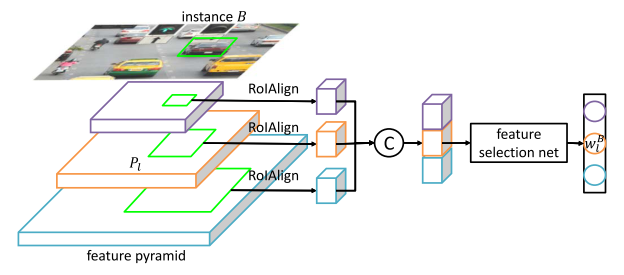
特征选择网络的设计：

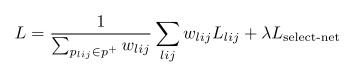
由三个3\*3的卷积层（没有padding）构成

接着ReLU函数和一个带softmax的全连接层

特征选择网络与检测器联合训练。 交叉熵损失用于优化，真实值是一个单热向量，指示哪个金字塔级别的损失最小，如 FSAF 模块中所定义







权重越大影响越大

**RoI**

**RoI（Region of Interest）**是从原始图像中提取的区域

**RoI 不是一个边界框**。它可能看起来像一个，但它只是一个进一步处理的proposal

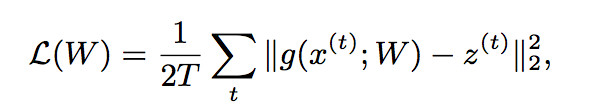
RoIAlign

**知识蒸馏**

分类 头部蒸馏和颈部蒸馏

知识蒸馏

imic的论文主要都是局限在较为简单的classification 的任务中，而对于较为复杂一些的detection任务，直接套用以前的方案则显得不行。本文提出了一种学习feature map来实现Object Detection任务上mimic的方法



W为小模型的weights，g(x;W) 为小模型的输出，z为学习的大模型的输出

对于feature map进行mimic

单纯使小模型学习大模型的feature map并不能work，原因在于feature map的维度太高，包含太多全局的信息，而对于仅有少量object的feature map，通常只有微弱的响应。因此，该文中提出了一个新的卷积网络mimic方法，即将学习整个feature map变为学习proposal采样之后的feature来简化任务。

了一个新的卷积网络mimic方法，即将学习整个feature map变为学习proposal采样之后的feature来简化任务

IMPROVE OBJECT DETECTION WITH FEATURE-BASED KNOWLEDGE DISTILLATION: TOWARDS ACCURATE AND EFFICIENT DETECTORS(ICLR2021)

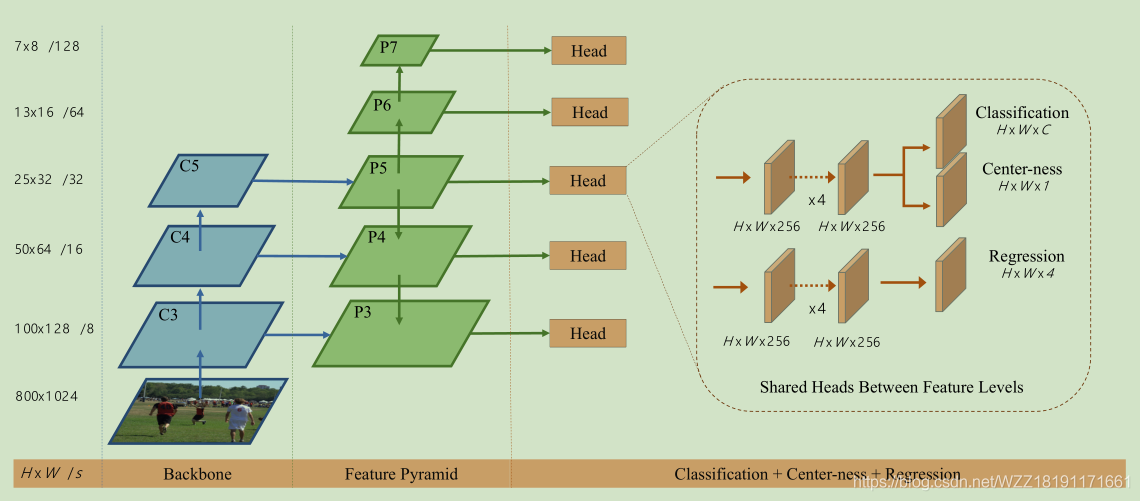
**FCOS**

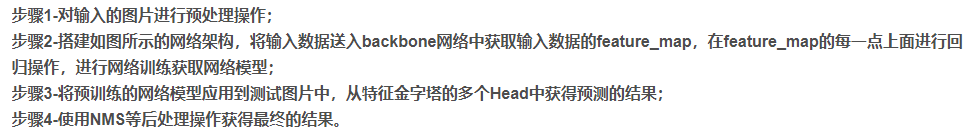
（补充：FCOS判断正样本的方法：将feature\_map中的每一个点(x,y)映射回原始的输入图片中；然后，如果这个映射回原始输入的点在相应的GT的BB范围之内，而且类别标签对应，我们将其作为训练的正样本块

FCOS详细分析：<https://blog.csdn.net/WZZ18191171661/article/details/89258086>）

最后，我们提出了我们提出的“中心性”分支，它有助于抑制检测到的低质量边界框并大幅提高整体性能

最后，我们提出了我们提出的“中心性”分支，它有助于抑制检测到的低质量边界框并大幅提高整体性能。





Q:损失函数是NMS之后的预测框与GTbox的偏差还是每个层输出的预测框和GTbox的偏差？

每个预测层输出的偏差

正负样本定义：

博文有助于理解

http://www.360doc.com/content/20/0428/22/99071\_909106981.shtml

Idea

在SAPD或者FSAF

融入AFSS或者Probabilistic Anchor Assignment with IoU Prediction for Object Detection

这类自适应选择正负样本的方法

目标检测补充知识：

正样本:即待检测的物体

负样本：背景

基于anchor的方法，训练过程是bbpx坐标位置的回归

Anchor-free的方法 训练过程是anchor point的中心点宽高四个参数的回归

正负样本的多少对模型的精度影响很大

正负样本对训练过程的作用

https://blog.csdn.net/clearch/article/details/80224223

不同anchor会有不同的·预测值，对于一张图片进行NMS，在冗余的框中实现互相抑制

ResNet