FCOS

FCOS: Fully Convolutional One-Stage Object Detection

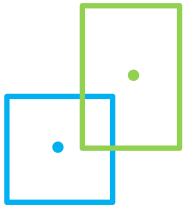
Zhi Tian   Chunhua Shen*   Hao Chen   Tong He
The University of Adelaide, Australia

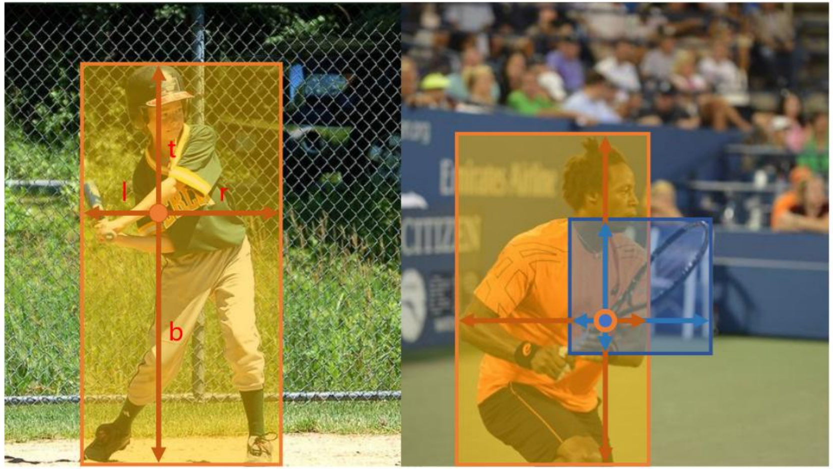
2019 CVPR

Anchor-Free

One-Stage

FCN-base



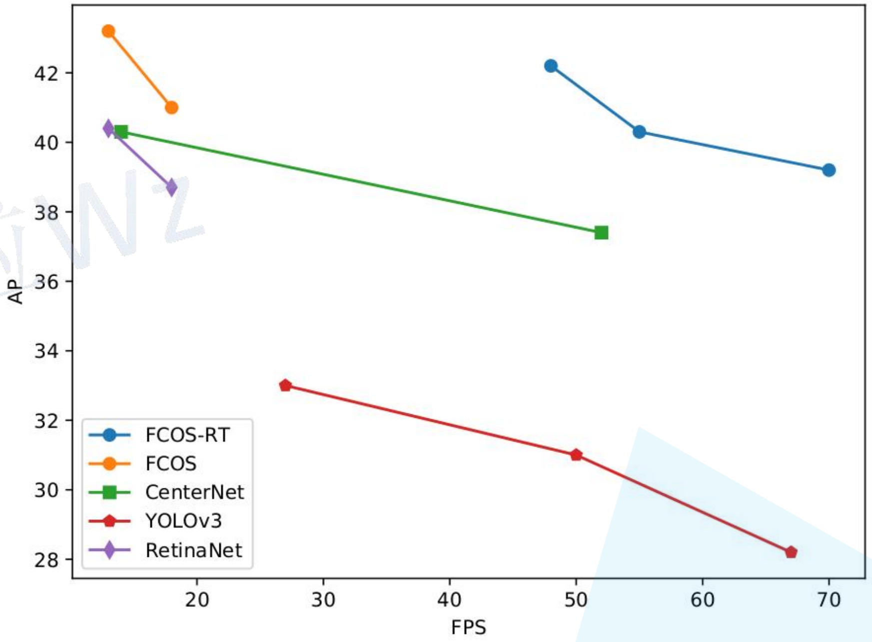


论文下载地址：https://arxiv.org/abs/1904.01355;   https://arxiv.org/abs/2006.09214
博文： https://blog.csdn.net/qq_37541097/article/details/124844726
公众号“阿喆学习小记”输入FCOS获取

FCOS

目录

- 0 前言
- 1 FCOS网络结构
- 2 正负样本的匹配
- 3 损失计算
- 4 其他
  - 4.1 Ambiguity问题
  - 4.2 Assigning objects to FPN



FCOS

前言

1. 检测器的性能和Anchor的size以及aspect ratio相关，比如在RetinaNet中改变Anchor能够产生约4%的AP变化。
2. 一般Anchor的size和aspect ratio是固定的，很难处理形状变化大的目标。而且迁移到其他任务中时，可能需要重新设计Anchor。
3. 为了达到更高的召回率，一般需要在图片中生成非常密集的Anchor Boxes。在训练时绝大部分的Anchor Boxes都会被分为负样本，这样会导致正负样本极度不均。
4. Anchor的引入使得网络在训练过程中更加的繁琐。

FCOS

网络结构

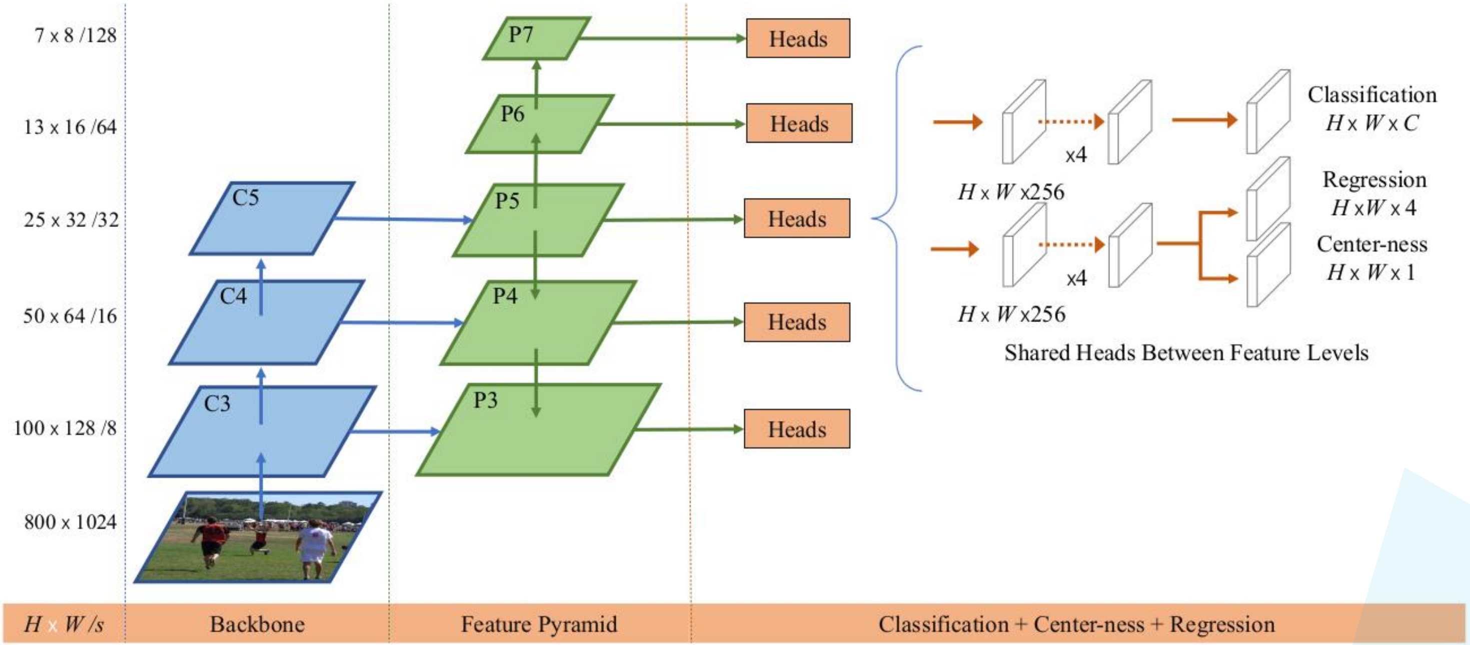
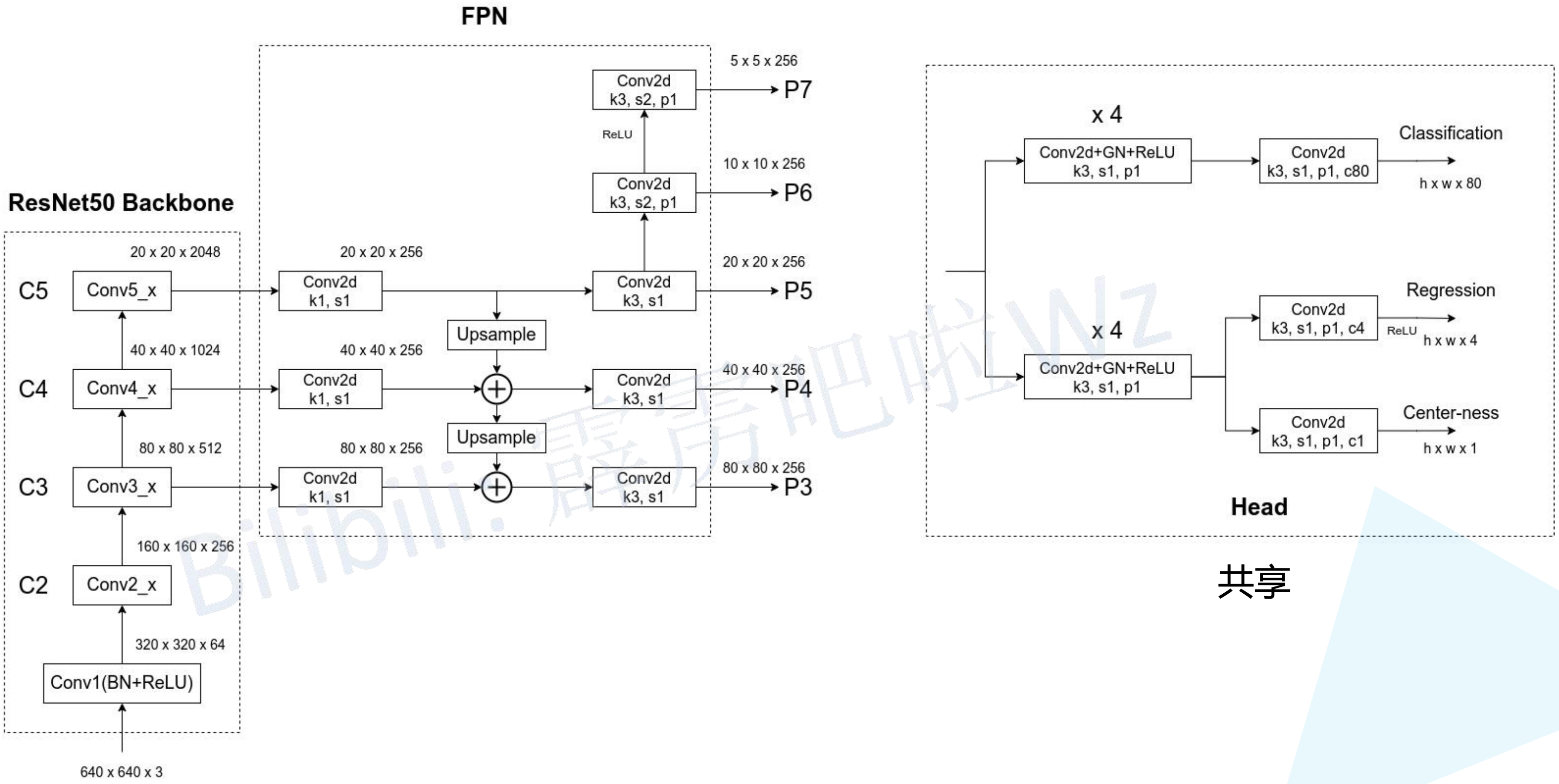


Fig. 2. **The network architecture of FCOS**, where C3, C4, and C5 denote the feature maps of the backbone network and P3 to P7 are the feature levels used for the final prediction. $H \times W$ is the height and width of feature maps. '$/s$' ($s = 8, 16, ..., 128$) is the down-sampling ratio of the feature maps at the level to the input image. As an example, all the numbers are computed with an 800×1024 input.

FCOS

网络结构

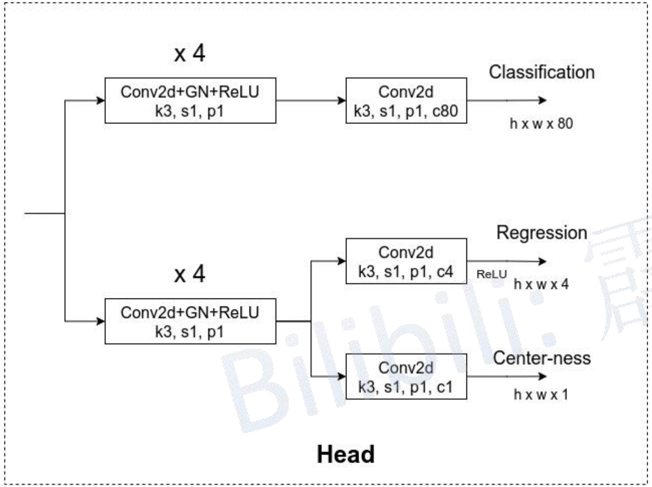


FPN
ResNet50 Backbone
C5
Conv5_x
20 x 20 x 2048
C4
Conv4_x
40 x 40 x 1024
C3
Conv3_x
80 x 80 x 512
C2
Conv2_x
160 x 160 x 256
Conv1(BN+ReLU)
320 x 320 x 64
640 x 640 x 3
Conv2d
k1, s1
20 x 20 x 256
40 x 40 x 256
80 x 80 x 256
Upsample
Upsample
Conv2d
k3, s1
Conv2d
k3, s2, p1
ReLU
5 x 5 x 256
P7
10 x 10 x 256
P6
20 x 20 x 256
P5
40 x 40 x 256
P4
80 x 80 x 256
P3
x 4
Conv2d+GN+ReLU
k3, s1, p1
Conv2d
k3, s1, p1, c80
Classification
h x w x 80
x 4
Conv2d+GN+ReLU
k3, s1, p1
Conv2d
k3, s1, p1, c4
ReLU
Regression
h x w x 4
Conv2d
k3, s1, p1, c1
Center-ness
h x w x 1
Head

共享

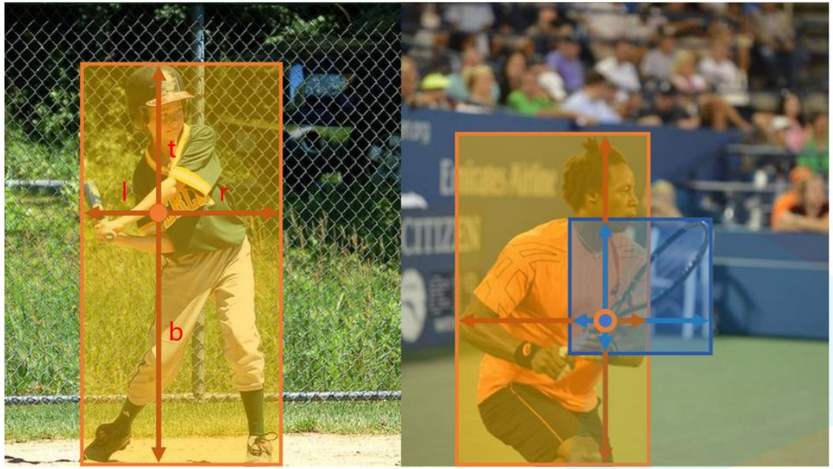
FCOS

网络结构



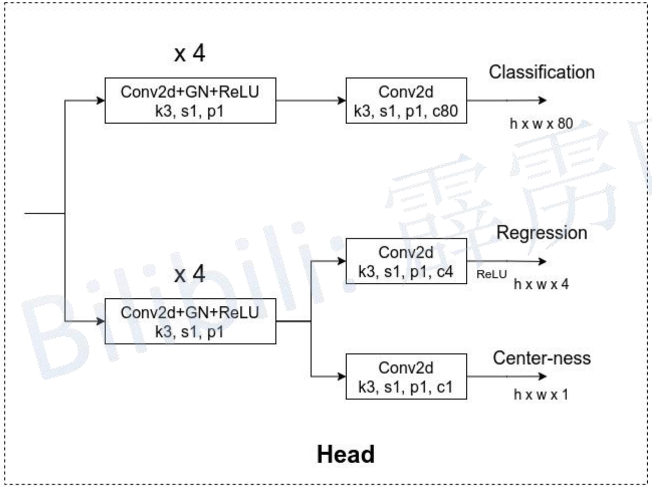
$$x_{min} = c_x - l \cdot s \quad y_{min} = c_y - t \cdot s$$

$$x_{max} = c_x + r \cdot s \quad y_{max} = c_y + b \cdot s$$



FCOS

网络结构

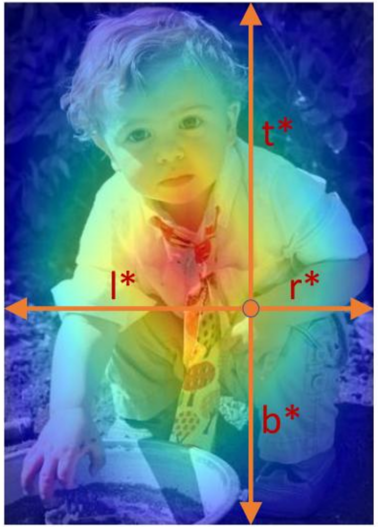


$$centerness^* = \sqrt{\frac{min(l^*, r^*)}{max(l^*, r^*)} \times \frac{min(t^*, b^*)}{max(t^*, b^*)}}$$

	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
None	33.5	52.6	35.2	20.8	38.5	42.6
center-ness†	33.5	52.4	35.1	20.8	37.8	42.8
center-ness	**37.1**	**55.9**	**39.8**	**21.3**	**41.0**	**47.8**

Table 4 – Ablation study for the proposed center-ness branch on `minival` split. "None" denotes that no center-ness is used. "center-ness†" denotes that using the center-ness computed from the predicted regression vector. "center-ness" is that using center-ness predicted from the proposed center-ness branch. The center-ness branch improves the detection performance under all metrics.

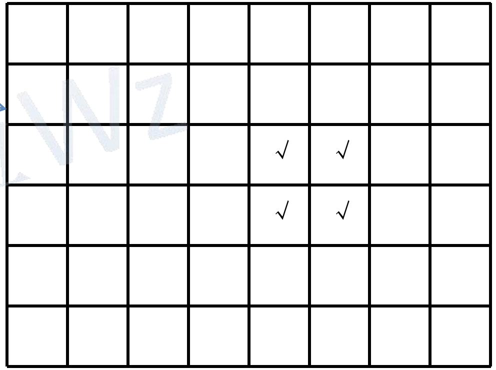
2019 version



FCOS

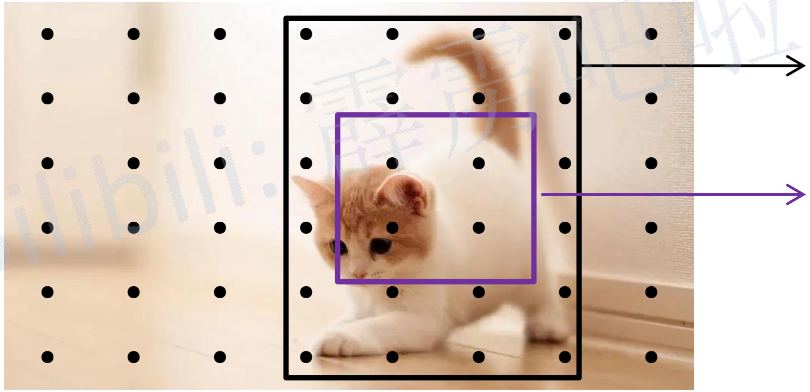
正负样本匹配

feature map



positive samples (2019 version)

positive samples (2020 version)



GT box

sub-box

$$(c_x - rs, c_y - rs, c_x + rs, cy + rs)$$

r	AP	AP$_{50}$	AP$_{75}$	AP$_S$	AP$_M$	AP$_L$
1.0	38.5	57.2	41.5	22.6	42.3	49.7
1.5	**38.9**	57.5	**42.2**	**23.1**	**42.7**	**50.2**
2.0	38.8	**57.7**	41.7	22.7	42.6	49.9

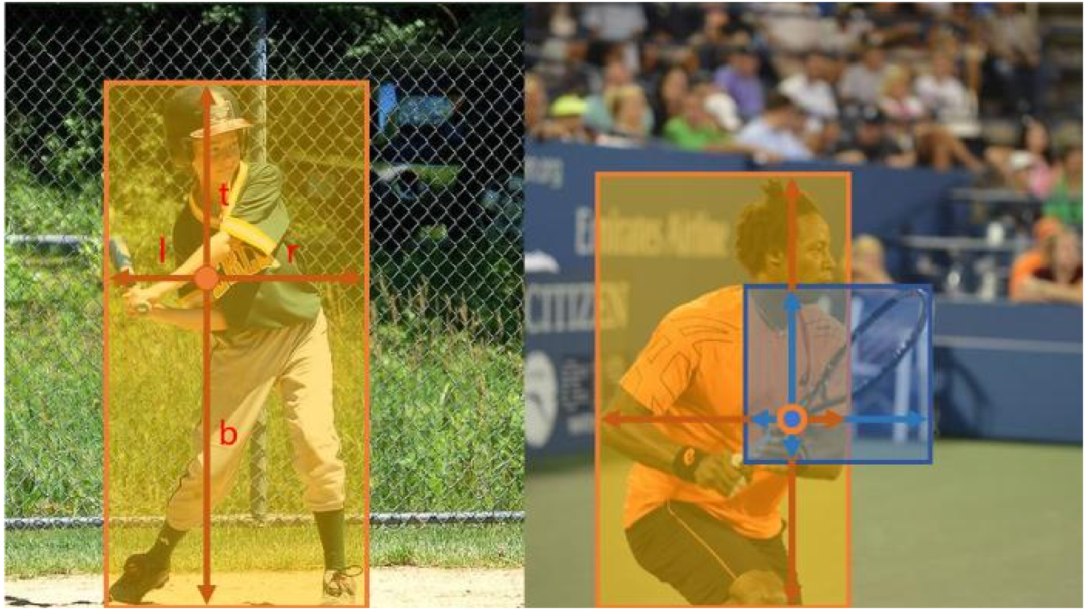
TABLE 6

Ablation study for the radius r of positive sample regions (defined in Section 2.1).

FCOS

正负样本匹配

Ambiguity问题



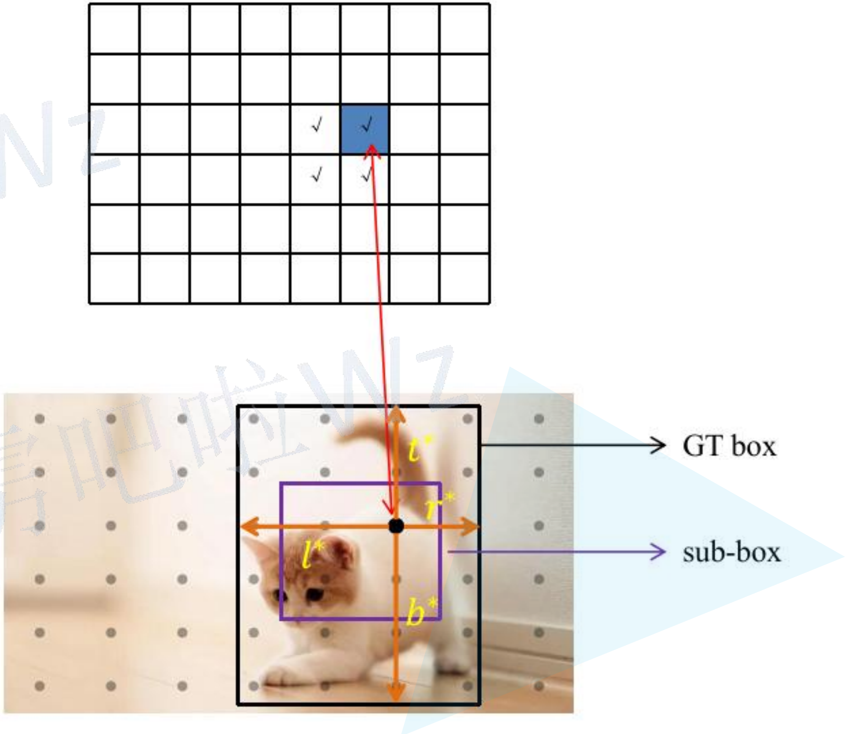
默认分配给面积Area最小的GT Box

FCOS

损失计算

$$L(\{p_{x,y}\},\{t_{x,y}\},\{s_{x,y}\}) = \frac{1}{N_{pos}}\sum_{x,y} \overset{\text{BCELoss + FocalLoss}}{L_{cls}(p_{x,y}, c^*_{x,y})} + \frac{1}{N_{pos}}\sum_{x,y} 1_{\{c^*_{x,y}>0\}} \overset{\text{GIoULoss}}{L_{reg}(t_{x,y}, t^*_{x,y})} + \frac{1}{N_{pos}}\sum_{x,y} 1_{\{c^*_{x,y}>0\}} \overset{\text{BCELoss}}{L_{ctrness}(s_{x,y}, s^*_{x,y})}$$

$$centerness^* = \sqrt{\frac{min(l^*, r^*)}{max(l^*, r^*)} \times \frac{min(t^*, b^*)}{max(t^*, b^*)}}$$



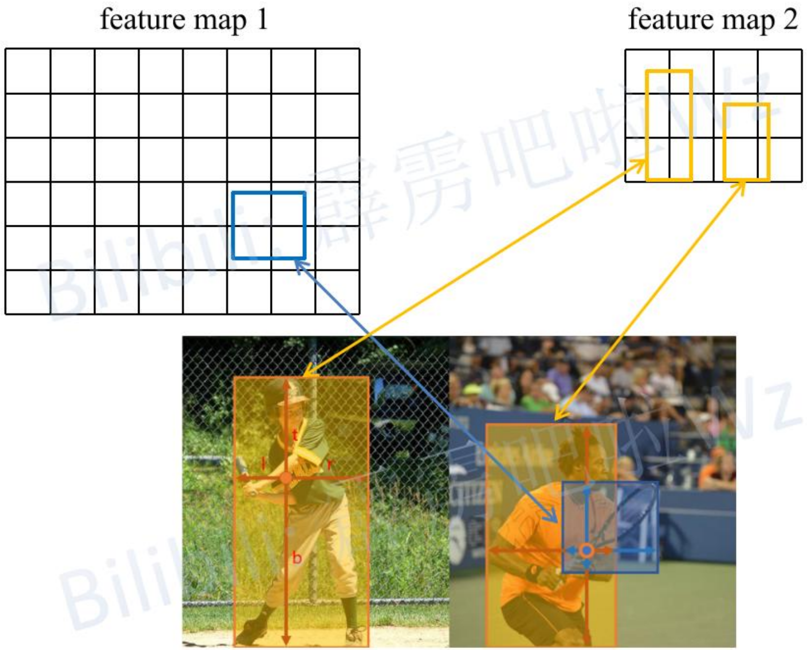
其中：

- $p_{x,y}$表示在特征图(x,y)点处预测的每个类别的score
- $c^*_{x,y}$表示在特征图(x,y)点对应的真实类别标签
- $1_{\{c^*_{x,y}>0\}}$当特征图(x,y)点被匹配为正样本时为1，否则为0
- $t_{x,y}$表示在特征图(x,y)点处预测的目标边界框信息
- $t^*_{x,y}$表示在特征图(x,y)点对应的真实目标边界框信息
- $s_{x,y}$表示在特征图(x,y)点处预测的 `center-ness`
- $s^*_{x,y}$表示在特征图(x,y)点对应的真实 `center-ness`

FCOS

Ambiguity问题

- 不使用FPN结构（仅在P4特征层上进行预测）会存在大量的ambiguous samples（大概占23.16%）
- 使用FPN结构ambiguous samples会大幅降低（大概占7.24%）
- 如果再采用center sampling匹配准则能够进一步降低ambiguous samples的比例（小于3%）

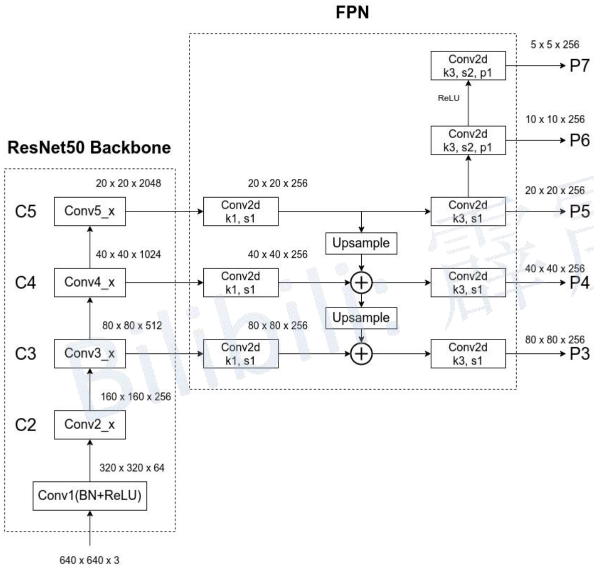


FCOS

Assigning objects to FPN

Strategy	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
FPN	37.7	56.6	40.6	22.2	40.9	49.7
$\sqrt{(h^* \times w^*)}/2$	37.6	56.5	40.6	22.4	41.6	47.3
$\max(h^*, w^*)/2$	38.1	57.0	41.3	22.5	41.8	48.7
$\max(l^*, t^*, r^*, b^*)$	**38.9**	**57.5**	**42.2**	**23.1**	**42.7**	**50.2**

TABLE 7



each level. More specifically, we first compute the regression targets l^*, t^*, r^* and b^* for each location on all feature levels. Next, if a location at feature level i satisfies $\max(l^*, t^*, r^*, b^*) \leq m_{i-1}$ or $\max(l^*, t^*, r^*, b^*) \geq m_i$, it is set as a negative sample and thus not required to regress a bounding box anymore. Here m_i is the maximum distance that feature level i needs to regress. In this work, m_2, m_3, m_4, m_5, m_6 and m_7 are set as $0, 64, 128, 256, 512$ and ∞, respectively. We

$$m_{i-1} < max(l^*, t^*, r^*, b^*) < m_i$$

$$P_i$$