

目标框PASCAL VOC标注格式



width = 1000

height = 654

PASCAL VOC格式的xml标记文件

```
<?xml version="1.0" encoding="ISO-8859-1"?>
- <annotation>
    <folder>bai</folder>
    <filename>trophy.jpg</filename>
    <path>/home/bai/trophy.jpg</path>
    - <source>
        <database>Unknown</database>
    </source>
    - <size>
        <width>1000</width>
        <height>654</height>
        <depth>3</depth>
    </size>
    <segmented>0</segmented>
    - <object>
        <name>trophy</name>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
        <difficult>0</difficult>
        - <bndbox>
            <xmin>187</xmin>
            <ymin>21</ymin>
            <xmax>403</xmax>
            <ymax>627</ymax>
        </bndbox>
    </object>
</annotation>
```

YOLO格式的txt标记文件

class_id	x	y	w	h
----------	---	---	---	---

2	0.295000	0.495413	0.216000	0.926606
---	----------	----------	----------	----------

$x = x_center / width = 295 / 1000 = 0.2950$

$y = y_center / height = 324 / 654 = 0.4954$

$w = (xmax - xmin) / width = 216 / 1000 = 0.2160$

$h = (ymax - ymin) / height = 606 / 654 = 0.9266$

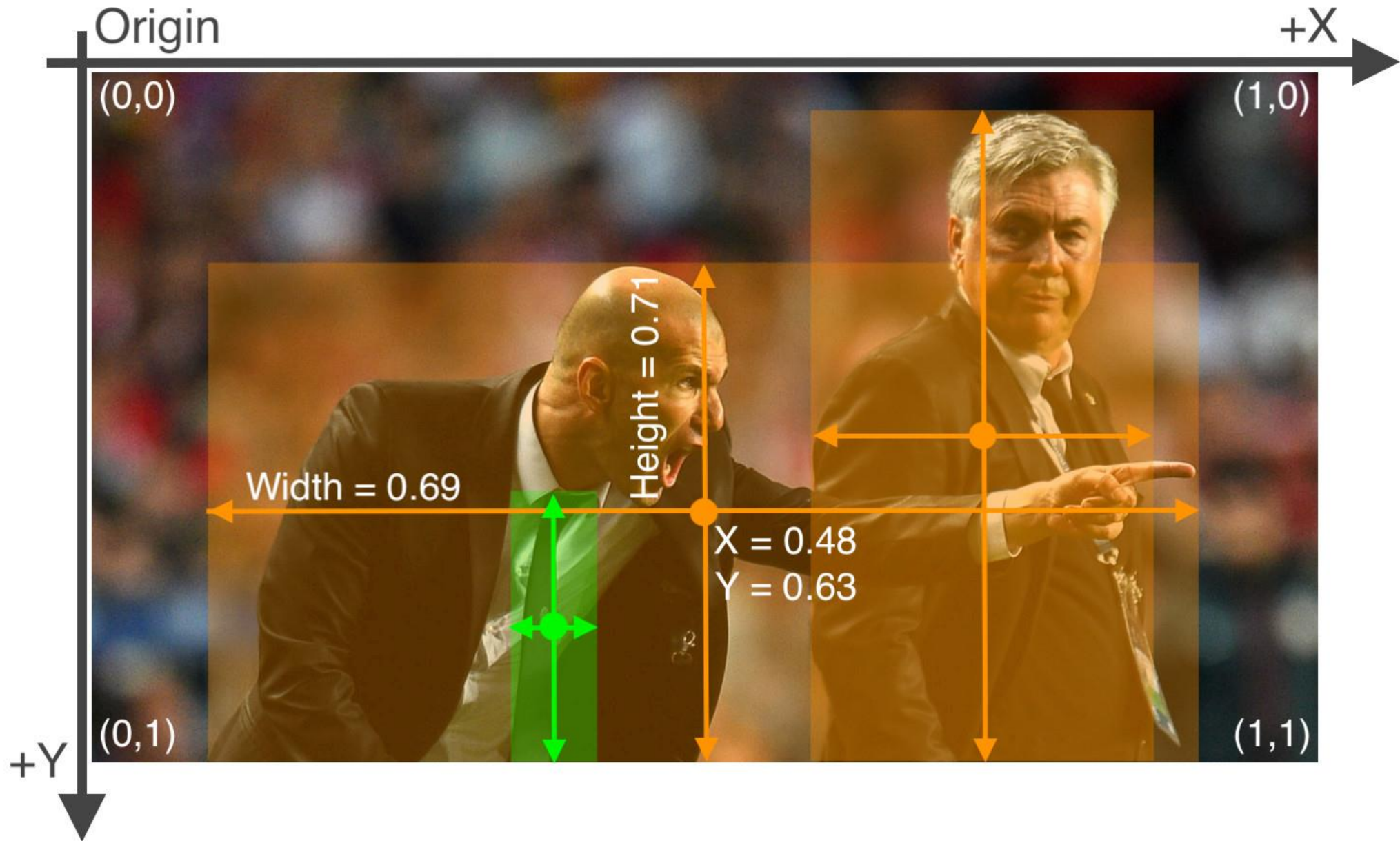
class_id: 类别的id编号

x: 目标的中心点x坐标 (横向) / 图片总宽度

y: 目标的中心的y坐标 (纵向) / 图片总高度

w: 目标框的宽带/图片总宽度

h: 目标框的高度/图片总高度



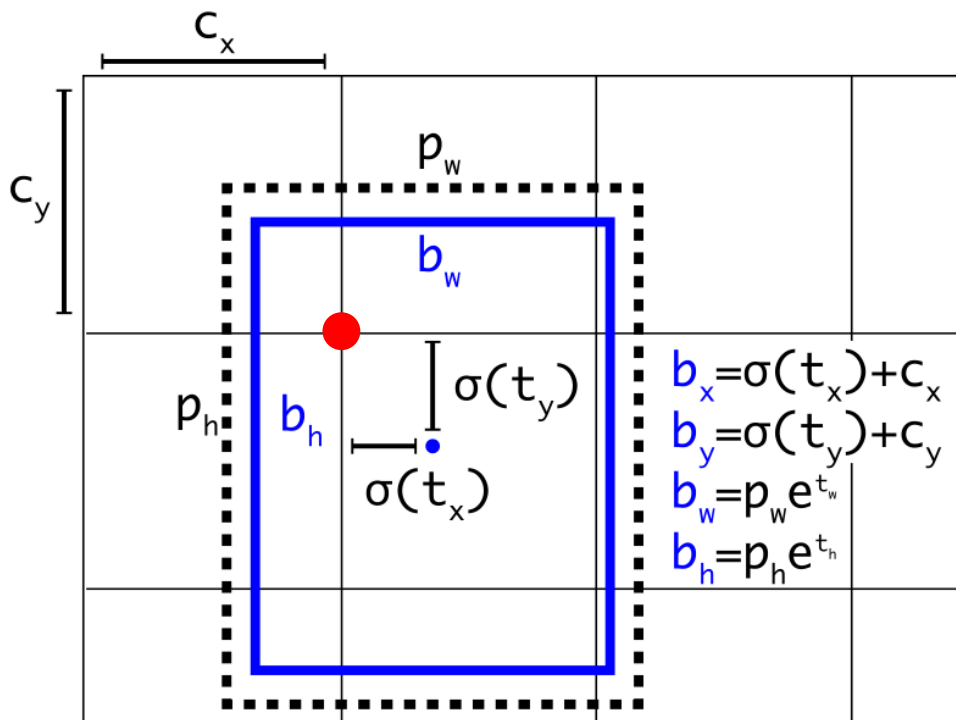
目标框回归

Anchor给出了目标宽高的初始值，需要回归的是目标真实宽高与初始宽高的**偏移量**

- 预测边界框中心点相对于对应cell左上角位置的相对偏移值
- 为了将边界框中心点约束在当前cell中，使用sigmoid函数处理偏移值，使预测偏移值在(0,1)范围内

YOLOv/v4目标框回归公式：

根据**边界框预测的4个offsets** t_x, t_y, t_w, t_h ，可以按如下公式计算出边界框实际位置（坐标值）和大小：



$$b_x = \sigma(t_x) + c_x$$

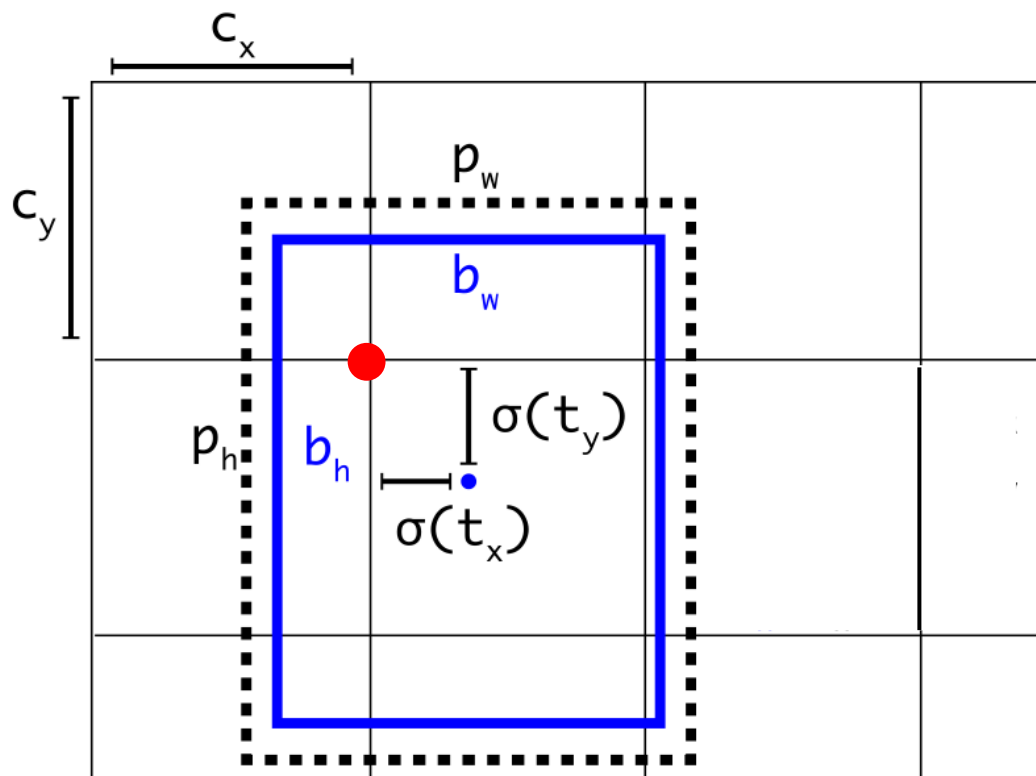
$$b_y = \sigma(t_y) + c_y$$

$$b_w = p_w e^{t_w}$$

$$b_h = p_h e^{t_h}$$

YOLOv5 目标框回归

YOLOv5采用了跨邻域网格的匹配策略，从而得到更多的正样本anchor，可加速收敛



$$b_x = 2\sigma(t_x) - 0.5 + c_x$$

$$b_y = 2\sigma(t_y) - 0.5 + c_y$$

$$b_w = p_w (2\sigma(t_w))^2$$

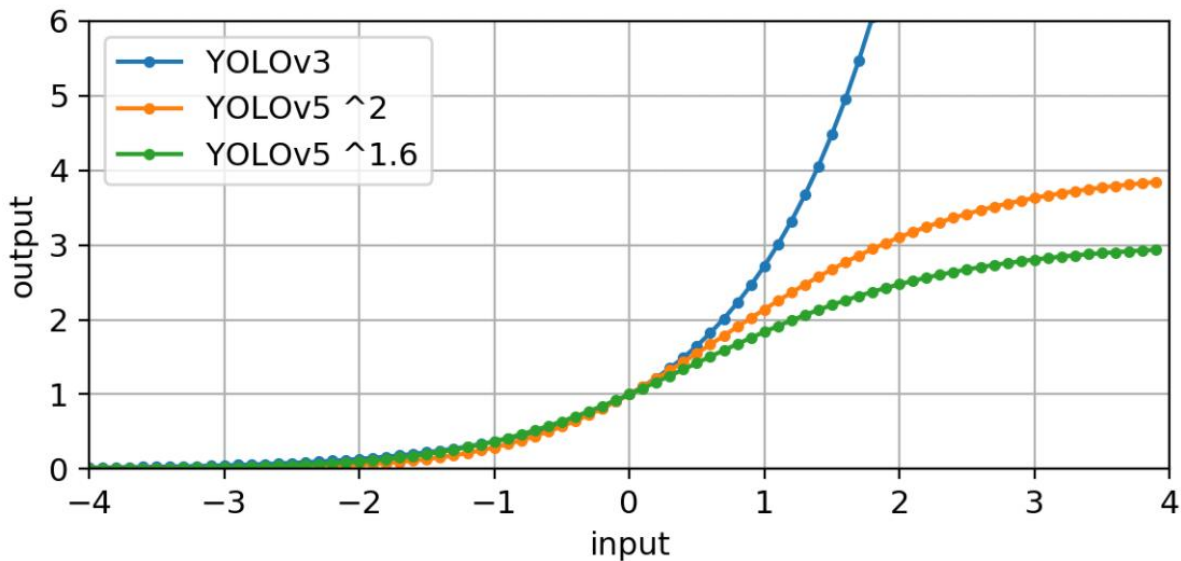
$$b_h = p_h (2\sigma(t_h))^2$$

原始的yolo / darknet框方程式存在严重缺陷。宽度和高度完全不受限制，因为它们只是 $\text{out} = \exp(\text{in})$ ，这很危险，因为它可能导致失控的梯度、不稳定、NaN损失并最终完全失去训练。

对于yolov5，确保通过sigmoid所有模型输出来修补此错误，同时还要确保中心点保持不变 $1 = \text{fcn}(0)$ ，因此模型的标称零输出将导致使用标称锚框大小。当前的方程式将锚点的倍数从最小0限制为最大4，并且锚点-目标匹配也已更新为基于宽度-高度倍数，标称上限阈值超参数为4.0。

[data/hyp.scratch.yaml](#)

`anchor_t: 4.0 # anchor-multiple threshold`



有些bbox由于和anchor匹配度不高，而认为是背景