

目标检测相关细节

叶亮

2020 年 10 月 23 日

1 Loss

1.1 IoU loss

1.2 GIoU loss

1.3 DIoU loss

Code optimization, the calculation of distance between two center points of bboxes.

$$\begin{aligned}(C_1 - C_2)^2 &= ((x1_{C1} + x2_{C1})/2 - (x1_{C2} + x2_{C2})/2)^2 \\ &= ((x1_{C1} + x2_{C1}) - (x1_{C2} + x2_{C2}))^2/4\end{aligned}\tag{1}$$

1.4 CIoU loss

1.5 BIoU loss

2 bbox & anchor

2.1 coder

在 RetinaNet、SSD、Cascade R-CNN 等网络中，网络预测的 bbox 都会进行 Delta xywh 编码，即将原始的 (xmin, ymin, xmax, ymax) 进行编码，计算相对距离，来减小回归的过拟合，提升回归的稳定性。

$$\begin{aligned}\delta_x &= (g_x - p_x)/p_w & \delta_y &= (g_y - p_y)/p_h \\ \delta_w &= \log g_w/p_w & \delta_h &= \log g_h/p_h\end{aligned}\tag{2}$$

上述公式计算得到的 δ 值通常很小，因为网络通常只对 p 进行少量微调，导致回归 loss 比分类 loss 小很多。为了提升学习的有效性， δ 通常需要经过均值和方差进行标准化。

$$\delta'_x = \frac{\delta_x - \mu_x}{\rho_x} \quad (3)$$

3 training

3.1 Optimizer

3.1.1 warmup

warmup 通常有三个方式:linear, constant, exp. 通常需要设置 warmup 的迭代数 $iter_{total}$ 和 warmup 的增加比率 ratio,

$$lr_t = lr_{constant} * ratio \quad (4)$$

$$\begin{aligned} lr_t &= lr_{const} * (1 - k), \\ k &= (1 - iter_t/iter_{total}) * (1 - ratio) \end{aligned} \quad (5)$$

$$lr_t = lr_{const} * k, \quad (6)$$

$$k = ratio^{1 - iter_t/iter_{total}} \quad (7)$$

4 YOLOv4

YOLOv4 的实现过程中,anchor 的生成以及 label 与 anchor 的对应关系的构建方法。

4.1 Anchor generation & target build

与 retinanet, rcnn 系列等 bbox 预测不同的是, yolo 模型推理得到的结果 (x,y,w,h) 需要经过如下公式进行编码:

$$b_x = \rho(t_x) + c_x \quad (8)$$

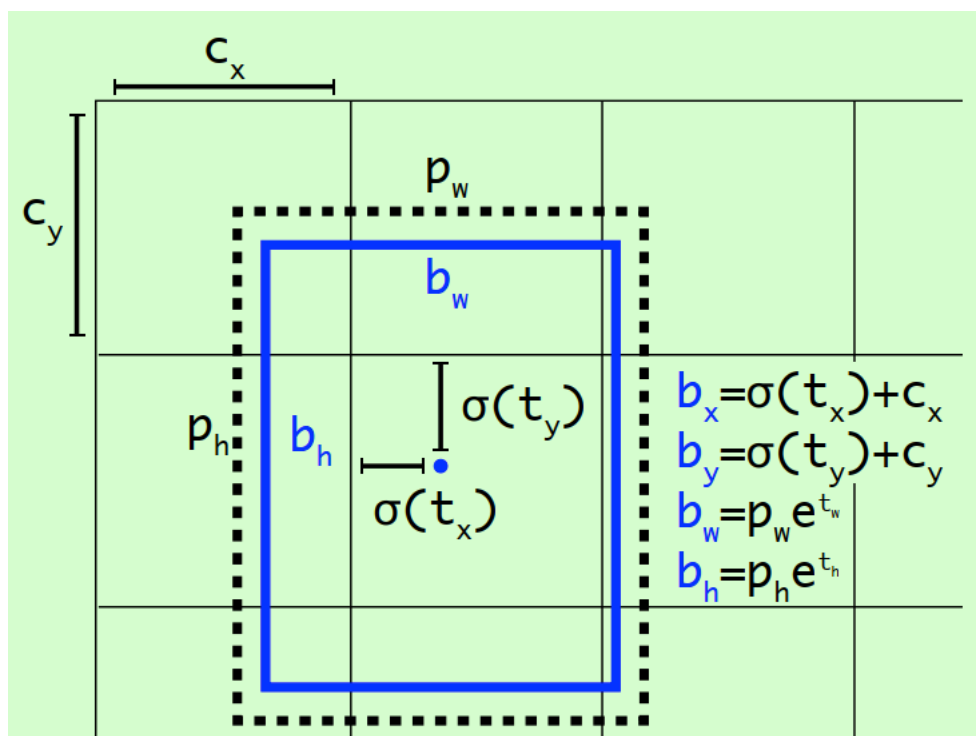


图 1: Yolo bbox prediction

5 backbone

5.1 Regularization

5.1.1 Dropout

原理,dropout 随机丢弃神经元 (全连接中输入神经元), 实现方式为 $keep_{prob}$, 每个神经元生成一个随机数 $k, k < keep_{prob}$ 即丢弃。优点, 该方法有利于分类中泛化能力的提升。

5.1.2 Drop Connect

5.1.3 Drop block

6 Refinedet

6.1 Anchor

Refine 中的 anchor 计算。对于每一个 feature map, 首先计算其 mesh grid, 然后计算每个框的中心点 $(x, y) = ((i+0.5)/featsize, (j+0.5)/featsize)$, 然后根据每个 feature map 对应的 anchor box 的大小, 计算 anchor 的长和宽 $WH_{ki} = box_k / imagesize$, where k means k th feature level, i means i th grid. e.g, 使用四个 feature level, $box = [32, 64, 128, 256]$, $imagesize = 320$, $featsize = [40, 20, 10, 5]$, $aspectratio = [2, 2, 2, 2]$, 那么最终生成 $40 * 40 * 3 + 20 * 20 * 3 + 10 * 10 * 3 + 5 * 5 * 3 = 6375$ 个 anchor。

7 Dataset

7.1 COCO Dataset

COCO dataset 全称 Common Objects in Context. 共有 80 个类。共有三种标注类型: object instance, object keypoints, image captions。使用 json 文件进行存储。

7.1.1 object instance

"info": info, "licenses": [license], "images": [image], "annotations": [annotation], "categories": [category]

8 Detector

8.1 CornerNet, CenterNet

8.1.1 targets computation

在 anchor free 网络中, 生成 gt bbox 的 target 时, 相应的高斯半径的计算方式如下。一共存在三种情况: 1, 生成的 target 与 gt bbox 有重叠部分, 其中一个 corner 在 gt box 内部, 另一个在 gt box 外部。

$$\frac{(w-r)*(h-r)}{w*h+(w+h)r-r^2} \geq iou \quad \Rightarrow \quad r^2 - (w+h)r + \frac{1-iou}{1+iou} * w * h \geq 0 \quad (9)$$

$$a = 1, \quad b = -(w+h), \quad c = \frac{1-iou}{1+iou} * w * h \quad (10)$$

$$r \leq \frac{-b - \sqrt{b^2 - 4*a*c}}{2*a} \quad (11)$$

2. 两个 corner 都在 gt box 里面

$$\frac{(w-2*r)*(h-2*r)}{w*h} \geq iou \quad \Rightarrow \quad 4r^2 - 2(w+h)r + (1-iou)*w*h \geq 0 \quad (12)$$

$$a = 4, \quad b = -2(w+h), \quad c = (1-iou)*w*h \quad (13)$$

$$r \leq \frac{-b - \sqrt{b^2 - 4*a*c}}{2*a} \quad (14)$$

3. 两个 corner 都在 gt box 外部

$$\frac{w * h}{(w + 2 * r) * (h + 2 * r)} \geq iou \quad \Rightarrow \quad 4 * iou * r^2 + 2 * iou * (w + h)r + (iou - 1) * w * h \leq 0 \quad (15)$$

$$a = 4 * iou, \quad b = 2 * iou * (w + h), \quad c = (iou - 1) * w * h \quad (16)$$

$$r \leq \frac{-b + \sqrt{b^2 - 4 * a * c}}{2 * a} \quad (17)$$

高斯核的计算

$$\exp \frac{-(x * x + y * y)}{2 * \sigma * \sigma} \quad (18)$$