

dalle 目标检测 深度学习 (Deep Learning)

# 请问 faster RCNN 和 SSD 中为什么用smooth L1 loss，和L2有什么区别？

如题显示全部

提问时间：2017-4-8 16:10:21

关注问题 写回答 邀请回答 好问题 6 添加评论 分享

8 个回答 默认排序

杨指北 无业游民 + 关注

475 人赞同了该回答  
创建于：2019-03-13 06:51:06 修改于：2019-03-13 17:30:43

为了从两个方面限制梯度：

- 1. 当预测框与 ground truth 差别过大时，梯度值不至于过大；
- 2. 当预测框与 ground truth 差别很小时，梯度值足够小。

考察如下几种损失函数，其中  $x$  为预测框与 groud truth 之间 elementwise 的差异：

$$L_2(x) = x^2 \tag{1}$$

$$L_1(x) = |x| \tag{2}$$

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \tag{3}$$

损失函数对  $x$  的导数分别为：

$$\frac{dL_2(x)}{dx} = 2x \tag{4}$$

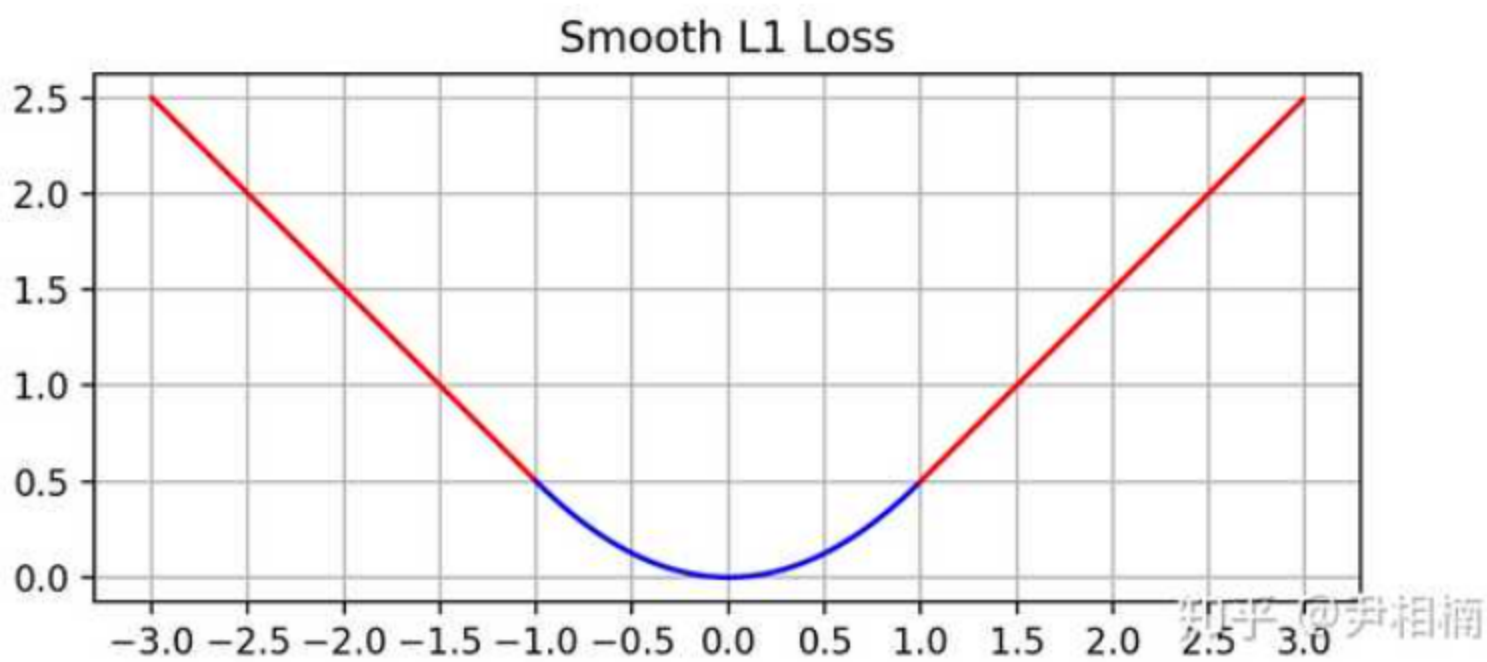
$$\frac{dL_1(x)}{dx} = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{otherwise} \end{cases} \tag{5}$$

$$\frac{d \text{smooth}_{L_1}}{dx} = \begin{cases} x & \text{if } |x| < 1 \\ \pm 1 & \text{otherwise} \end{cases} \tag{6}$$

观察 (4)，当  $x$  增大时  $L_2$  损失对  $x$  的导数也增大。这就导致训练初期，预测值与 ground truth 差异过于大时，损失函数对预测值的梯度十分大，训练不稳定。

根据方程 (5)， $L_1$  对  $x$  的导数为常数。这就导致训练后期，预测值与 ground truth 差异很小时， $L_1$  损失对预测值的导数的绝对值仍然为 1，而 learning rate 如果不变，损失函数将在稳定值附近波动，难以继续收敛以达到更高精度。

最后观察 (6)， $\text{smooth}_{L_1}$  在  $x$  较小时，对  $x$  的梯度也会变小，而在  $x$  很大时，对  $x$  的梯度的绝对值达到上限 1，也不会太大以至于破坏网络参数。 $\text{smooth}_{L_1}$  完美地避开了  $L_1$  和  $L_2$  损失的缺陷。其函数图像如下：



由图中可以看出，它在远离坐标原点处，图像和  $L_1$  loss 很接近，而在坐标原点附近，转折十分平滑，不像  $L_1$  loss 有个尖角，因此叫做 smooth  $L_1$  loss。

编辑于 2019-03-13 17:30



璎珞  
NLPer

+ 关注

38 人赞同了该回答

创建于: 2017-11-22 17:36:30 修改于: 2017-11-22 17:36:31

搬运工，觉得有帮助：

当预测值与目标值相差很大时，梯度容易爆炸，因为梯度里包含了  $x-t$ 。所以rgb在Fast RCNN里提出了SmoothL1Loss。

$$L_{\text{loc}}(t^u, v) = \sum_{i \in \{x, y, w, h\}} \text{smooth}_{L_1}(t_i^u - v_i), \quad (2)$$

in which

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases} \quad (3)$$

当差值太大时, 原先L2梯度里的x-t被替换成了±1, 这样就避免了梯度爆炸, 也就是它更加健壮.

原文链接: [Single Bounding Box Regression](#)

编辑于 2017-11-22 17:36

▲ 赞同 38 ▼

● 5 条评论

✈ 分享

★ 收藏

♥ 喜欢

...