

OHEM 交叉熵损失

范仁义

230725

一、课件及资料

► 代码、课件及各种资料位置：

fry404006308/fry_course_materials: 范仁义录播课资料,
https://github.com/fry404006308/fry_course_materials

二、OHEM 全称及原理

► 原理

挖掘困难样本，
忽略简单样本

► 也就是说，只有困难样品才参与损失计算

 GPT-4 >

Context cleared

OhemCELoss 的全称（包括中文全称）及对应的原理

OhemCELoss 是一个损失函数，全称是 Online Hard Example Mining Cross-Entropy Loss（在线难样本挖掘交叉熵损失）。

英文全称: Online Hard Example Mining Cross-Entropy Loss

中文全称: 在线难样本挖掘交叉熵损失

这个损失函数结合了交叉熵损失（Cross-Entropy Loss）和在线难样本挖掘（Online Hard Example Mining, OHEM）的概念。

交叉熵损失（Cross-Entropy Loss）：这是一种常见的损失函数，用于衡量模型的预测概率分布与真实分布之间的差异。在分类问题中，它常常被用来优化模型的性能。

...

在线难样本挖掘（Online Hard Example Mining, OHEM）：这是一种训练策略，主要用于解决类别不平衡的问题。在训练深度学习模型时，难样本（即模型预测错误的样本）通常包含更多的信息，可以帮助模型学习到更强的特征。因此，OHEM 的策略是在训练过程中选择一部分难样本来进行学习，忽略一些容易的样本，从而更好地优化模型。

具体来说，OhemCELoss 在计算损失时，会先按照预测结果的损失值排序，然后选择损失最大的一部分（即最难的样本）来计算最终的损失。这样可以让模型在训练过程中更加关注难分类的样本，从而提高模型的性能。

三、来源

- Focal-Loss那篇论文里面提到
- 分割网络BiSeNetV2中用到

Online Hard Example Mining (OHEM): [31] proposed to improve training of two-stage detectors by constructing minibatches using high-loss examples. Specifically, in OHEM each example is scored by its loss, non-maximum suppression (nms) is then applied, and a minibatch is constructed with the highest-loss examples. The nms threshold and batch size are tunable parameters. Like the focal loss, OHEM puts more emphasis on misclassified examples, but unlike FL, OHEM completely discards easy examples. We also implement a variant of OHEM used in SSD [22]: after applying nms to all examples, the minibatch is constructed to enforce a 1:3 ratio between positives and negatives to help ensure each minibatch has enough positives.

We test both OHEM variants in our setting of one-stage detection which has large class imbalance. Results for the original OHEM strategy and the ‘OHEM 1:3’ strategy for selected batch sizes and nms thresholds are shown in Table 1d. These results use ResNet-101, our baseline trained with FL achieves 36.0 AP for this setting. In contrast, the best setting for OHEM (no 1:3 ratio, batch size 128, nms of .5) achieves 32.8 AP. This is a gap of 3.2 AP, showing FL is more effective than OHEM for training dense detectors. We note that we tried other parameter setting and variants for OHEM but did not achieve better results.

四、OHEM步骤流程

- ▶ 假设我取一个预测值概率标准0.7,
- ▶ 如果预测值大于0.7的可以看成简单样本, 不参与损失计算,
- ▶ 如果预测值小于0.7的, 看成困难样本, 参与损失计算
- ▶ 当点不够的时候, 最少保证取 n_{\min} (比如1/16) 个点

五、代码

范头