

《An Efficient Approach to Escalate the Speed of Training Convolution Neural Networks》阅读笔记

zxp

May 4, 2024

1 论文内容

《一种提高卷积神经网络训练速度的嵌入式设备方法》这篇论文和我们做的东西关系不大，论文指出现在模型越来越大，从最早的8层已经进化到有100多层，嵌入式等计算能力小的设备越来越没法吃到深度学习的红利，因此需要优化模型

1.1 论文中提到的背景

论文指出训练时间依赖数据集大小，有种优化方向是整理数据集将一些不重要的数据去除，训练的数据集少了训练时间少了，精简数据集增加效率。另一种是裁剪模型，将过滤器变小，这样计算少了，模型参数也少了需要的训练时间也少了。模型中有部分参数对精度的影响是十分有限的，即使减少这些参数模型的精度并不会变小很多，现在很多减小模型就是裁剪模型减少参数，有各种各样的方式去压缩模型。

1.2 论文考虑的方式

论文两种方式都考虑，先将数据集中接近的数据聚合，删除接近的数据。然后使用软裁剪的方式一次裁剪所有层，之所以是软裁剪和硬裁剪不同的是将逻辑上删了过滤器，逻辑上不管，总模型大小不变。

2 心得

与我们的关系不大，相同的是背景，嵌入式设备计算能力不够，不过这篇论文

是对模型层面进行改变。