

BN LN IN GN

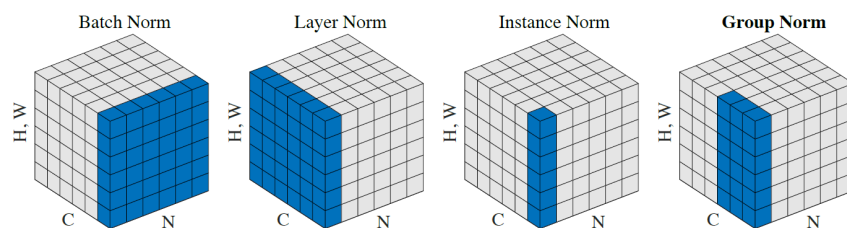


Figure 2. Normalization methods. Each subplot shows a feature map tensor. The pixels in blue are normalized by the same mean and variance, computed by aggregating the values of these pixels. Group Norm is illustrated using a group number of 2.

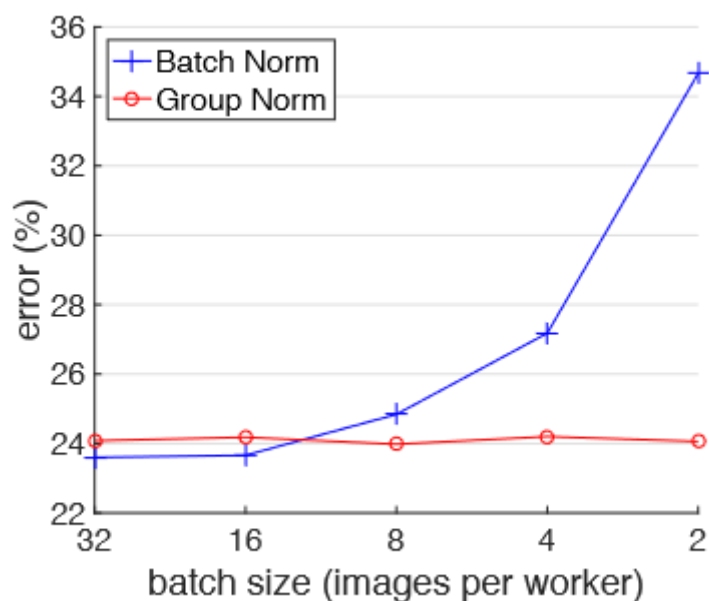
batch norm

最后的结果可以看成在C个(N, W, H) slices上求normalize，比如说有一个(8, 32, 32, 256)的feature输出，会在(8, 32, 32, 1)形状在每个feature点上求对应normalize

layer normalzie

简介

主要是因为当bs很小的时候BN效果会很差，尤其是一些检测和分割的任务， $bs > 32$ 一般比较有效。



可以发现GN的测试集错误率非常稳定。

GN做的事情主要是将feature map channel按照一定方式分组，然后按照组进行feature normalize。GN的计算和batch size无关。

GN在小batch效果非常好，常规batch也仅比BN差0.5%左右，而且GN可以自然适应pre-train的finetune。

还有一些其他的normalize方式比如LN和IN，但这些方法都不适合visual recognition