|  |  |
| --- | --- |
| BN层合并原理 | 文件类型：技术文档 |
| 文件编号： |
| 面向的部门： 研发中心 |
| 保密等级：高 |
| 作者： 刘凯 |
| 日期：2019-2-22 |
| 版本：1.0V |

BN层合并原理



目 录

[卷积层和BN层合并 1](#_Toc1737488)

[FC层和BN层合并 1](#_Toc1737489)

在训练深度网络模型时，BN（Batch Normalization）层能够加速网络收敛，并且能够控制过拟合，一般放在卷积层之后。BN 层将数据归一化后，能够有效解决梯度消失与梯度爆炸问题。虽然 BN 层在训练时起到了积极作用，然而，在网络前向推断时多了一些层的运算，影响了模型的性能，且占用了更多的内存或者显存空间。目前，很多先进的网络模型（ResNet，MobileNet，Xception，ShuffleNet 等）都使用了BN技术，因此，我们有必要将 BN 层的参数合并到卷积层，来提升模型前向推断的速度。

## 卷积层和BN层合并

因为卷积层(conv层)有padding的存在，BN层在conv层之前的网络无法合并，但BN层在conv层后的网络可以合并。

卷积层的输入为*X*，权重为*W*，偏置为*B*，输出：



BN层的均值为，方差为，缩放因子和偏置分别为，则输出为：



其中，为一个小量，在InsightFace中。因此，conv层和BN层合并后的系数为：



其中：



## FC层和BN层合并

与conv层不同，FC层可以与前置和后置的BN层合并。

FC层的权重和偏置为W和b，输入为X，则输出为：



设FC前的BN层参数为，后置的BN层参数为，则与前置BN合并的参数为：



与后置的BN合并的参数为：

