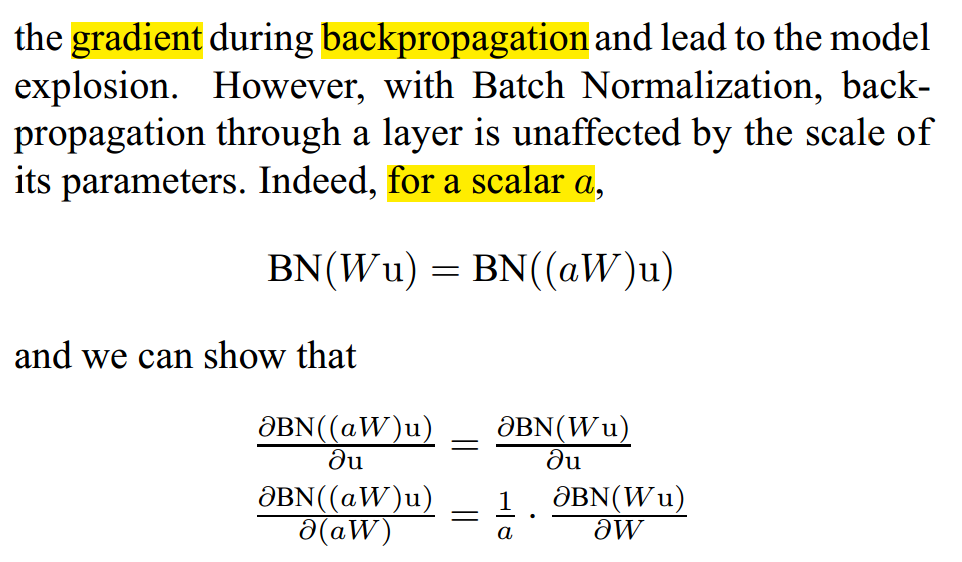


在卷积神经网络上用bn 层的时候，每一张特征图采用一对 γ β 参数，也就是一个 filter对应一对，因为卷积层的特性是权值共享

Bn 层对 W 的 scale 影响不明显， 原因是 W 的 scale只影响损失函数对当前权重的梯度，不会影响反向传播（即之后的梯度），同时大的 alpha会导致小的梯度



论文推测 bn层最终会让 该层的雅可比矩阵值趋向 1（梯度为1 适合训练）