**梯度弥散详解**

什么是梯度弥散？

答：每一个节点大致都是矩阵乘的方式训练参数，一个block经过多个节点，层层嵌套，梯度流过一个blcok之后，就已经下降了几个量级，传到前一层的梯度将会变得很小，这就是梯度弥散。

从公式中分析：①链式法则的累乘效应 ②极小梯度在传递链上的连续出现；

经过的层次越多，梯度弥散越严重，导致浅层部分的网络权重参数得不到很好的训练。

解决梯度弥散问题——残差结构

## 答：resNet残差在block外加上“捷径”，称“skip connection”，h(x)+x再求梯度时无论h(x)层数有多深，都会有“1”,正是由于多了这条捷径，来自深层的梯度能直接畅通无阻地通过，去到上一层，使得浅层的网络层参数等到有效的训练。DenseNet不仅实现相邻层的残差，也可实现跨层级的残差。

总而言之，是在后传过程中，skip connection则能在后传过程中更好地把梯度传到更浅的层次

解决梯度弥散问题——BN

答：BN通过规范化输入数据改变数据分布，在前传过程中消除梯度弥散。即规定好分布范围，避免输入block后，输出分布过小，影响梯度求解结果。