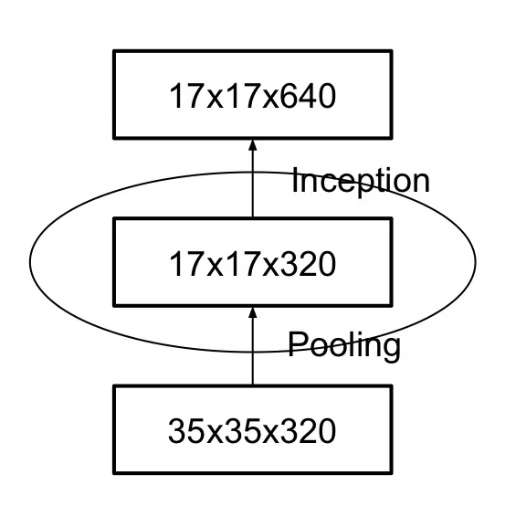
# 背景

提出了四个通用的网络设计准则和优化思路：

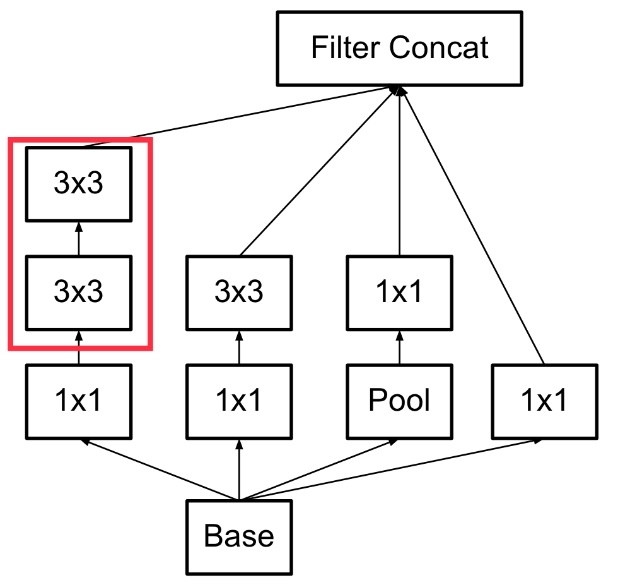
①从输入到输出，特征图的尺寸应该是逐渐变小的，不然会引起特征表达瓶颈；



如上图所示的结构就背离了第一原则。

②高维信息更容易被网络处理。增加卷积后的激活函数的次数，得到高维稀疏特征；

③可以先将具有大感受野的将卷积映射到低维空间，不会引起很多，甚至不会引起任何在特征表达能力上的损失；



如上图利用两个3×3卷积代替一个5×5的卷积。

④均衡考虑网络的深度和宽度。

# 方法

①分解大卷积核

邻近的激活区域是高度关联的，意味着输出的特征图里有冗余的部分，因此考虑将大卷积特征聚合之前，将其降维处理，且保留非线性激活函数，增加非线性的同时也提高了特征表达能力。

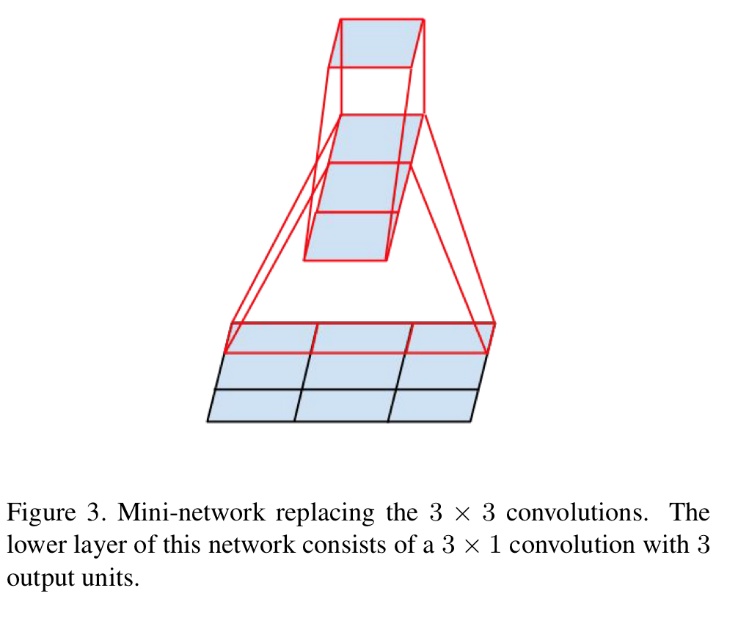
将大卷积分解减少了计算量。假设输入特征图尺寸为(C,H,W)，卷积个数都为c个经过一个5×5的卷积，计算量为：

H×W×C×5×5×c=25HWCc

经过两个3×3卷积，计算量为：

(H×W×C×3×3×c)×2=18HWCc

②非对称分解



如上图所示，将一个3×3卷积核分解为一个3×1卷积和一个1×3卷积。

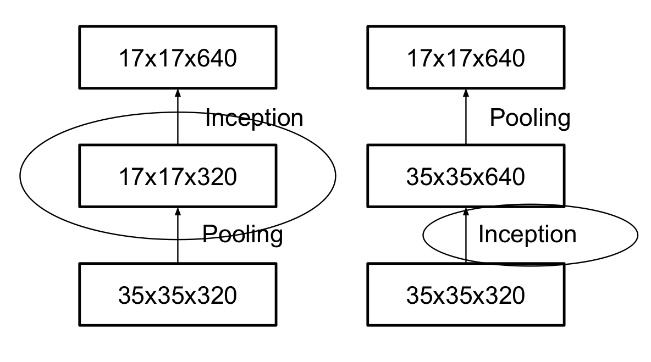
更通俗地来讲可以将一个n×n的卷积分解为一个n×1卷积和一个1×n卷积。在特征尺寸介于12~20之间使用这种卷积分解效果最好。

减少了计算量而且n×1卷积和一个1×n卷积之间的非线性激活函数增加了网络的表征能力。

③辅助分类器

在Inception\_v1里提到的辅助分类器可以使模型收敛地更快，其实通过实验发现用不用辅助分类器差别不大，只是使用辅助分类器的精度会相对高一些。而且去掉浅层的辅助分类器也不会有什么副作用。如果辅助分类器中使用了BN或者dropout等正则化方法，得到的效果更好。

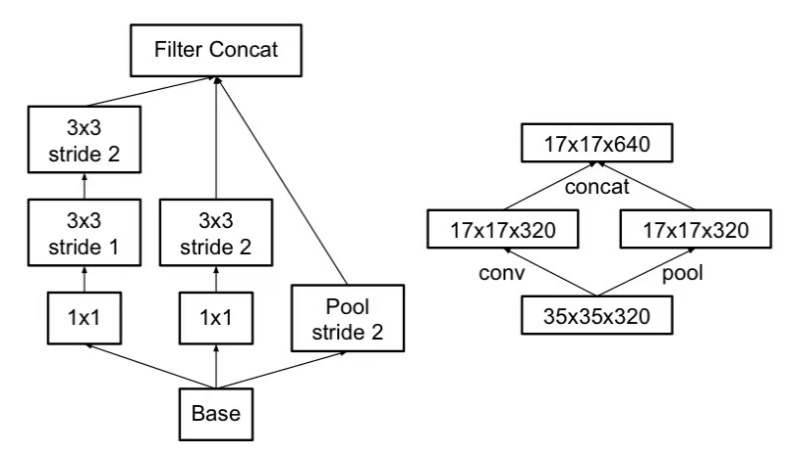
④高效降低尺寸（使特征图变小变厚）



上图中，左图利用步长为2的卷积下采样，造成信息丢失，带来特征表达瓶颈问题，而右图存在的问题是计算量大。

假设输入特征图尺寸为，输出的特征图尺寸为，如果先升维(卷积核为n，卷积核个数为2k)再池化，其计算量有：





采用并行的方式解决上述问题，左边两路卷积操作，右边一路池化操作，即避免了特征瓶颈又减少了计算量。

⑤标签平滑操作

通常的分类任务是将图像的真实标签值转换成one-hot编码，然后根据softmax值计算交叉熵，优化任务就是使得交叉熵函数最小。这样存在一个问题，就是当softmax得到的置信度越大时，交叉熵越小，因此最小交叉熵会使得正确类别的分数趋于无穷，使得模型对于预测结果过于自信，泛化能力差。

因此引入了标签平滑参数以及标签先验分布来解决这些问题。

原先的标签函数分布：

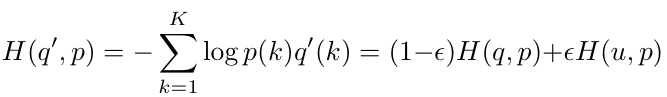


可以写作

标签平滑后变为：



交叉熵可以表示为：



为预测值与真实标签值的交叉熵，为预测值与先验分布的交叉熵。相当于在真是标签上增加噪声，让预测值不要过度集中于概率较高的类别

Inception\_v2指的是使用了新提出的技术中的一种或多种的Inception模块。而Inception\_v3指的是这些技术全用了的Inception模块

# 总结

相比于Inception\_v1的结构针对于不同深度设计了不同的结构，不是像v1简单地重复相同的模块结构。也解决了v1里存在的特征瓶颈问题，以更柔和的方式让特征图的变小变厚。

Inception结构感觉就是在保证表达能力的同时减少参数量。将原本的乘积操作以并行的方式分散为加和操作，以后自己在设计网络时，如果参数量大的话，可以考虑利用这种并行的方式减少计算量。

# 补充

🖊神经网络是将像素空间映射到特征空间，增加激活函数的次数可以增加这种非线性映射能力，使得网络更容易解析特征，加快网络的训练。