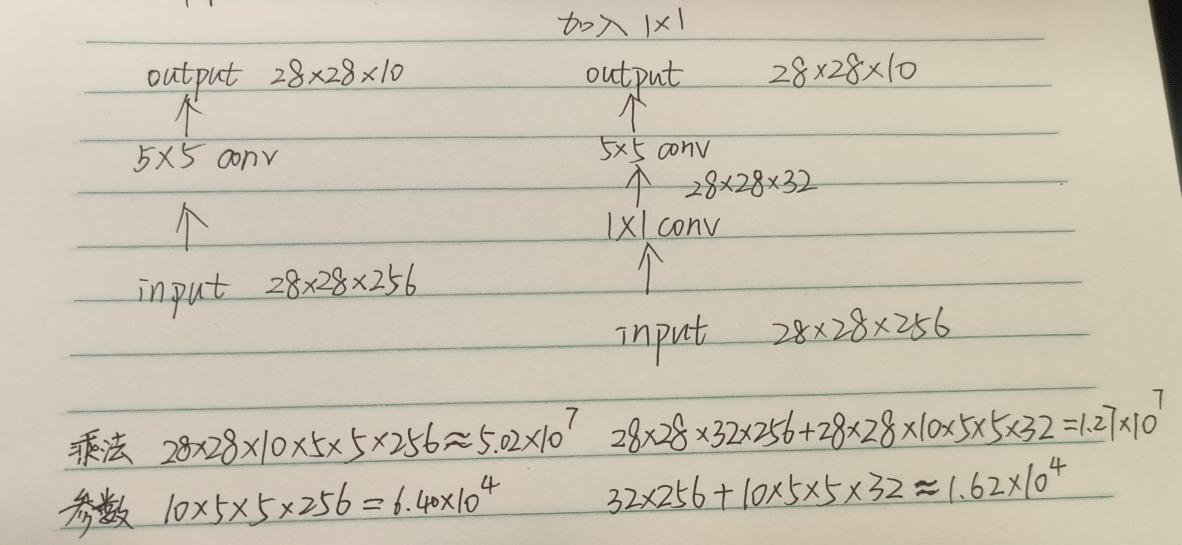
A6.1(5分)分析卷积神经网络中用1x1的卷积核的作用：

使用1×1卷积能够缩减维度(减小channel)，形成“瓶颈层”，起到减少参数、降低计算量的作用。

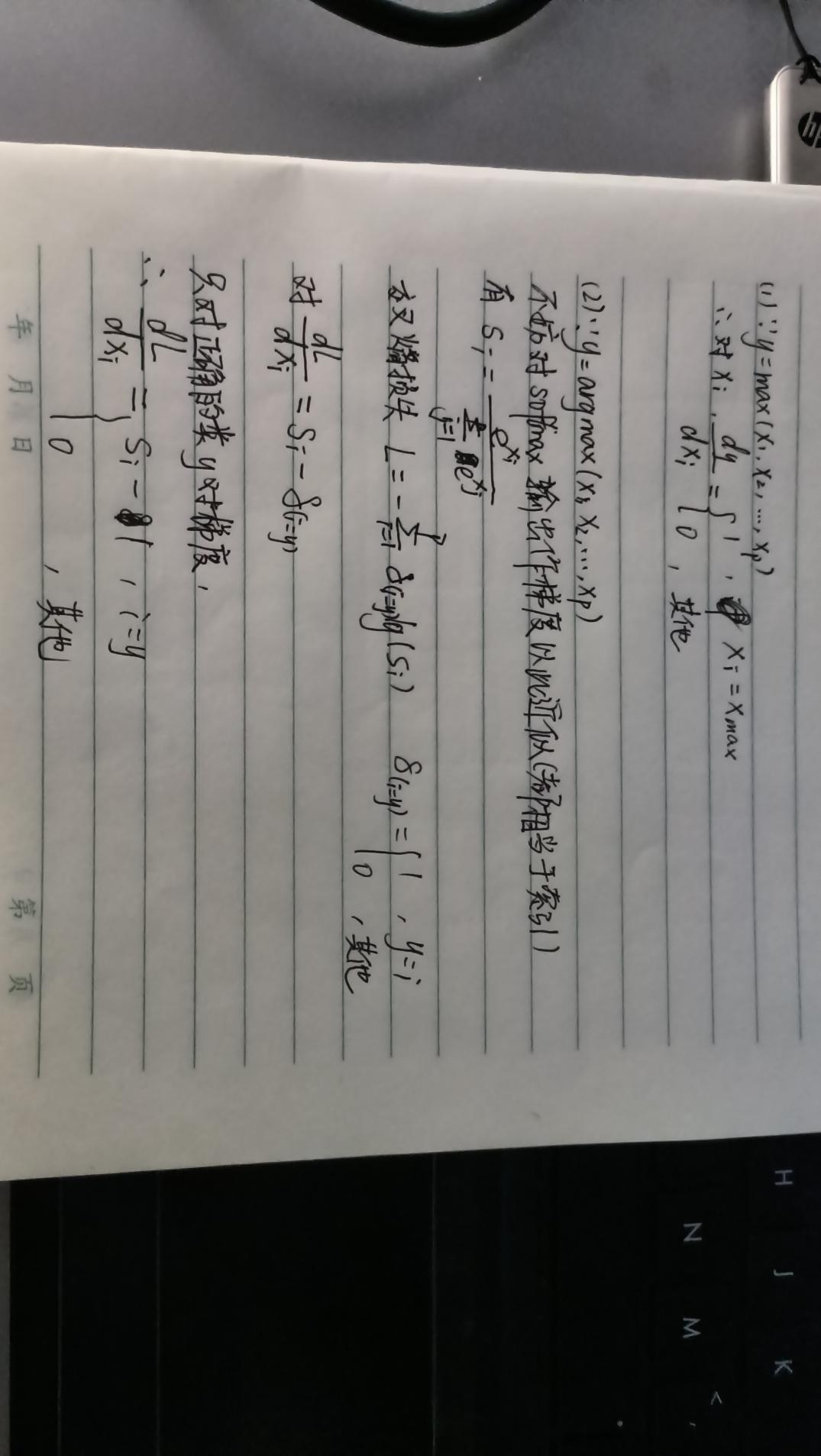
以下图为例，使用1×1卷积后，乘法次数与参数数量均减小了数倍。



此外，在1×1卷积后加入ReLu增添非线性，可以起到提高泛化能力的作用。

总而言之，这种先降维度后升维度的瓶颈型结构能够跨通道（跨不同特征图信息）聚合，进一步可以起到降维（或者升维）的作用，减少参数，即利用深层次提高抽象能力，同时节约计算量和参数的数量。

A6.2(5分)计算函数y = max(x1,…xp)和函数y = argmax(x1,…,xp)的梯度。



A6.3(5分)推导LSTM网络中参数的梯度，并分析其避免梯度消失的效果。

1. 参数的梯度

①在前向传播中，通过LSTM的门和记忆单元的运算，计算出输出和神经元状态。

②在反向传播中，计算损失相对于参数的梯度，使用链式法则，将梯度从损失传播到各个门和记忆单元。

具体的各个门的计算公式在A4中已经给出：

输入门：f(i)=

输出门：f(o)=

遗忘门：f(t)=

1. 避免梯度消失

①遗忘门决定了神经元状态中哪些信息需要被遗忘，有助于保留长期依赖关系，避免过早地忘记重要信息。

②输入门控制着新输入信息进入神经元状态的程度，有助于防止梯度消失，允许网络在长序列中选择性地更新记忆，有助于维持梯度的有效传播。

③神经元状态的设计允许信息直接流动，避免梯度在反向传播过程中消失的问题。

A6.4(5分)当将自注意力模型作为神经网络的一层使用时，分析它和卷积层以及循环层在建模长距离依赖关系的效率和计算复杂度方面的差异。

（1）建模长距离依赖关系：

①卷积层：通过局部感受野和权值共享的方式捕捉局部特征，在处理长距离依赖关系时信息传播可能受限，效率也随之下降。

②循环层：通过时间递归操作捕捉长距离依赖关系，容易面临梯度消失或梯度爆炸问题，长序列的处理效率可能低下。

③自注意力模型：直接在所有位置上建立全局关联，能够更好地捕捉长距离依赖关系；通过计算不同位置之间的注意力权重，更灵活地调整不同位置的关注程度，对输入序列的不同部分进行加权。相对来讲效率较高。

（2）计算复杂度：

①卷积层：卷积层的计算复杂度主要取决于卷积核的大小、通道数和层的深度，一般来讲计算复杂度较小。

②循环层：循环层的计算复杂度主要与序列长度和隐藏状态的维度有关，处理长序列时的递归计算可能会导致较高的计算成本。

③自注意力模型：自注意力模型的计算复杂度与序列长度和特征维度有关，但相比循环层，自注意力模型可以并行计算不同位置之间的关系，因此在某些情况下可能更高效。

综上，卷积层参数共享降低计算成本和复杂度，在处理长距离依赖关系时可能受限，效率较低；循环层建模长距离依赖关系的计算复杂度较高且效率低下；自注意力模型在捕捉长距离依赖关系上表现较好，计算复杂度高但效率也相对较高。