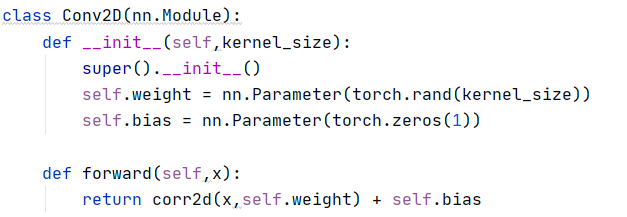
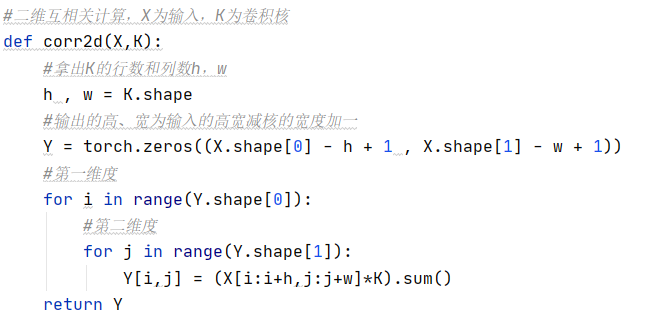
**卷积层：两个性质**

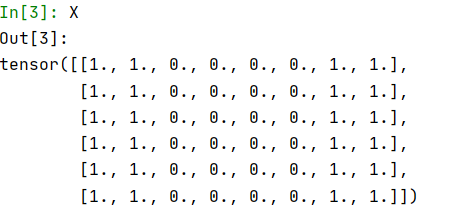
1.平移不变性（translation invariance）：不管检测对象出现在图像中的哪个位置，神经网络的前面几层应该对相同的图像区域具有相似的反应，即为“平移不变性”。

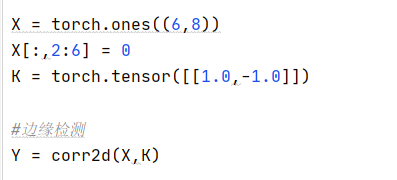
2.局部性（locality）：神经网络的前面几层应该只探索输入图像中的局部区域，而不过度在意图像中相隔较远区域的关系，这就是“局部性”原则。最终，可以聚合这些局部特征，以在整个图像级别进行预测。

总结：卷积层就是全连接层加上

代码：

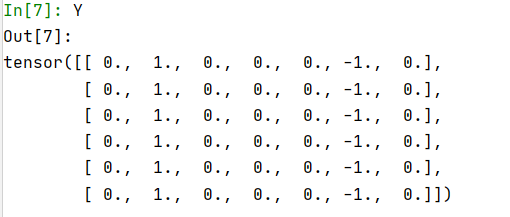
输出的高、宽为输入的高宽减核的宽度加一（步幅为1）

这个K可以用于边缘检测：例如如下矩阵



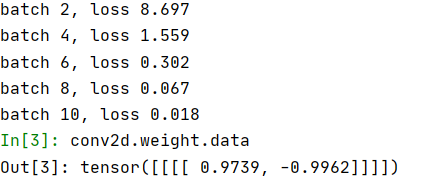


结果：从1到0的边缘为1，从0到1边缘显示为-1

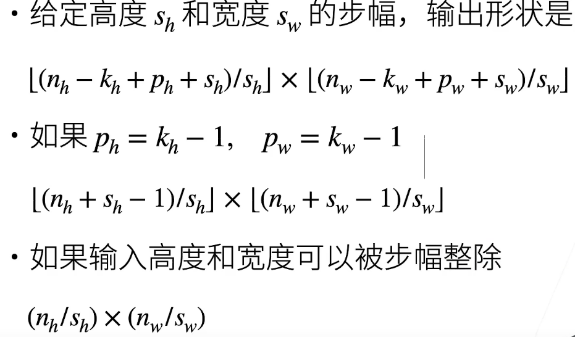


学习上述卷积核：

结果如下



故可以通过人造数据集学习卷积核的大小

n为输入大小，k为卷积核大小，p为padding大小

**卷积层里多输入多输出通道：**

输出通道数是卷积层的超参数

每个输入通道有独立的二维卷积核，所有通道结果相加得到一个输出通道结果

每个输出通道有独立的三维卷积核

**池化层：**

平均池化和最大池化

池化层返回窗口中最大值或平均值

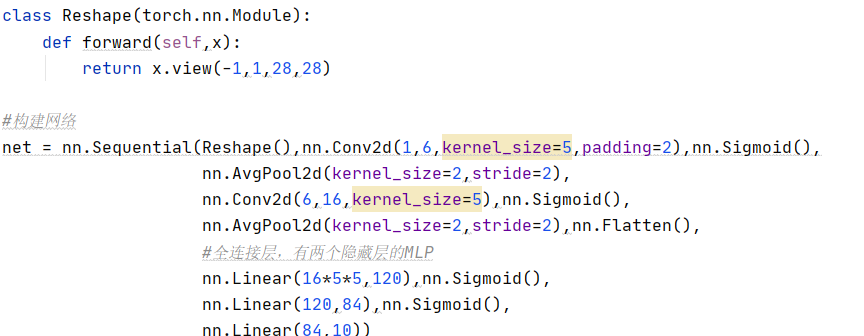
缓解卷积层位置敏感性

同样有窗口大小、填充和步幅作为超参数

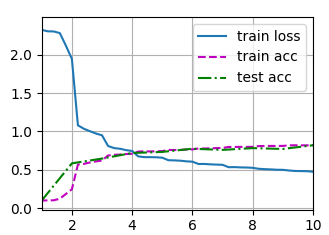
**LeNet实现：**

构建模型：

第一层为卷积层，1个通道输入，6个通道输出，卷积核为5\*5，sigmoid作激活函数，其次是2\*2平均池化层，然后又是一层卷积层，6通道输入，16通道输出，一层2\*2平均池化层，然后是全连接层，有两个隐藏层的MLP，10个分类输出



使用Fashion-MNIST数据集，使用gpu运行此次计算，优化器为SGD，学习率为0.9，结果如下



训练10轮

可以看出和MLP差距不大，但是LeNet几乎

没有过拟合，MLP是会有一点

AlexNet实现

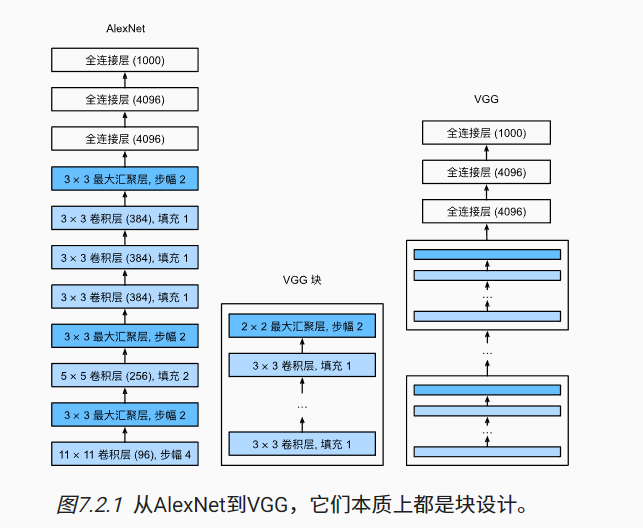
AlexNet是更深更大的LeNet

它使用ImageNet数据集，输出为1000类

但是这里使用Fashion-MNIST数据集，输出为10类

训练学习率为0.01，训练10轮，结果如下

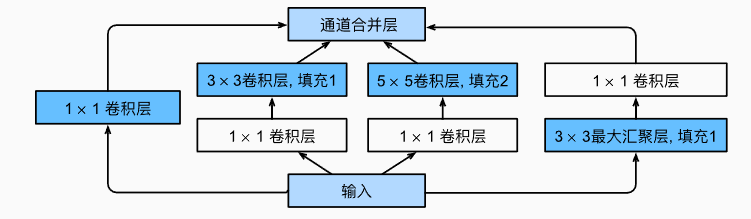
比LeNet表现好很多

**VGG块：**

模型比AlexNet更复杂，但是性能更好一点

GoogLeNet：

关键：Inception块



Batch\_Norm：批量归一化

1.在模型训练过程中，批量归一化利用小批量的均值和标准差，学习出适合的偏移和缩放，不断调整神经网络的中间输出，使整个神经网络各层的中间输出值更加稳定（对于全连接层，作用在每一个特征维（每一列）上；对于卷积层，拉成一个矩阵之后，作用在输出维）

2.可以加速收敛速度（使用了批量归一化之后，学习率可以调的比较大，用更大的学习率训练可以提高训练速度，不会出现学习率过大时，上层梯度爆炸，梯度较小时，下层梯度消失的情况。将每一层的输入变成差不多的分布之后，就可以使用大一点的学习率），但一般不改变模型精度

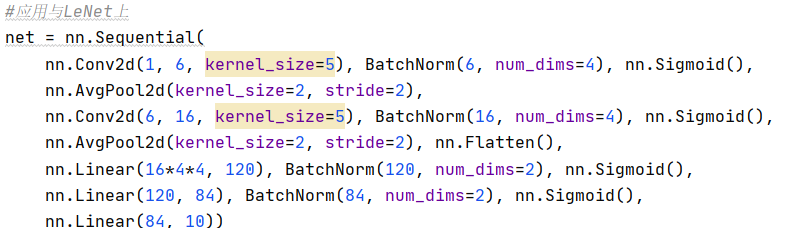
3.批量归一化在全连接层和卷积层的使用略有不同

4.批量归一化层和暂退层一样，在训练模式和预测模式下计算不同

5.批量归一化有许多有益的副作用，主要是正则化。另外一方面，原论文中关于批量归一化“减少内部协变量偏移”的原始动机似乎不是一个有效的解释

代码实现：

在LeNet上应用batch\_norm：



在卷积层后，全连接层后使用BN

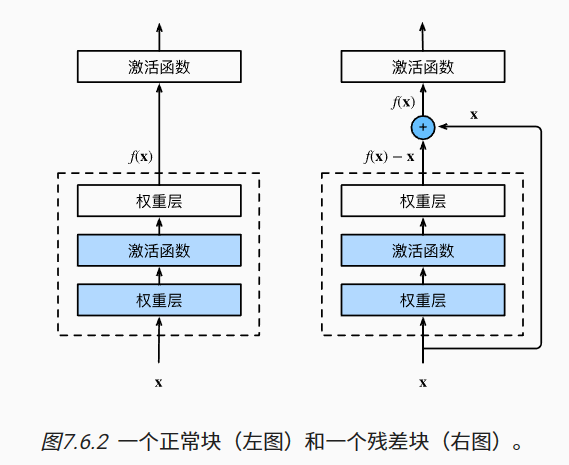
结果：收敛的更快了，lr为1.0，训练10轮，batch\_size 为256



这个是没有用batch\_size的：



ResNet：

残差网络的核心思想是：每个附加层都应该更容易地包含原始函数作为其元素之一

1.之前增加模型深度的方法都是层层堆叠的方法，ResNet的思想是在堆叠层数的同时不会增加模型的复杂度

2.上图中左侧表示一个正常块，右侧表示一个残差块

x：原始输入f(x)：理想映射（也是激活函数的输入）

3.对于正常块中来说，虚线框中的部分需要直接拟合出理想映射 f(x)；而对于残差块来说，同样的虚线框中的部分需要拟合出残差映射 f(x) - x

4.残差映射在现实中往往更容易优化

5、如果以恒等映射 f(x) = x 作为所想要学出的理想映射 f(x)，则只需要将残差块中虚线框内加权运算的权重和偏置参数设置为 0，f(x) 就变成恒等映射了

6.在实际中，当理想映射 f(x) 极接近于恒等映射时，残差映射易于捕捉恒等映射的细微波动

7.在残差块中，输入可以通过跨层数据线路更快地向前传播

总结：

1.残差块使得很深的网络更加容易训练（不管网络有多深，因为有跨层数据通路连接的存在，使得始终能够包含小的网络，因为跳转连接的存在，所以会先将下层的小型网络训练好再去训练更深层次的网络），甚至可以训练一千层的网络（只要内存足够，优化算法就能够实现）

2.学习嵌套函数是神经网络的理想情况，在深层神经网络中，学习另一层作为恒等映射比较容易

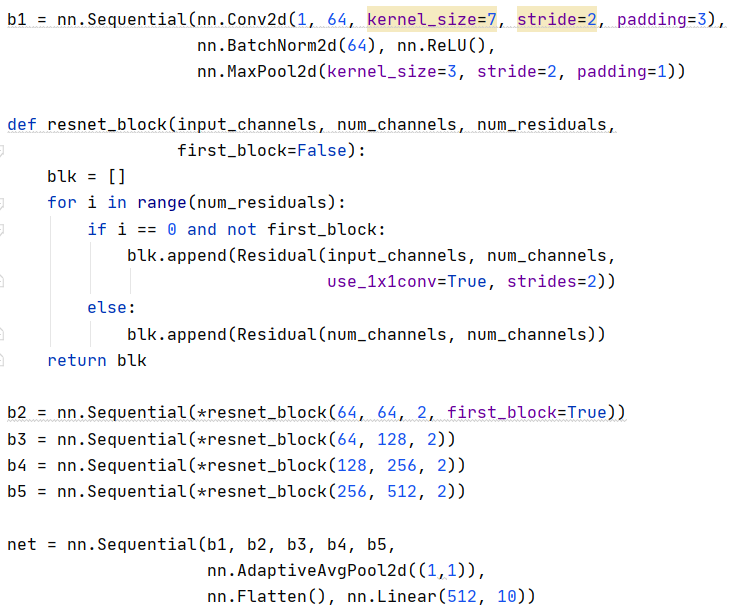
3.残差映射可以更容易地学习同一函数，例如将权重层中的参数近似为零

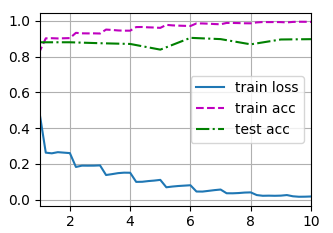
4.利用残差块可以训练出一个有效的深层神经网络：输入可以通过层间的残余连接更快地向前传播

5.残差网络对随后的深层神经网络的设计产生了深远影响，无论是卷积类网络还是全连接类网络，几乎现在所有的网络都会用到，因为只有这样才能够让网络搭建的更深

代码：







学习率为0.05，训练10轮，batchsize为256

结果为：



train acc很高