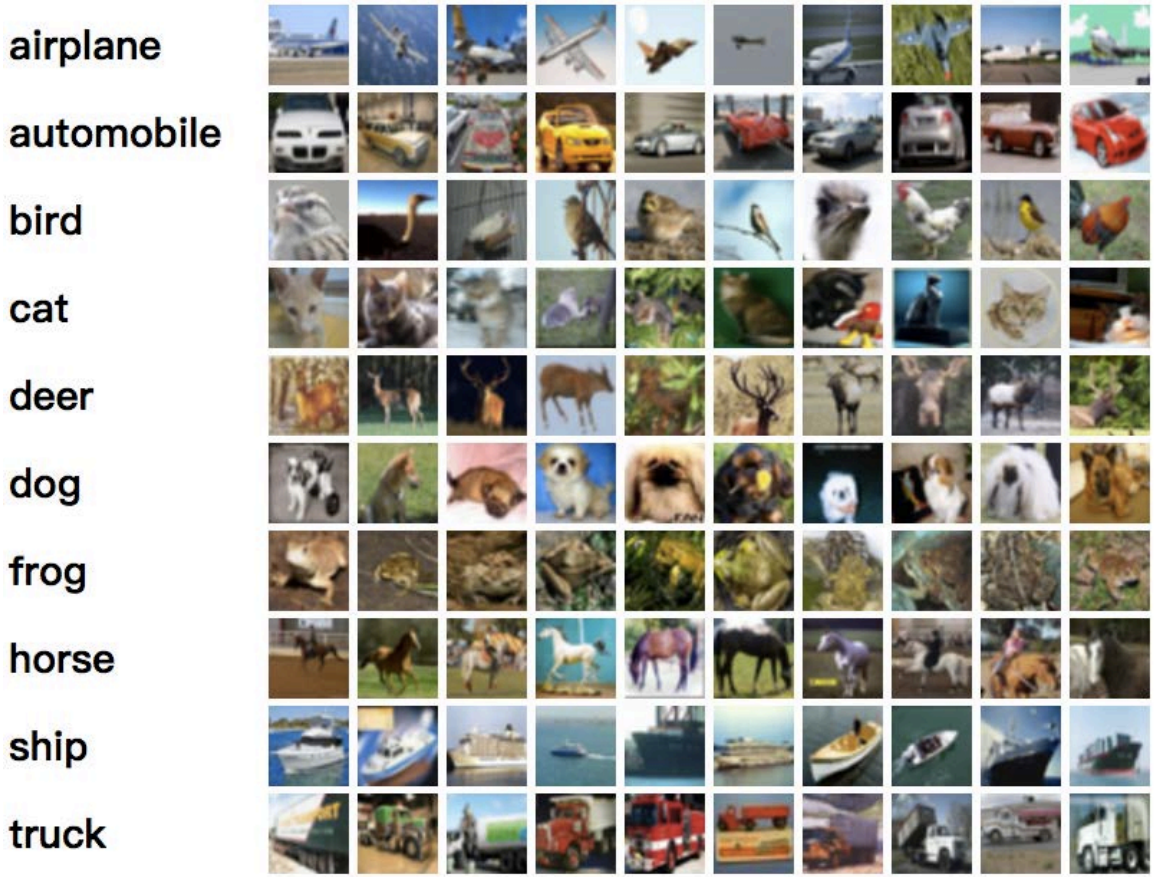
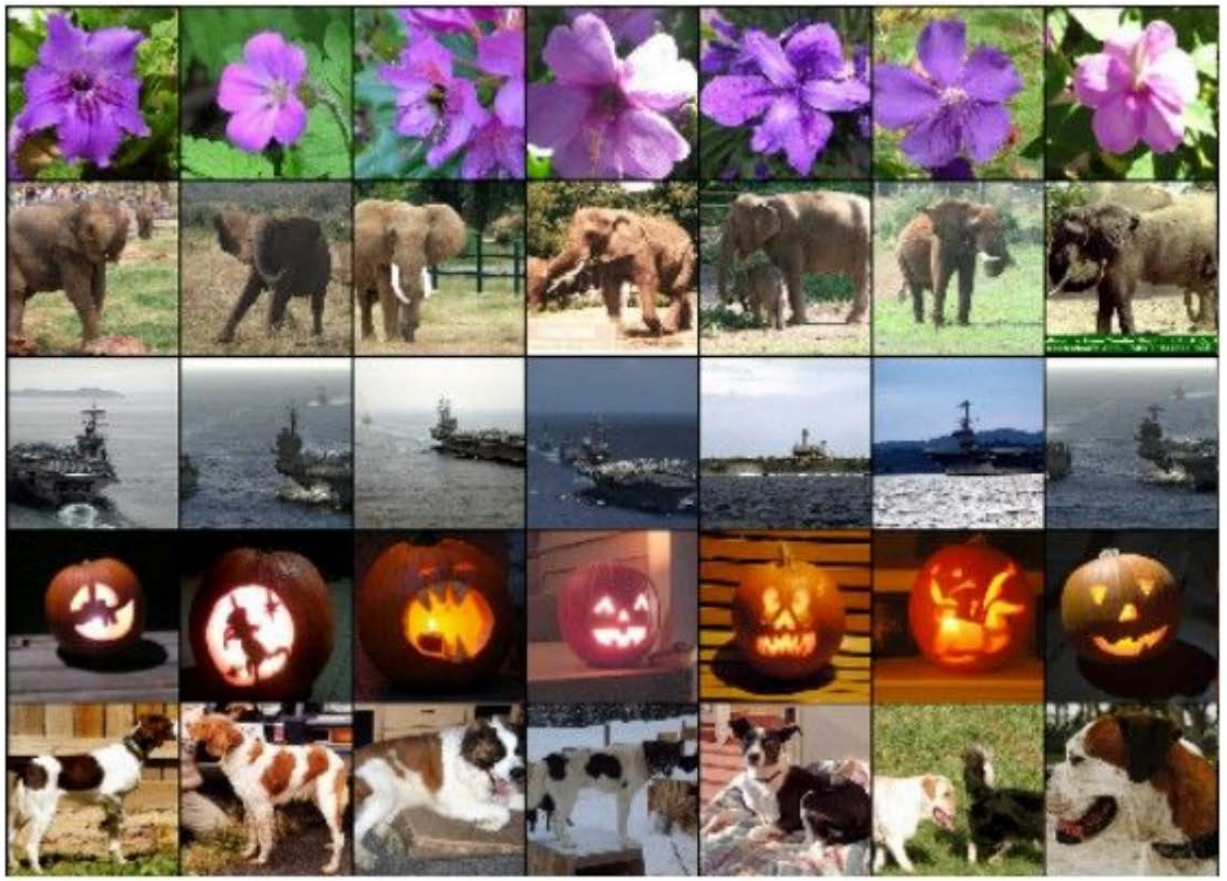
Lecture9
深度特征学习

无处不在的卷积神经网络

图像分类

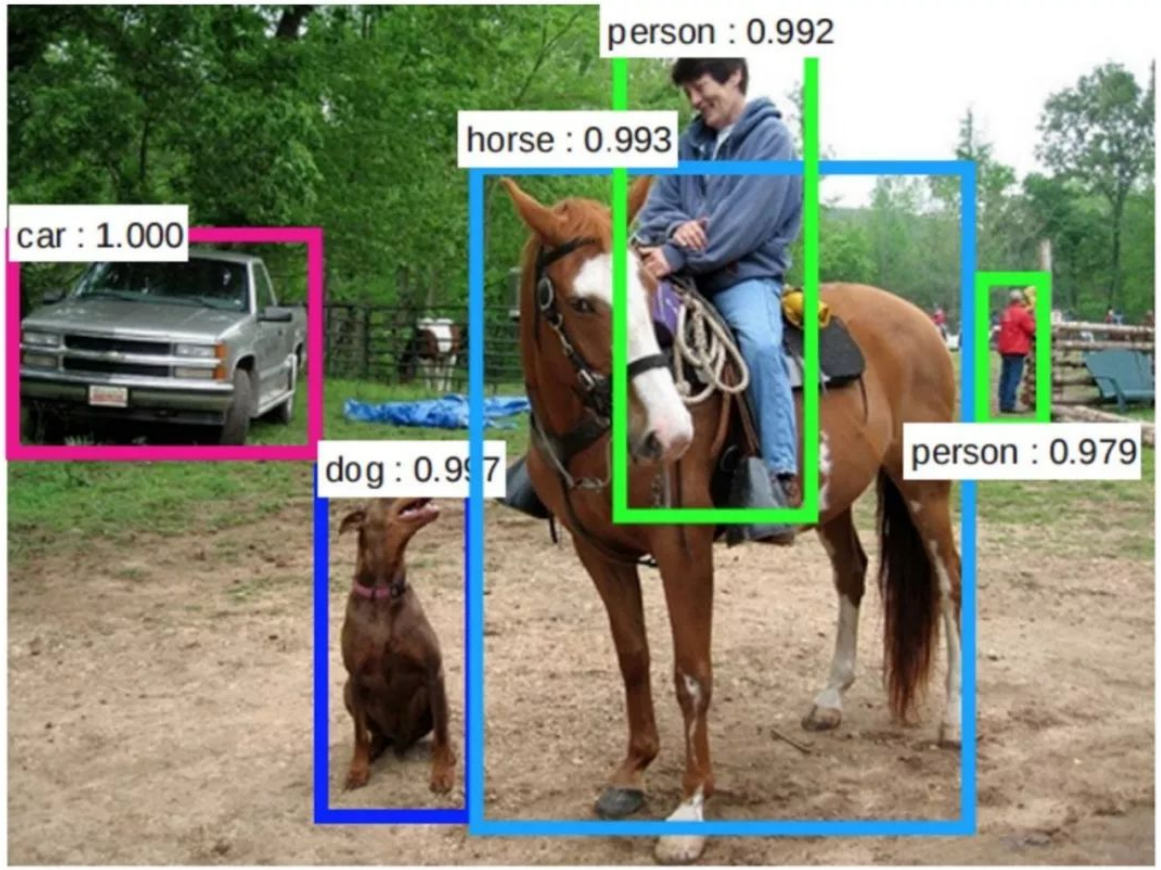


图像检索



无处不在的卷积神经网络

目标检测



语义分割

无处不在的卷积神经网络

A white teddy bear sitting in the grass

A man in a baseball uniform throwing a ball

A woman is holding a cat in her hand

图像描述

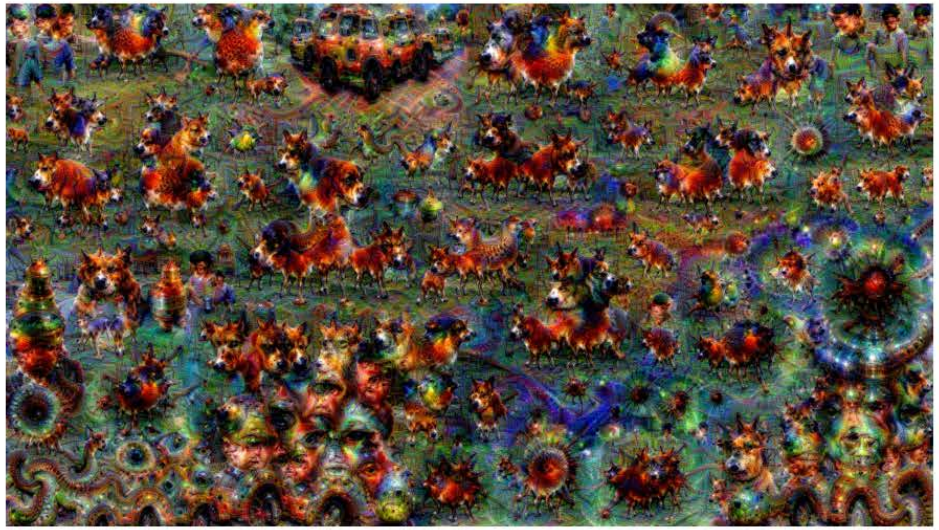
A man riding a wave on top of a surfboard

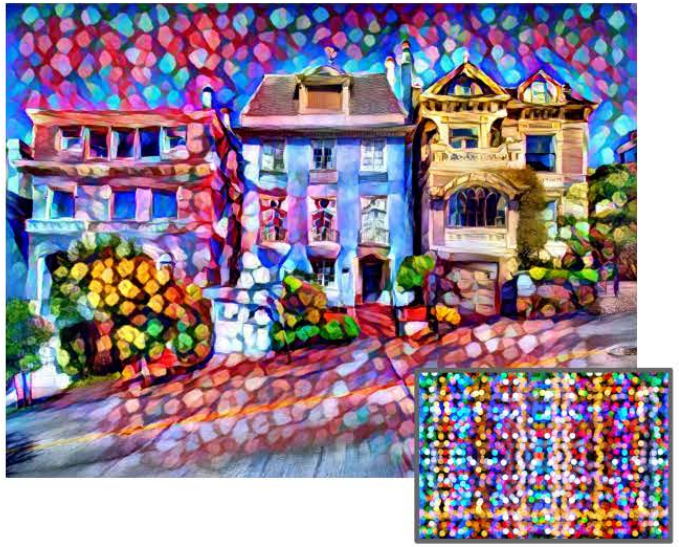
A cat sitting on a suitcase on the floor

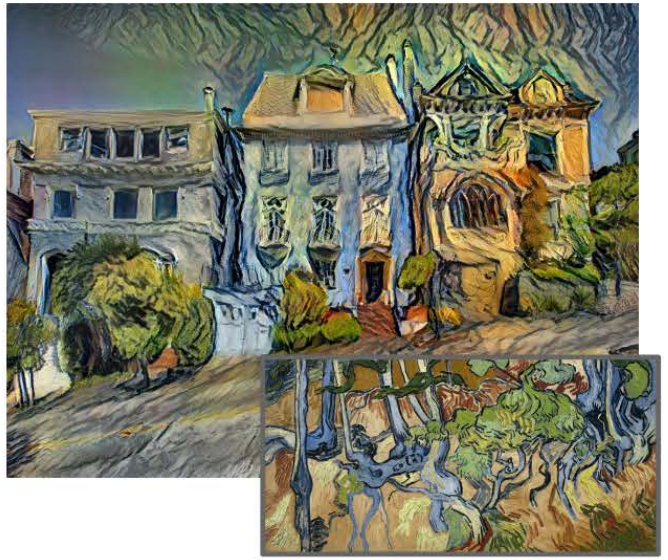
A woman standing on a beach holding a surfboard

无处不在的卷积神经网络

风格迁移



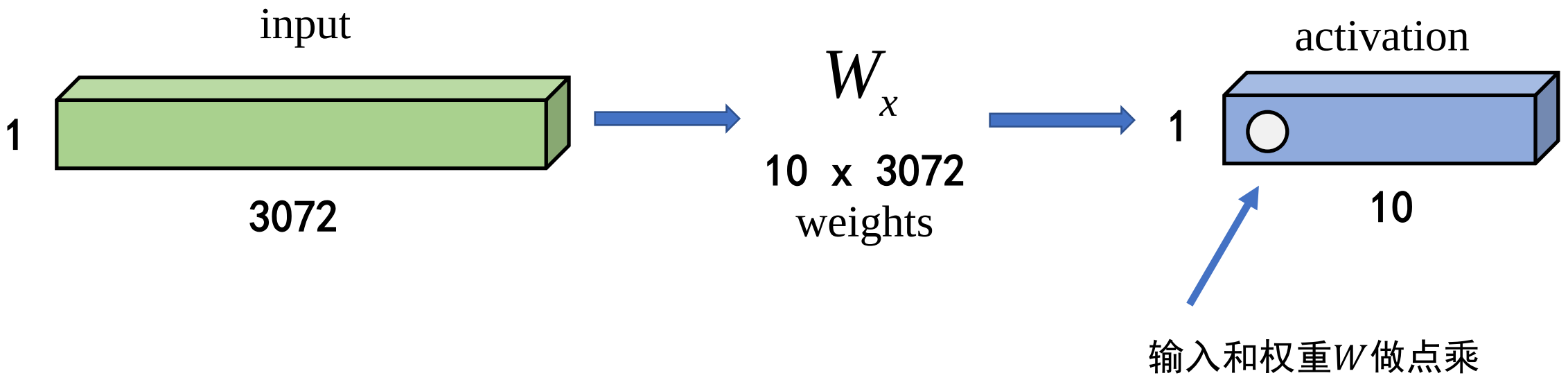




卷积神经网络
Convolutional Neural Networks

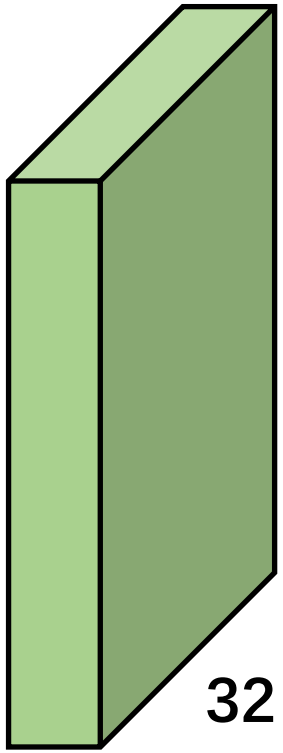
简单回顾:全连接层（Fully Connected Layer）

32x32x3 image → 拉伸为 3072x1



卷积层（Convolution Layer）

32x32x3 image → 保留空间信息

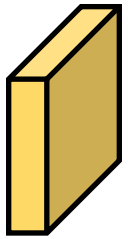


32 height

32 width

3 depth

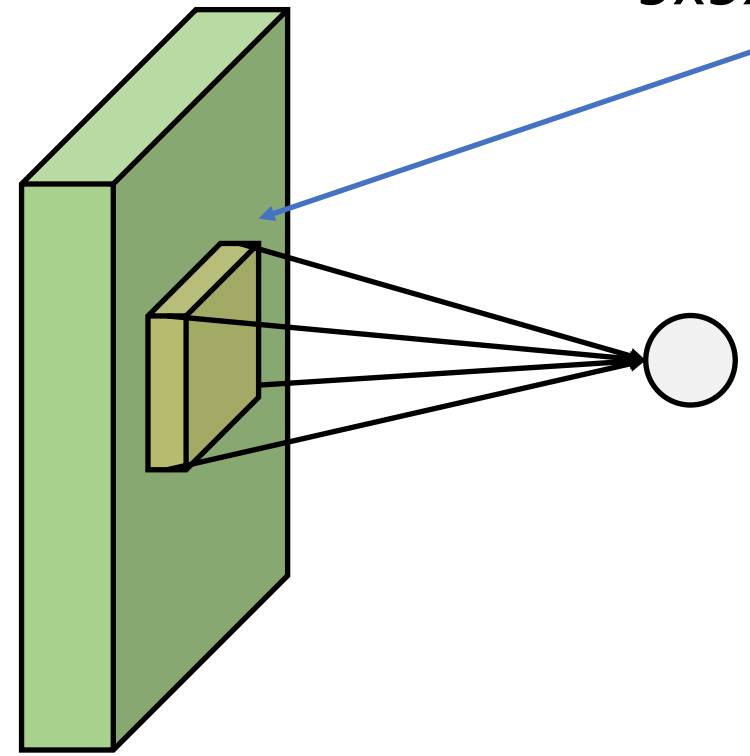
5x5x3 filter



用卷积核（filter）对原始图像进行卷积

卷积：在原始图像空间维度上进行滑动计算点积

卷积层（Convolution Layer）



32x32x3 image

5x5x3 filter ω

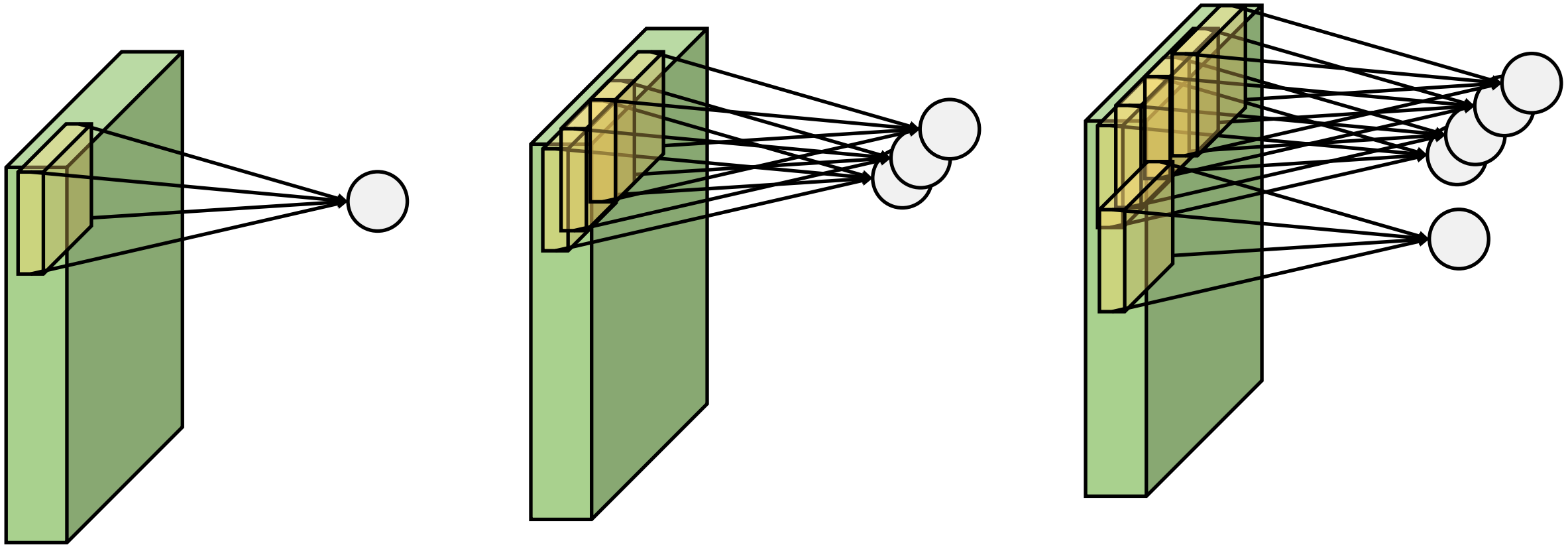
通道数相等

卷积核与原始图像中相同大小的区域进行点乘得到的结果

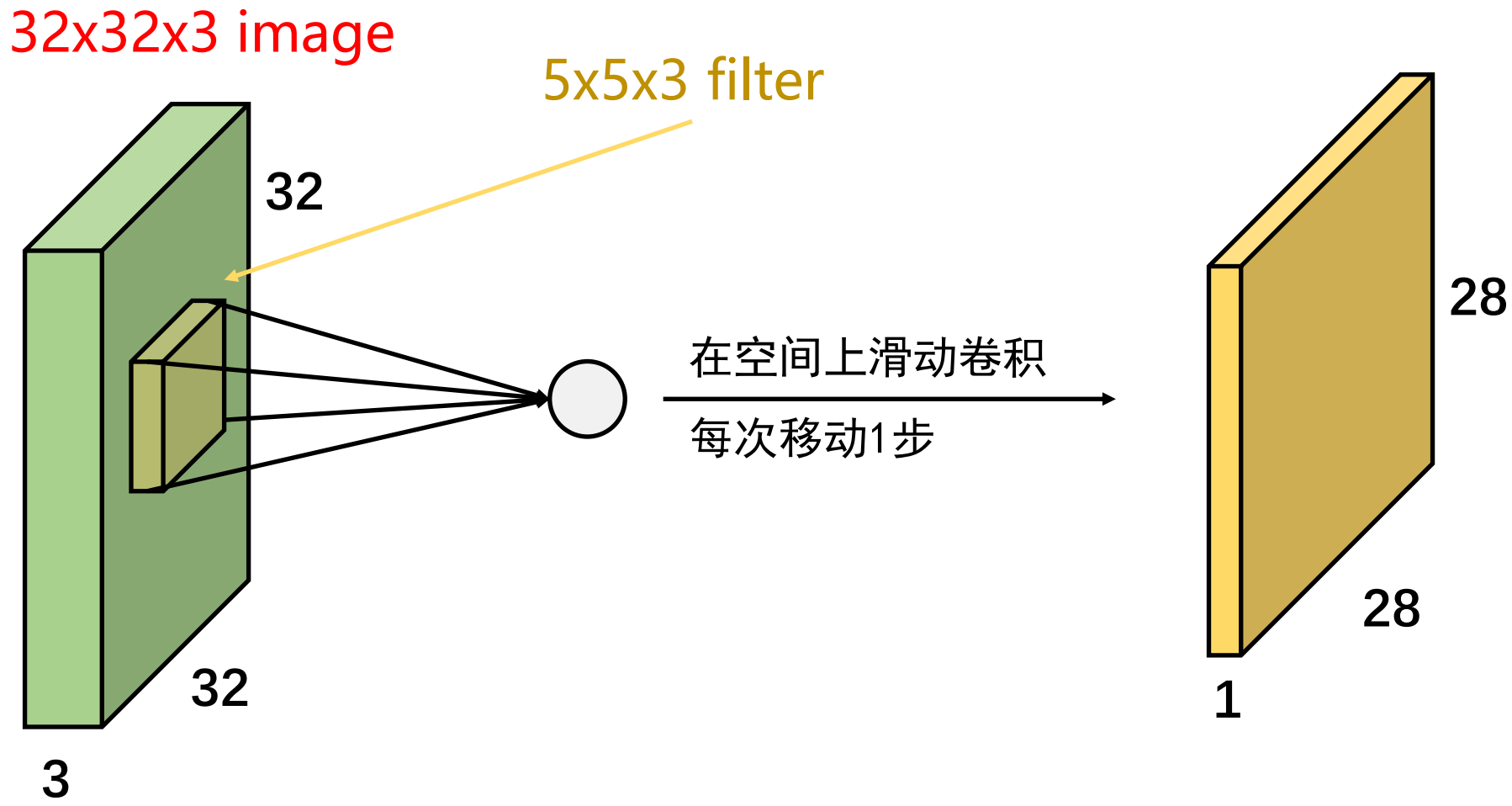
$$\omega^T x + b$$

偏置

卷积层（Convolution Layer）

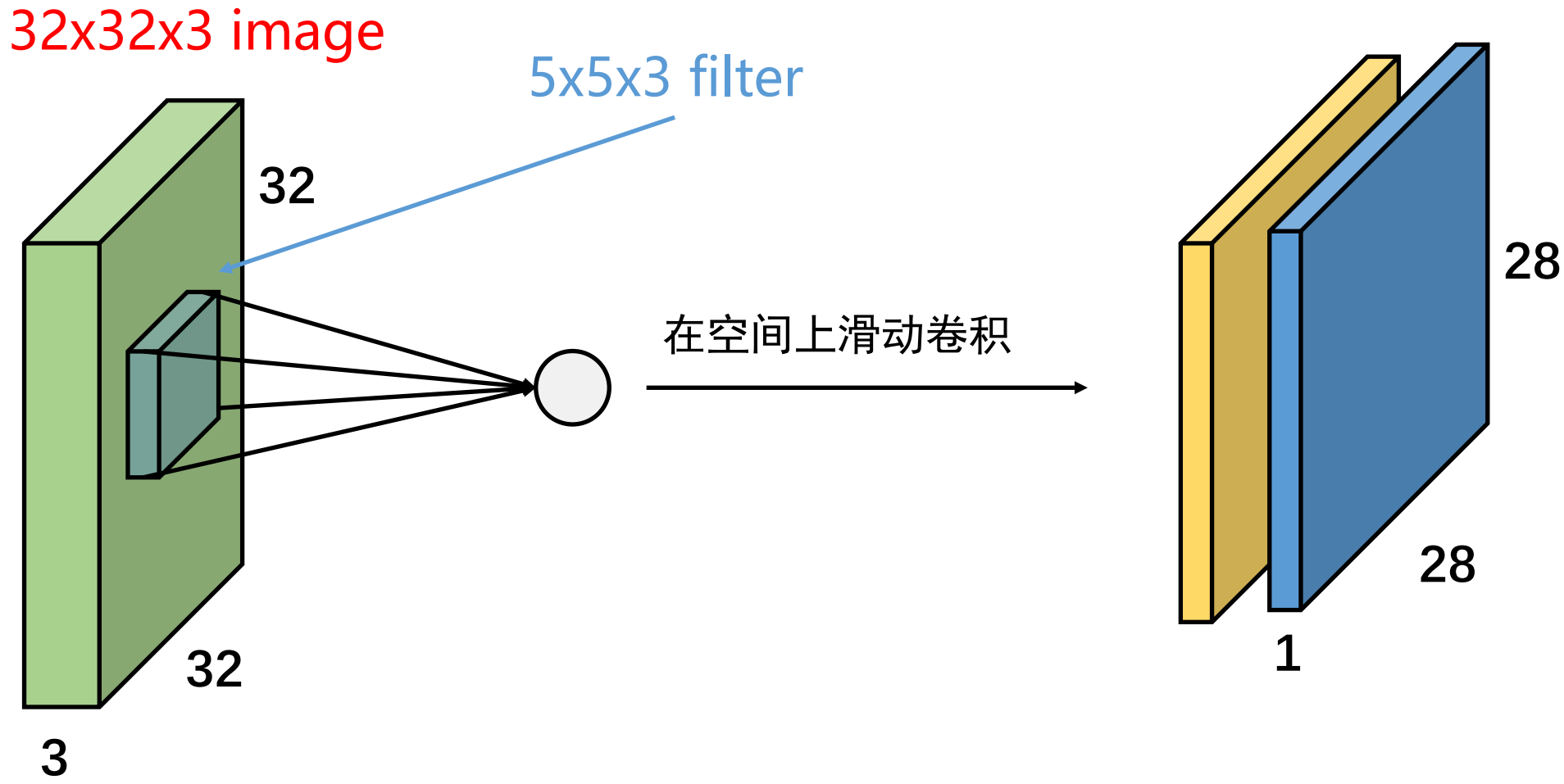


卷积层（Convolution Layer）

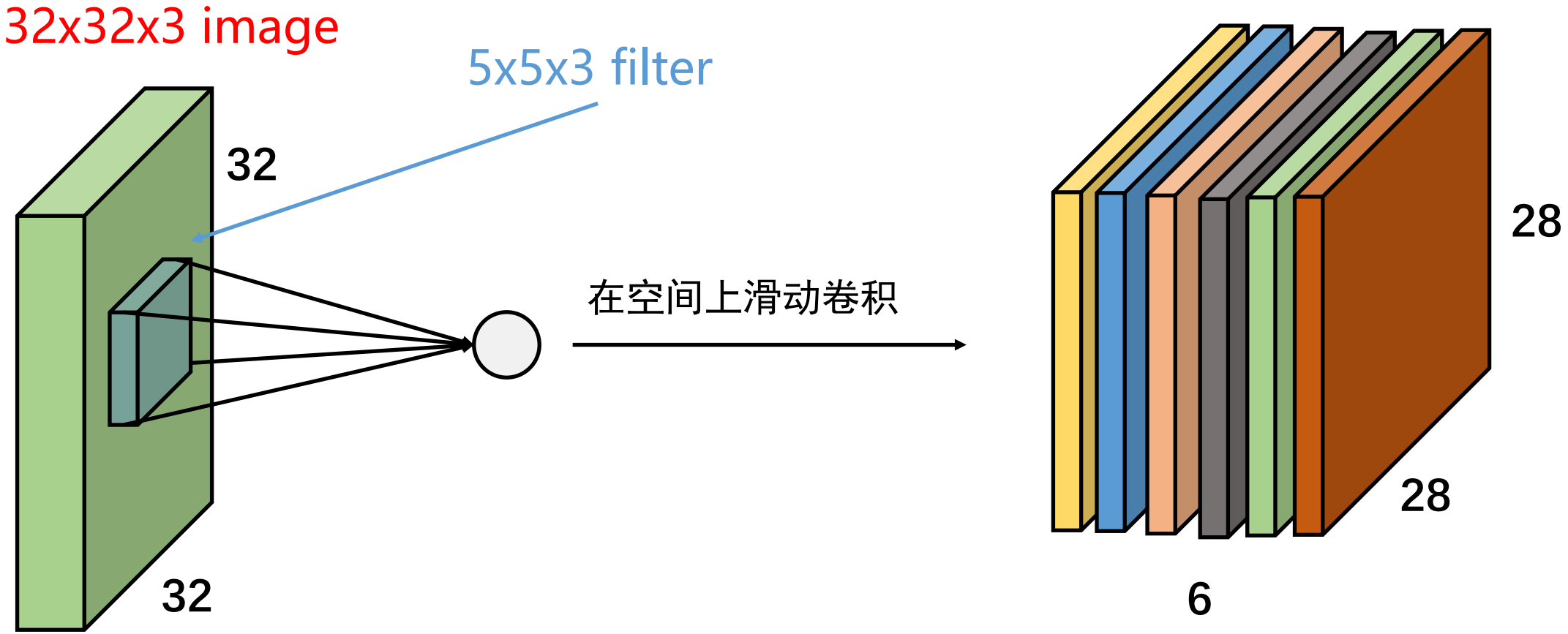


卷积层（Convolution Layer）

使用第二个不同的蓝色卷积核

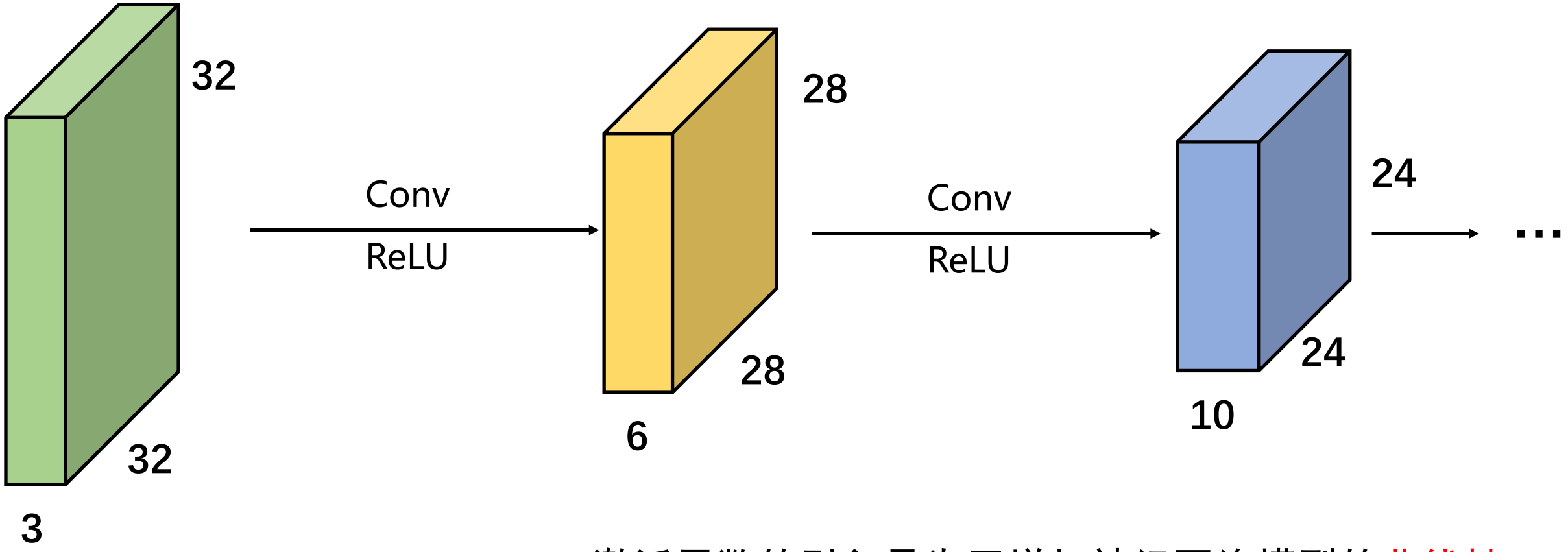


举例：若有6个5x5大小的卷积核，则得到6张不同的特征图



3　将特征图堆叠到一块，得到大小为28x28x6的新特征

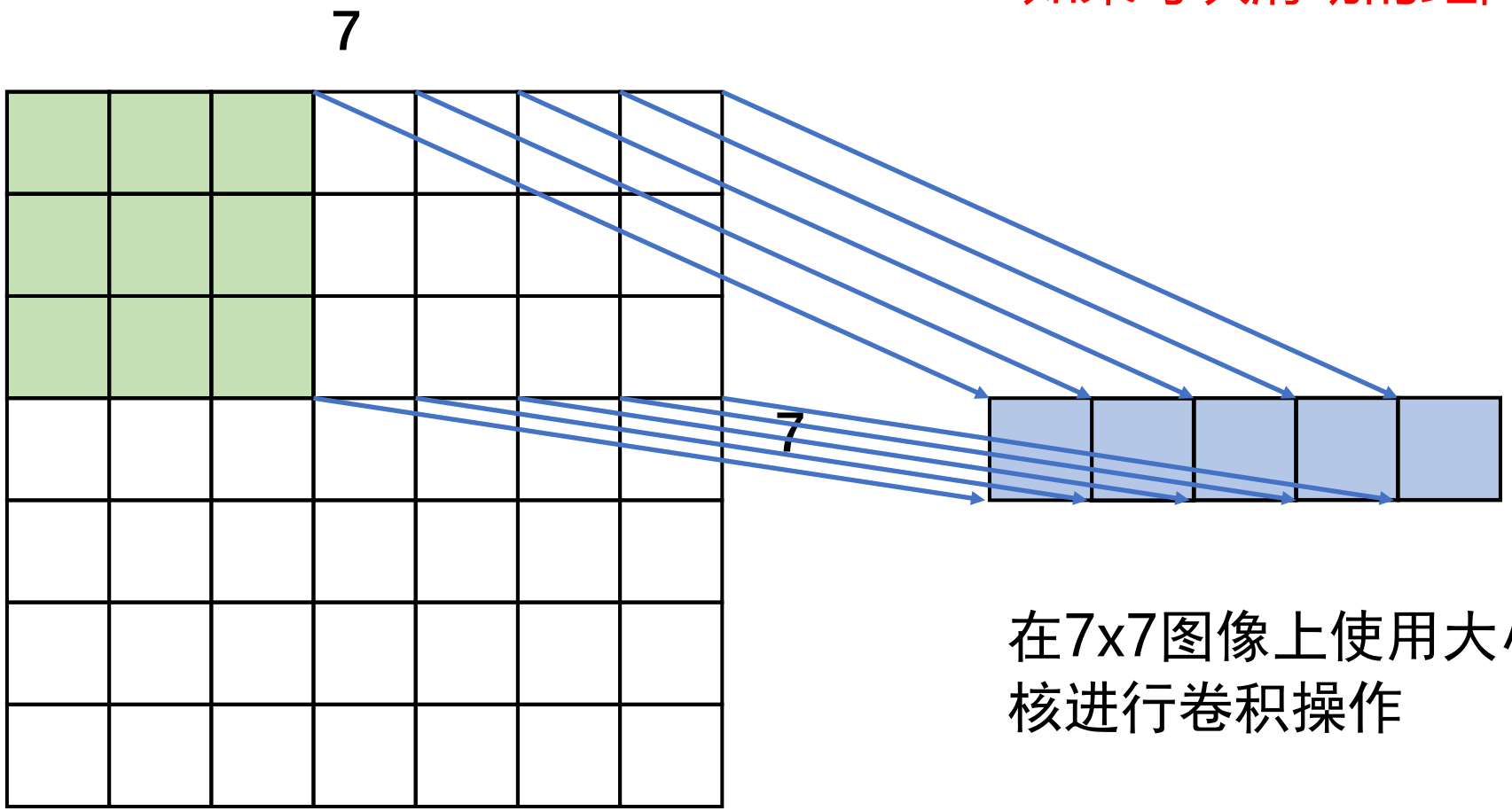
卷积操作常常和ReLU激活函数共同使用　　ReLU: $f(x) = \max(0, x)$



激活函数的引入是为了增加神经网络模型的非线性

空间维度视角下的卷积操作：

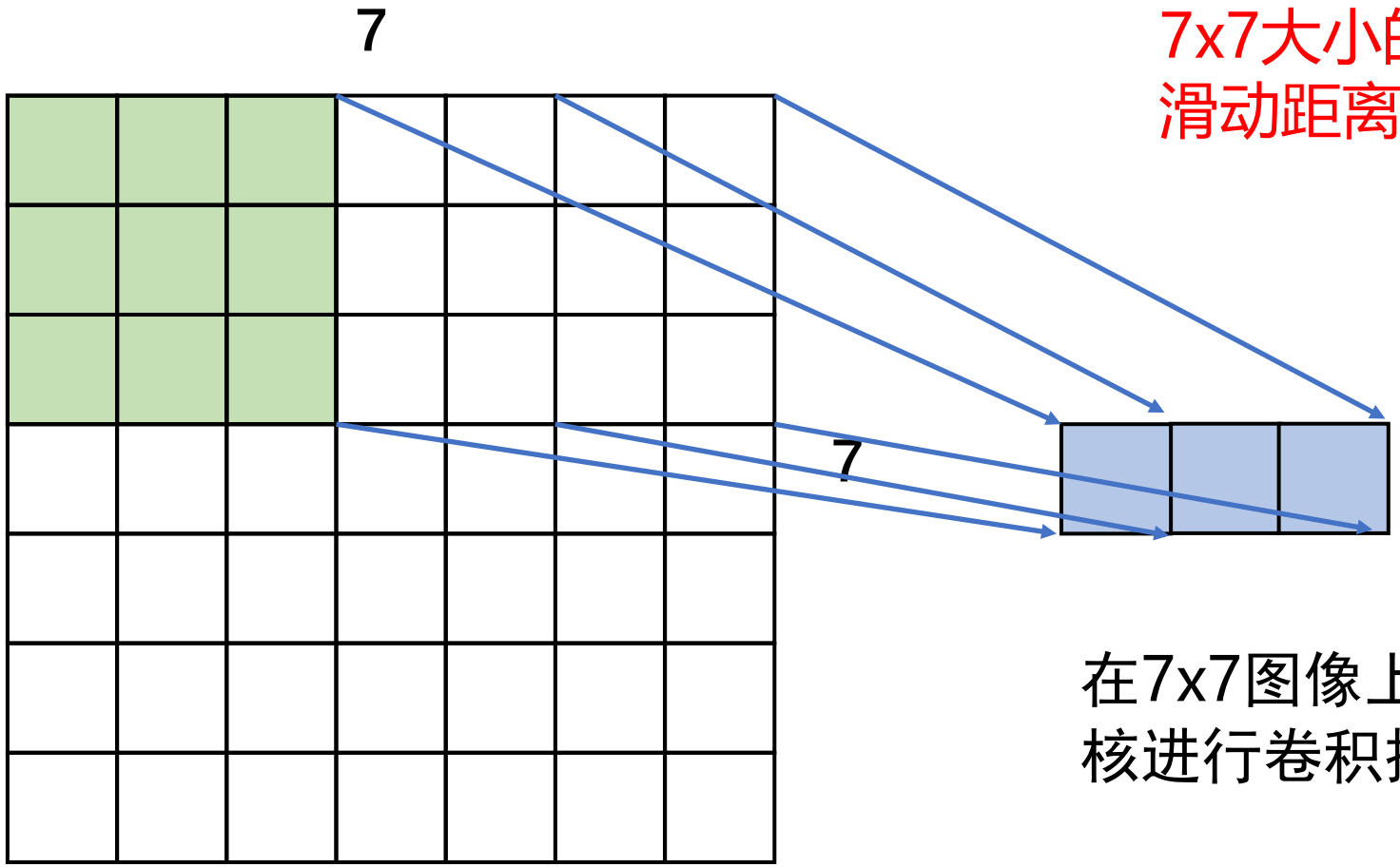
如果每次滑动的距离变大会怎样?



在7x7图像上使用大小为3x3的卷积核进行卷积操作

→ 输出大小为5x5

空间维度视角下的卷积操作：

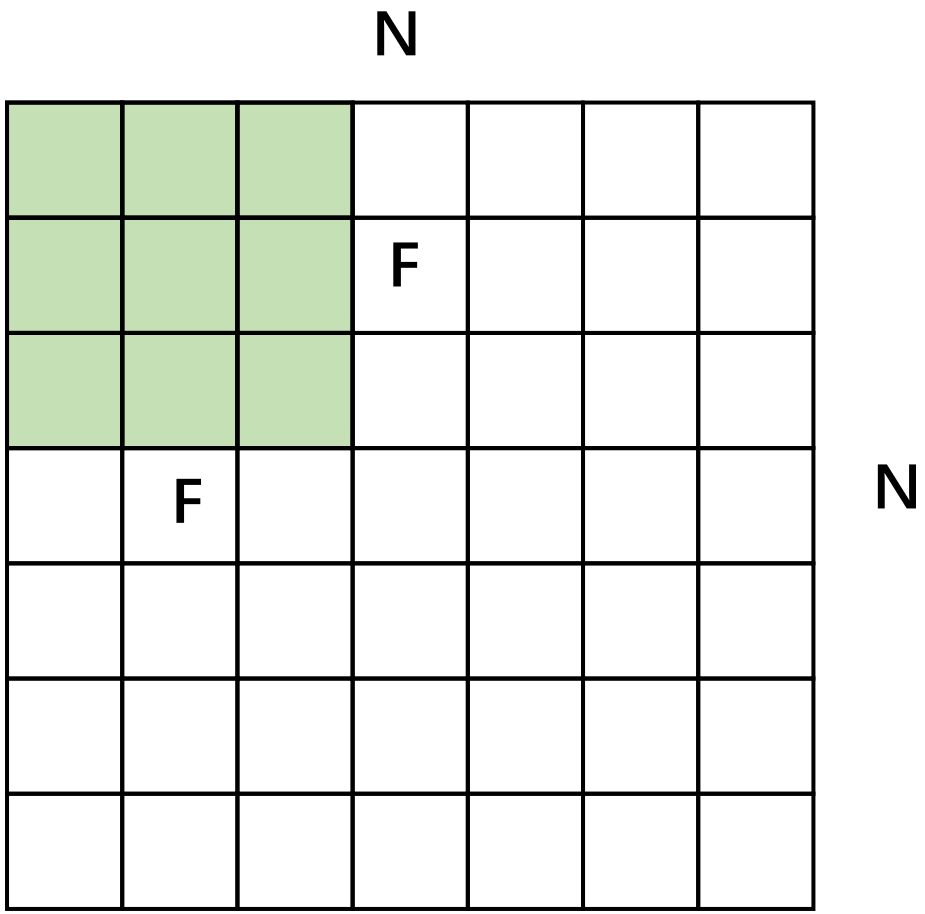


滑动的距离能变成3吗?

7x7大小的输入无法使用步长(单次滑动距离)为3的3x3卷积核

在7x7图像上使用大小为3x3的卷积核进行卷积操作，每次移动两格

→ 输出大小为3x3！



卷积操作输出尺寸计算公式：

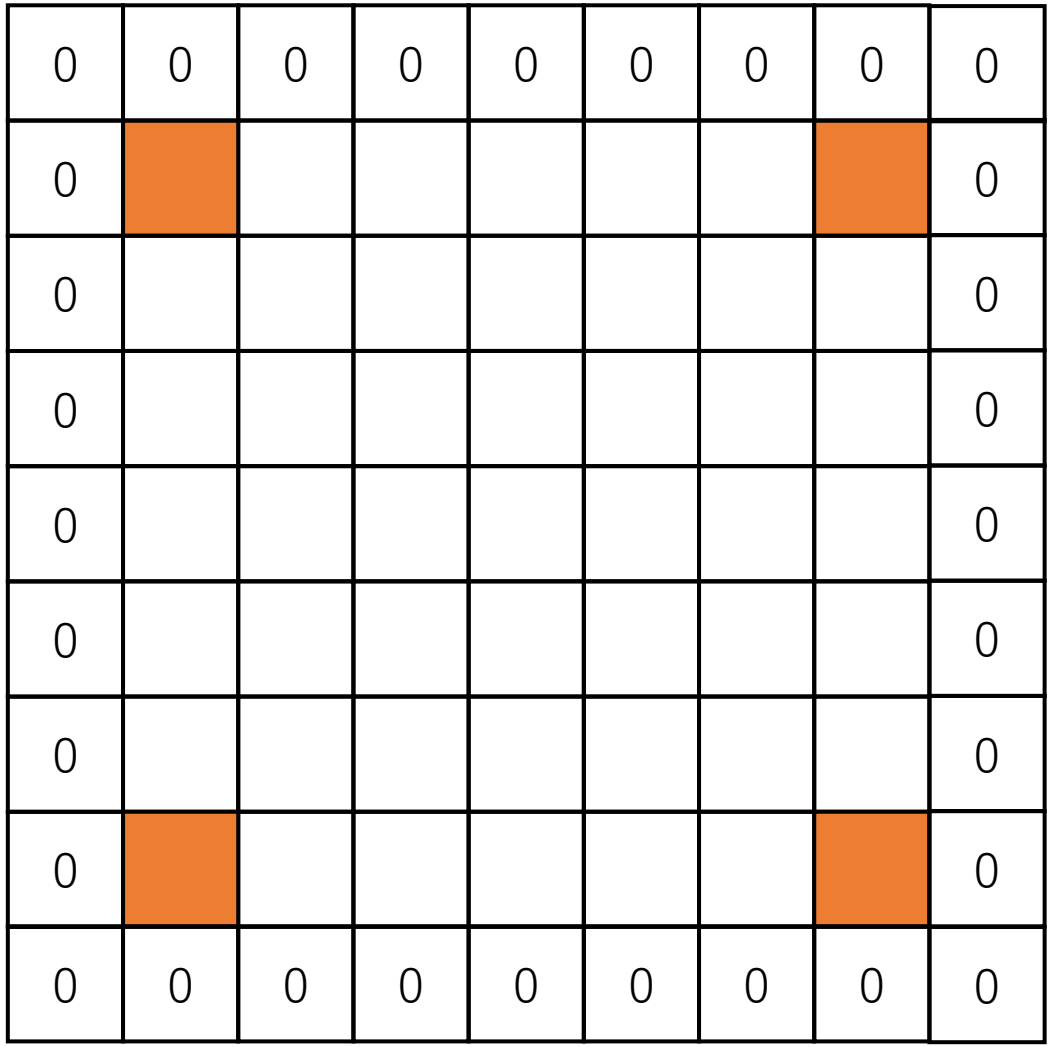
（N-F）/步长+1

举例：当 N=7， F=3：

步长为1：(7-3)/1+1=5

步长为2：(7-3)/2+1=3

步长为3：(7-3)/3+1=2.33



遇到的问题：

1.图像角落的像素点都只进行了一次计算

2.图像尺寸经过卷积后缩小过快

解决方法：

Padding：用 ‘0’ 填充图像边缘

0	0	0	0	0	0	0	0	0
0								0
0								0
0								0
0								0
0								0
0								0
0								0
0	0	0	0	0	0	0	0	0

在7x7图像上使用大小为3x3的卷积核进行卷积操作(步长为1)，padding设置为1，输出尺寸是多少？

➔ 输出大小为7x7
尺寸没有减小！

卷积操作输出尺寸计算公式：

(N−F+2P)/步长+1

N=7，F=3，P=1 => (7−3+2)/1+1=7

小测时间：在32x32x3图像上使用10个大小为5x5的卷积核进行卷积操作

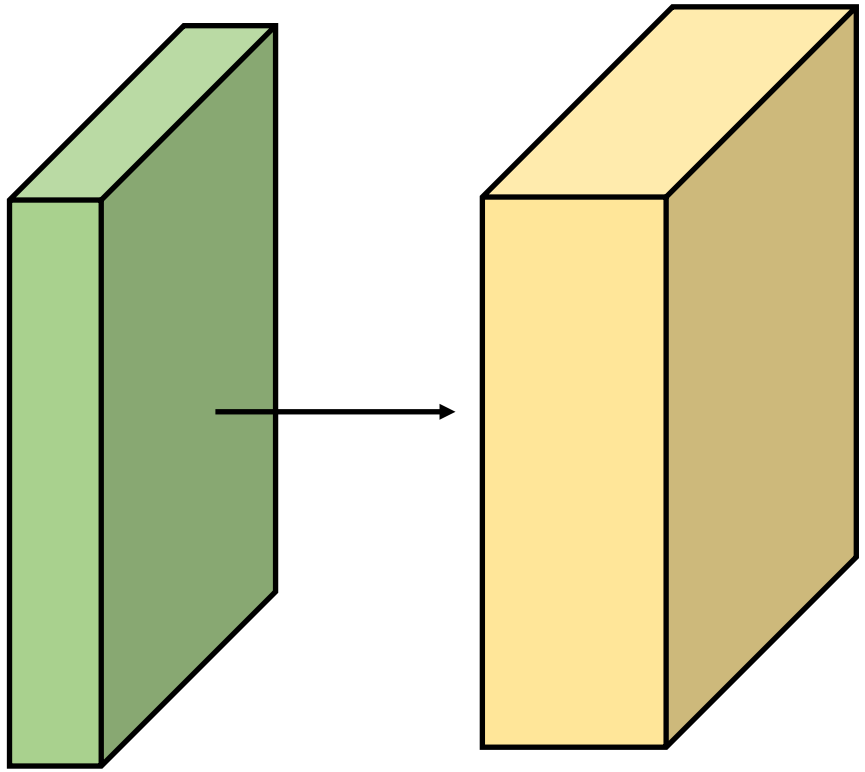
其中步长为1，padding设置为2，请问卷积后输出尺寸是多少？

提示：卷积操作输出尺寸计算公式：(N-F+2P)/步长+1

输出尺寸为：

(32-5+2*2)/1+1=32，由于有10个卷积核

故最终输出大小为32x32x10

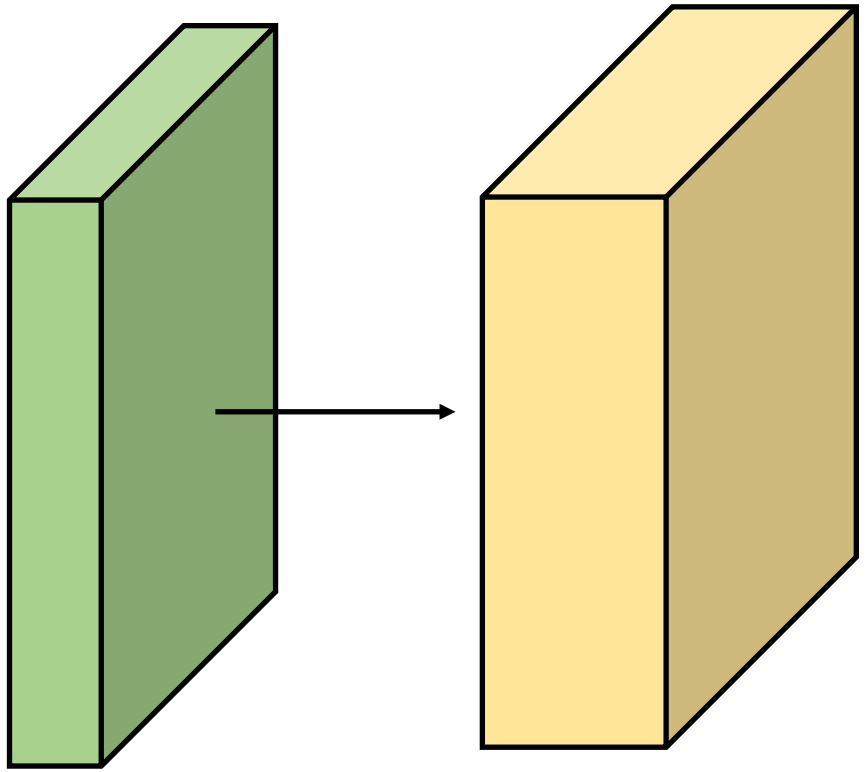


小测时间：在32x32x3图像上使用10个大小为5x5的卷积核进行卷积操作其中步长为1，padding设置为2，请问该卷积层的参数量是多少？

偏置

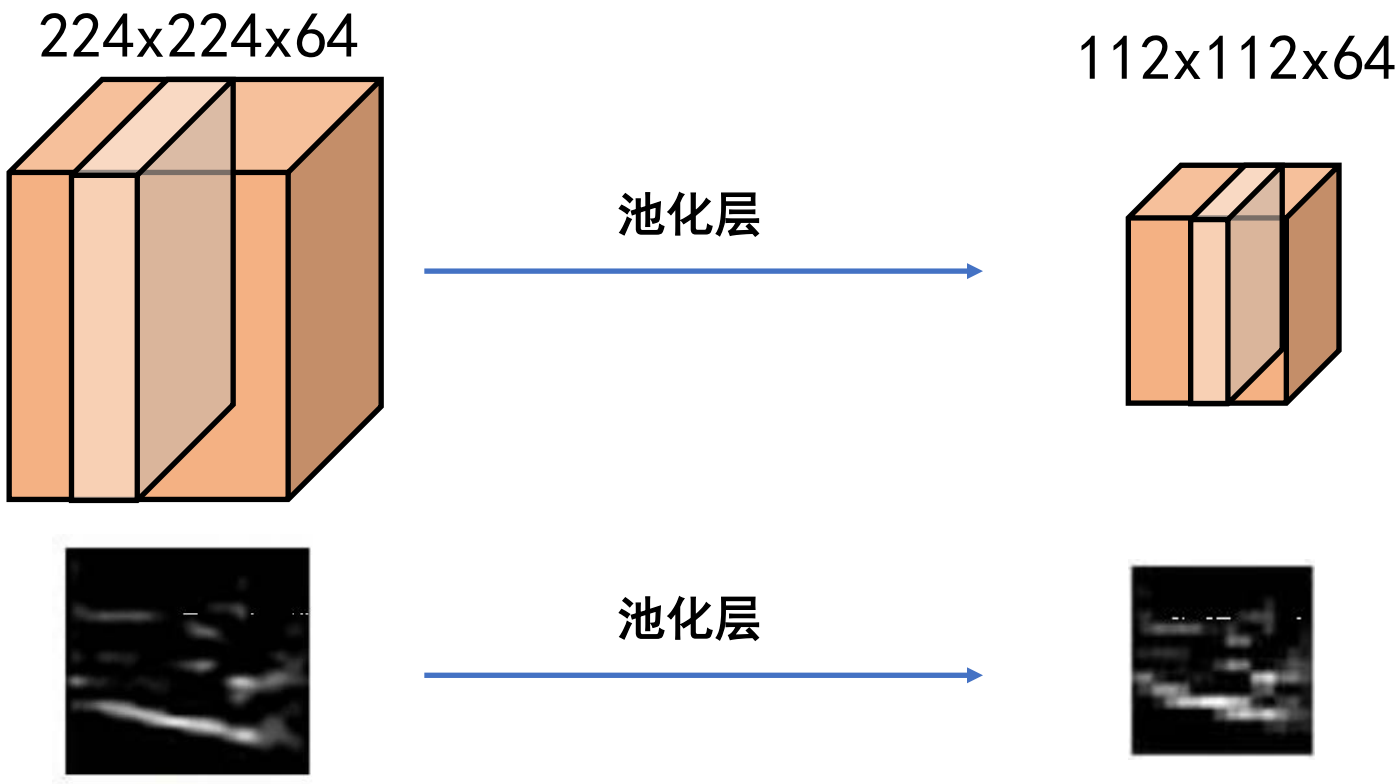
每个卷积核参数为：5x5x3+1=76

故总参数量为76x10=760

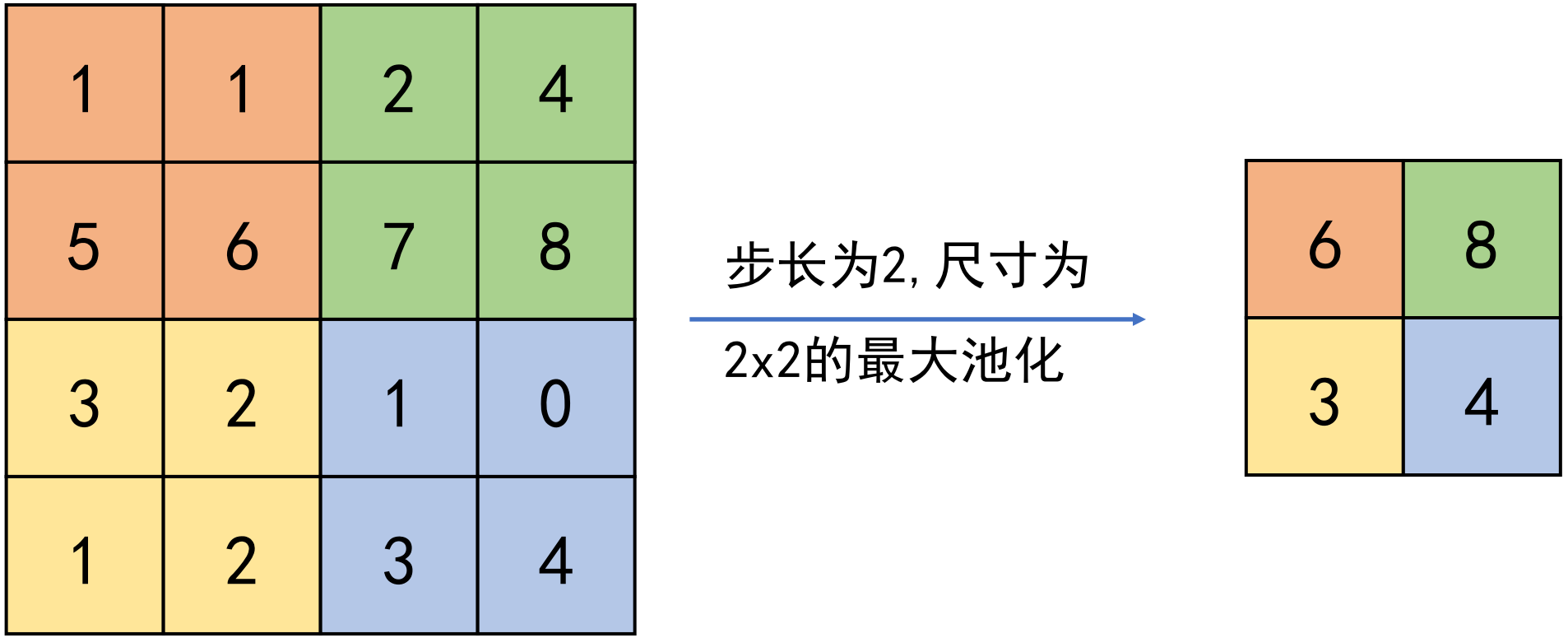


池化层(Pooling layer)

–使学习到的特征尺寸变得更小，从而更容易进行处理
（过大尺寸对显存要求过高）
–在每个特征图上单独进行操作



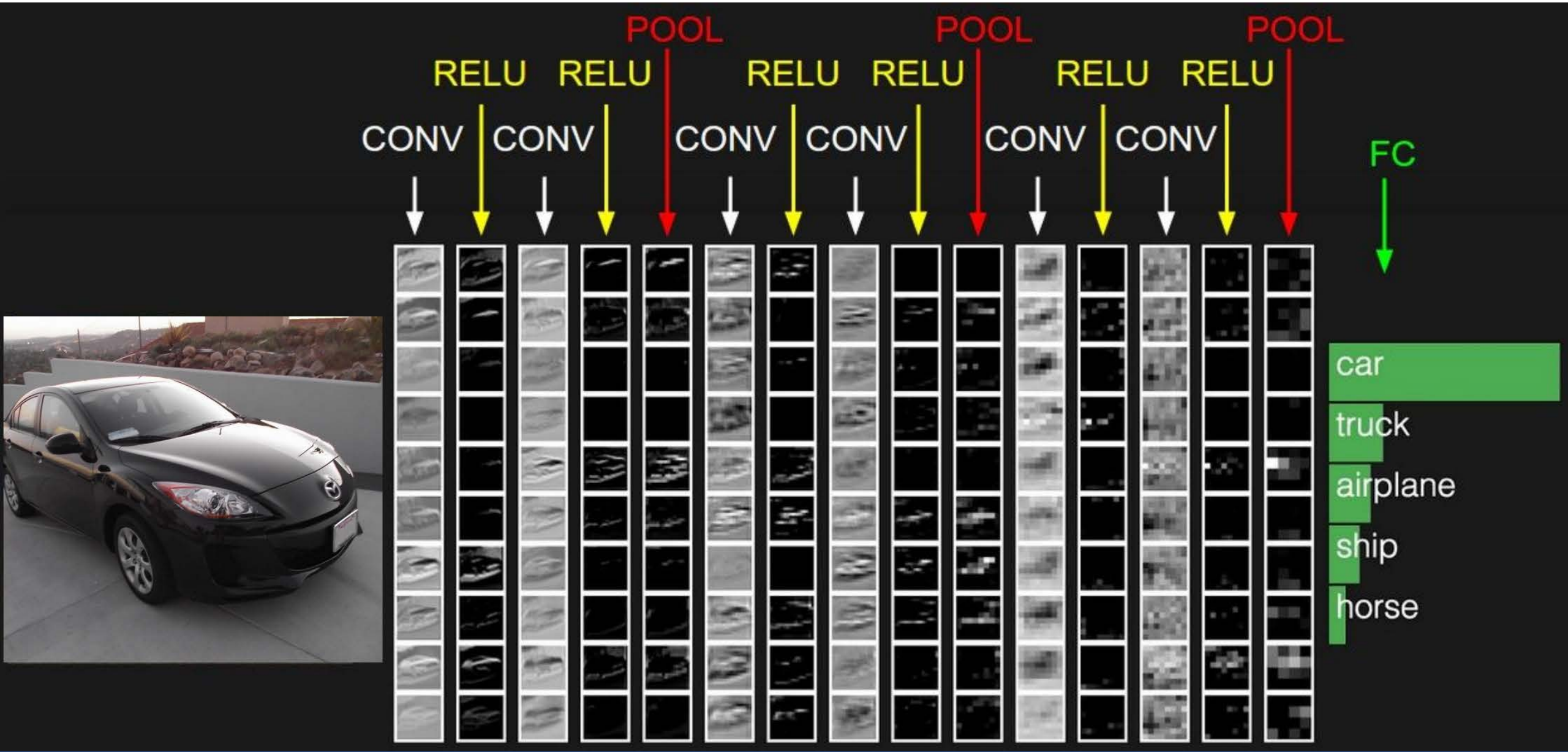
最大池化(Max Pooling)



平均池化(Average Pooling)

以图像分类为例：

ConvNet demo: training on CIFAR-10

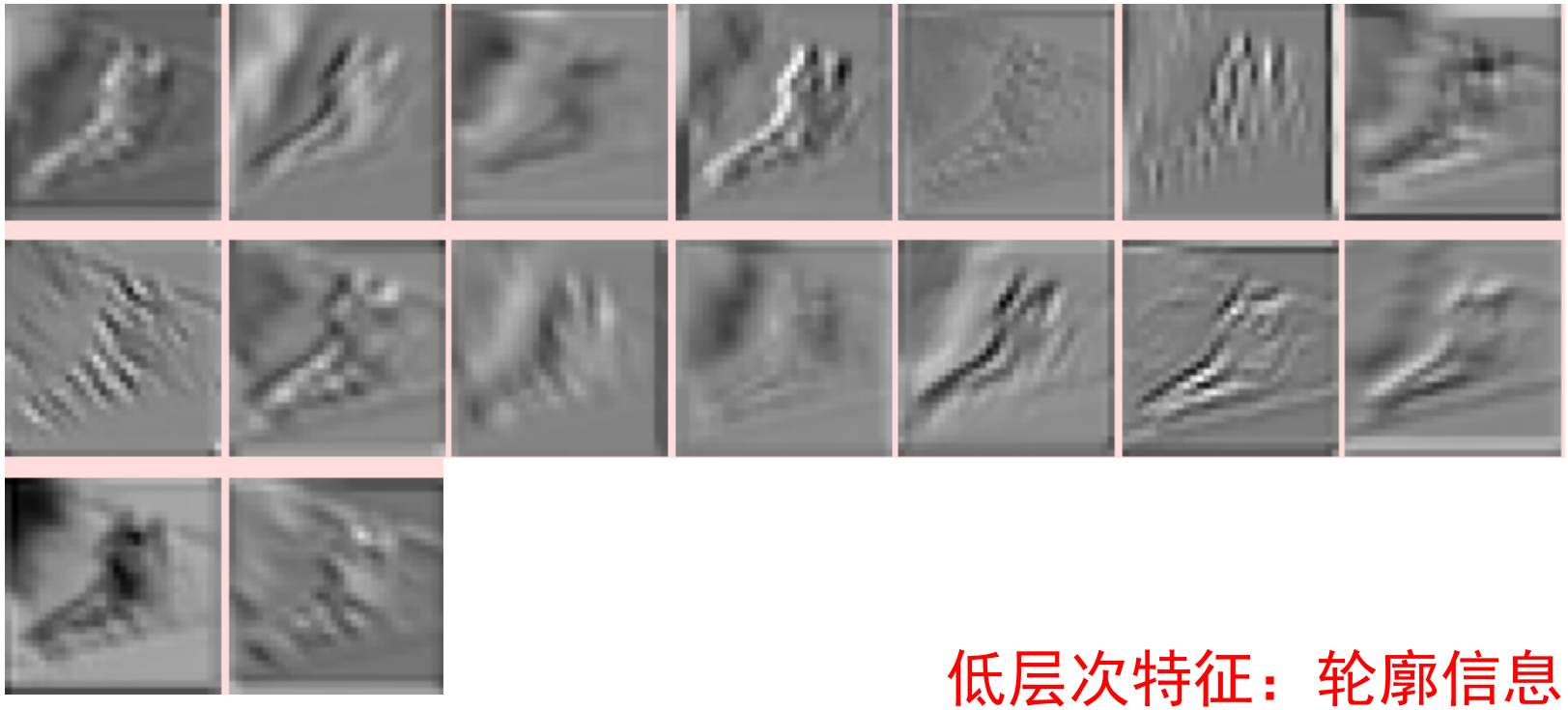


以图像分类为例：

ConvNet demo: training on CIFAR-10

输入图像(32x32x3)

经过第一层卷积得到的特征图(32x32x16)



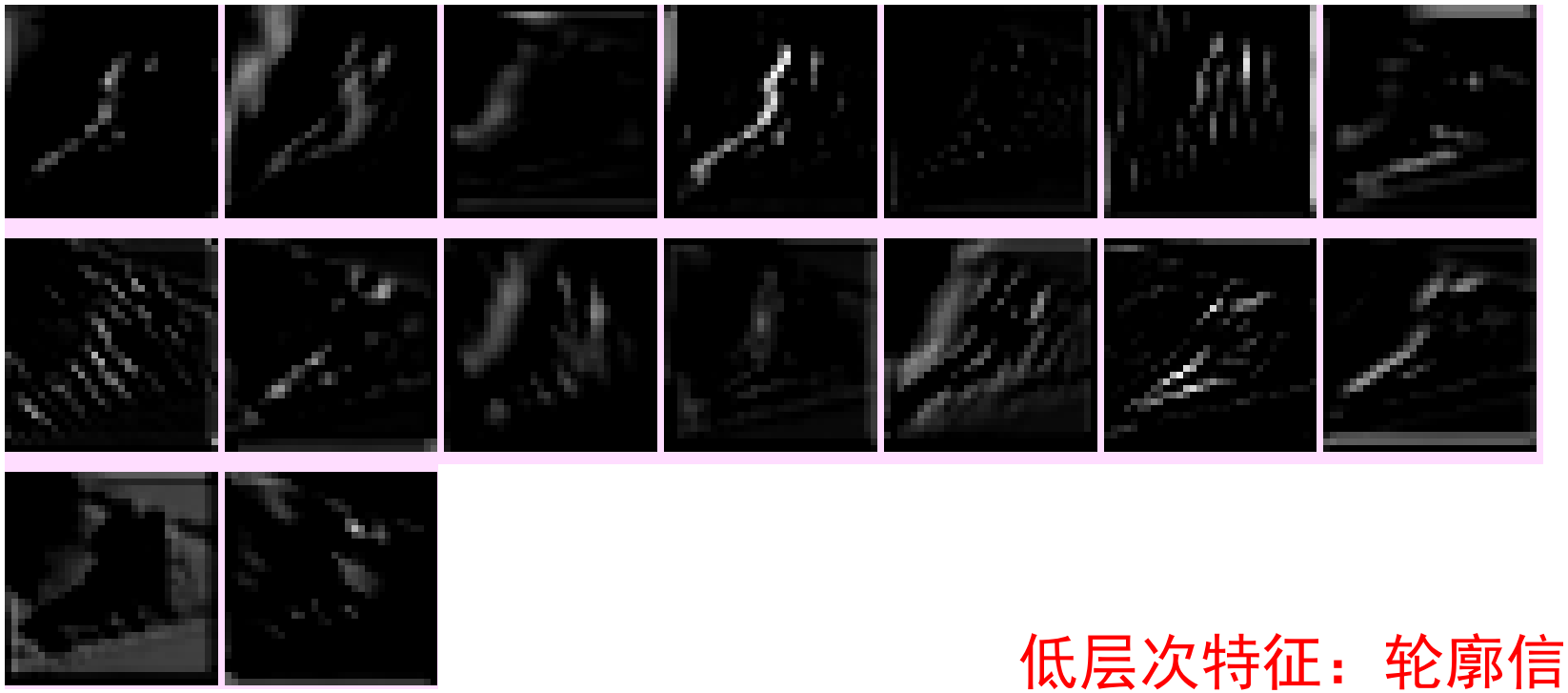
低层次特征：轮廓信息

以图像分类为例：

ConvNet demo: training on CIFAR-10

输入图像(32x32x3)

经过第一层卷积得到的特征图通过ReLU激活函数(32x32x16)



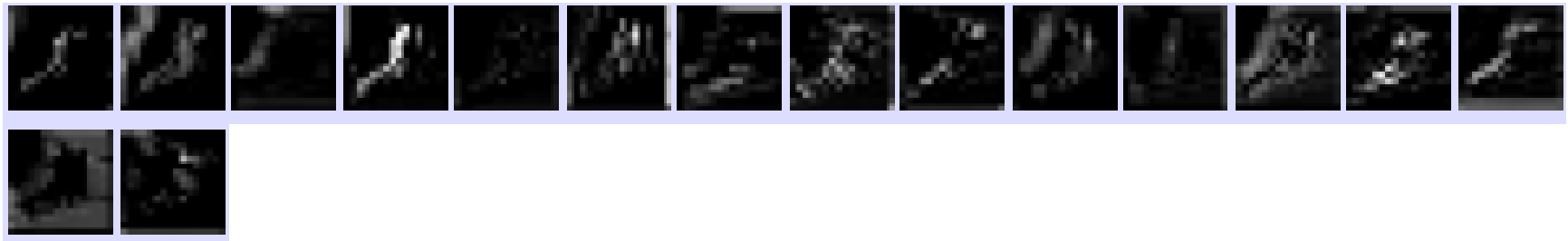
低层次特征：轮廓信息

以图像分类为例：

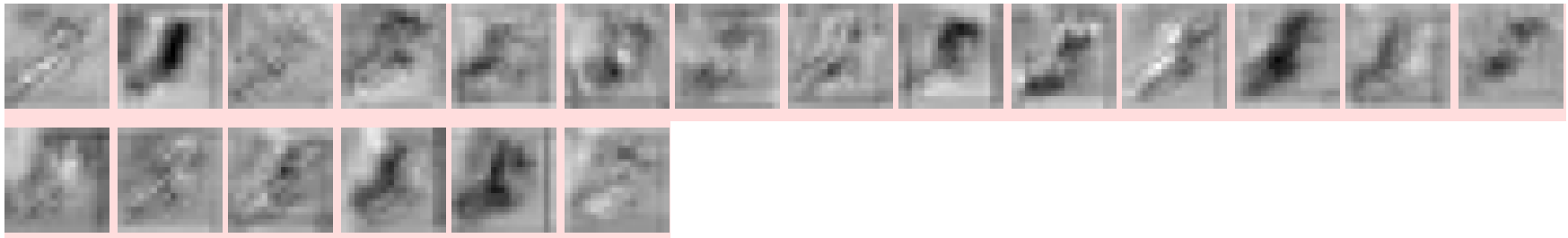
ConvNet demo: training on CIFAR-10

输入图像(32x32x3)

经过池化层(16x16x16)



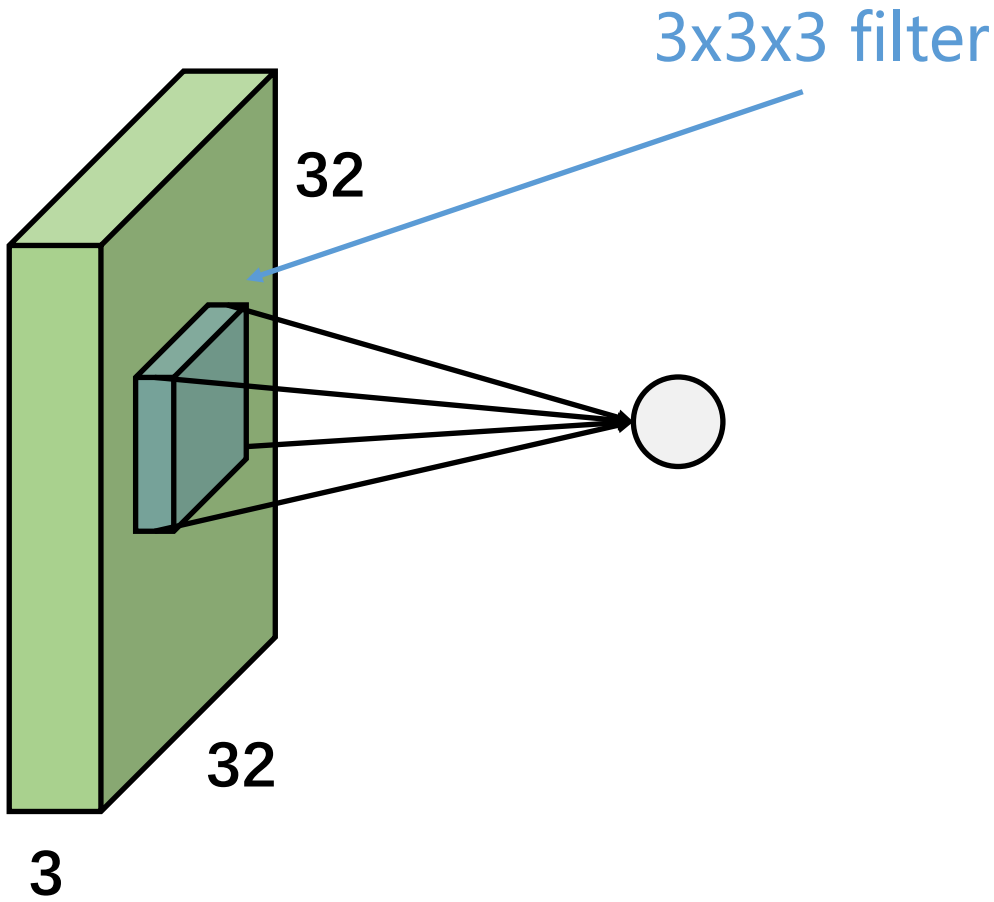
经过第二层卷积层
(16x16x20)



高层次特征：将简单特征组合成复杂一点的特征

感受野(Receptive Field)

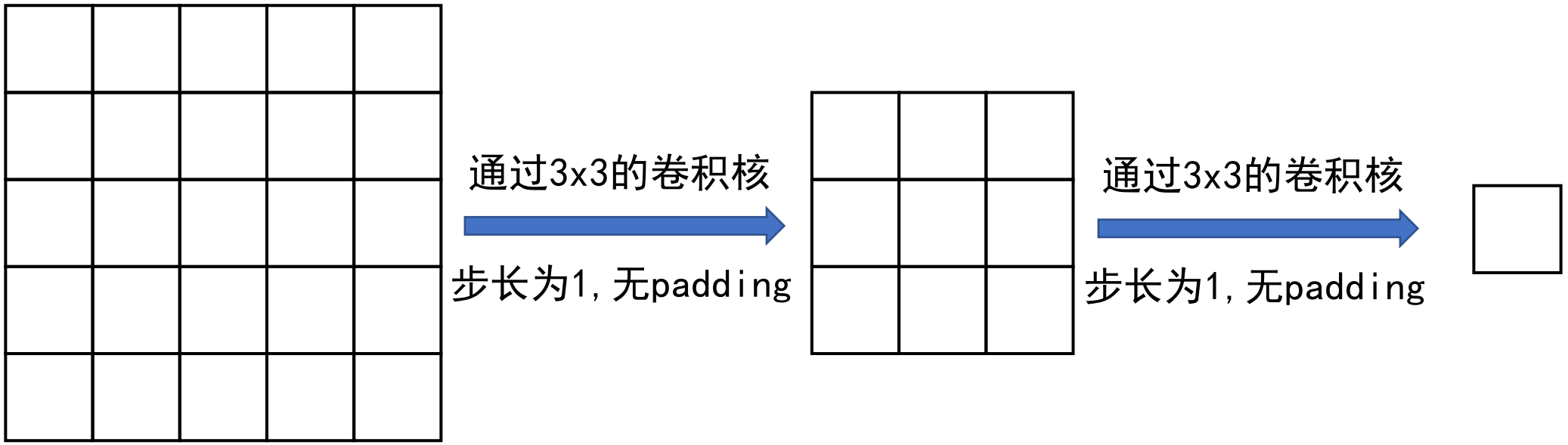
定义：卷积神经网络每一层输出特征图上的像素点在输入图像上映射区域的大小。
通俗地讲，就是特征图上一个点对应输入图上的区域。



该卷积核对应的感受野就是3x3

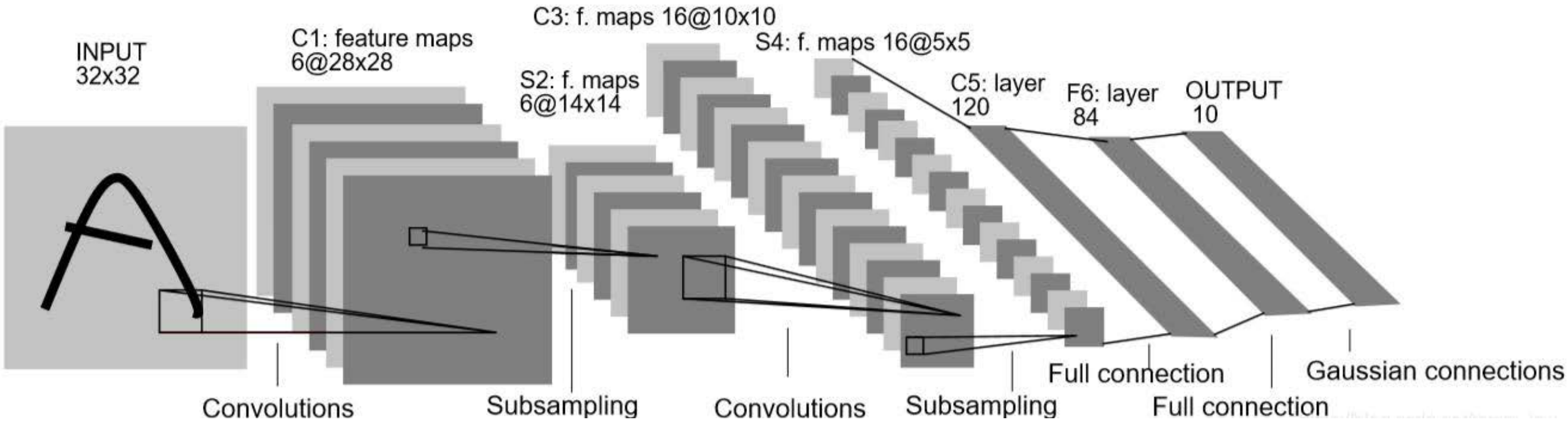
感受野(Receptive Field)

定义：卷积神经网络每一层输出特征图上的像素点在输入图像上映射区域的大小。
通俗地讲，就是特征图上一个点对应输入图上的区域。

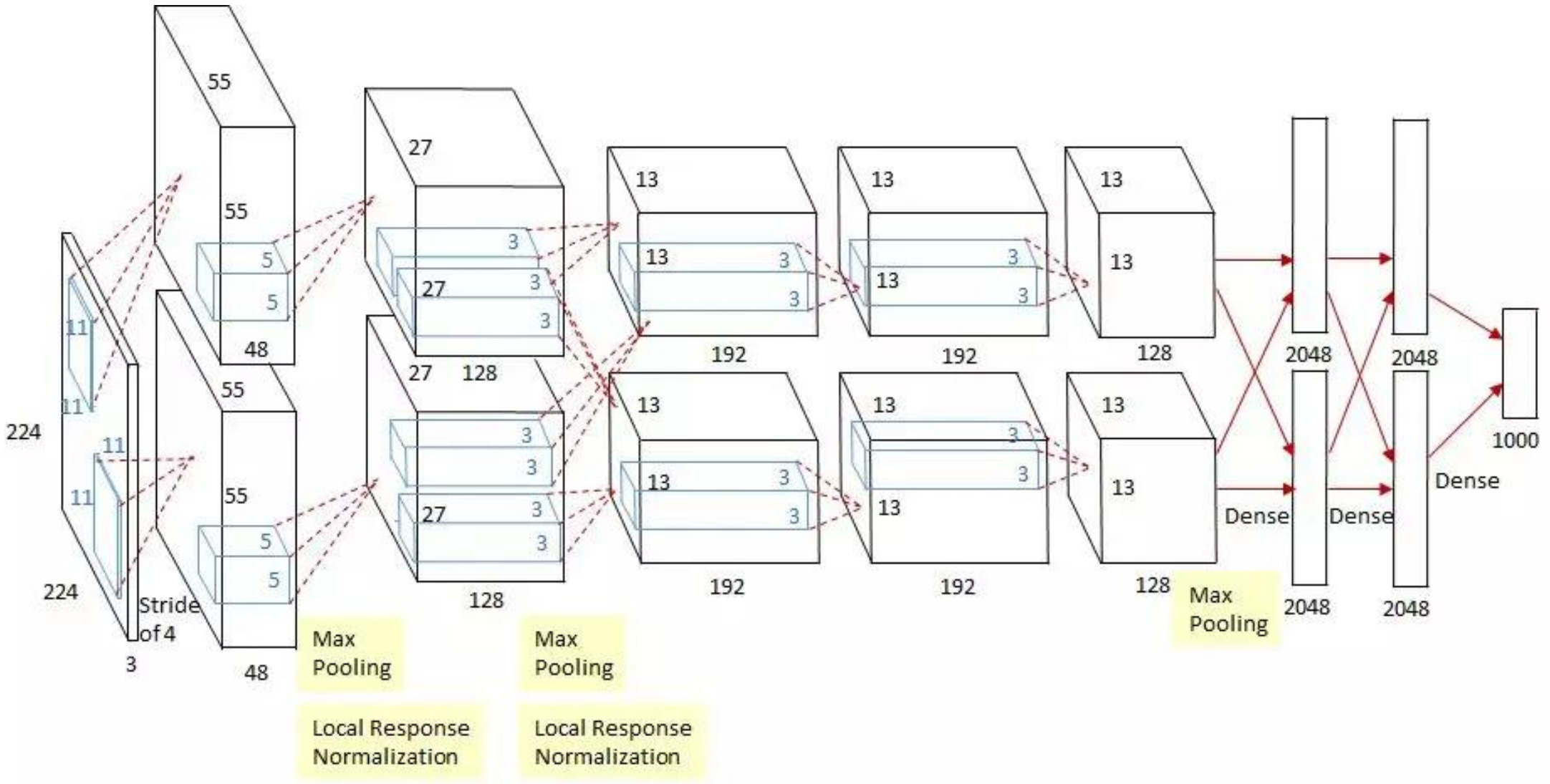


两层3x3卷积操作对应的感受野就是5x5

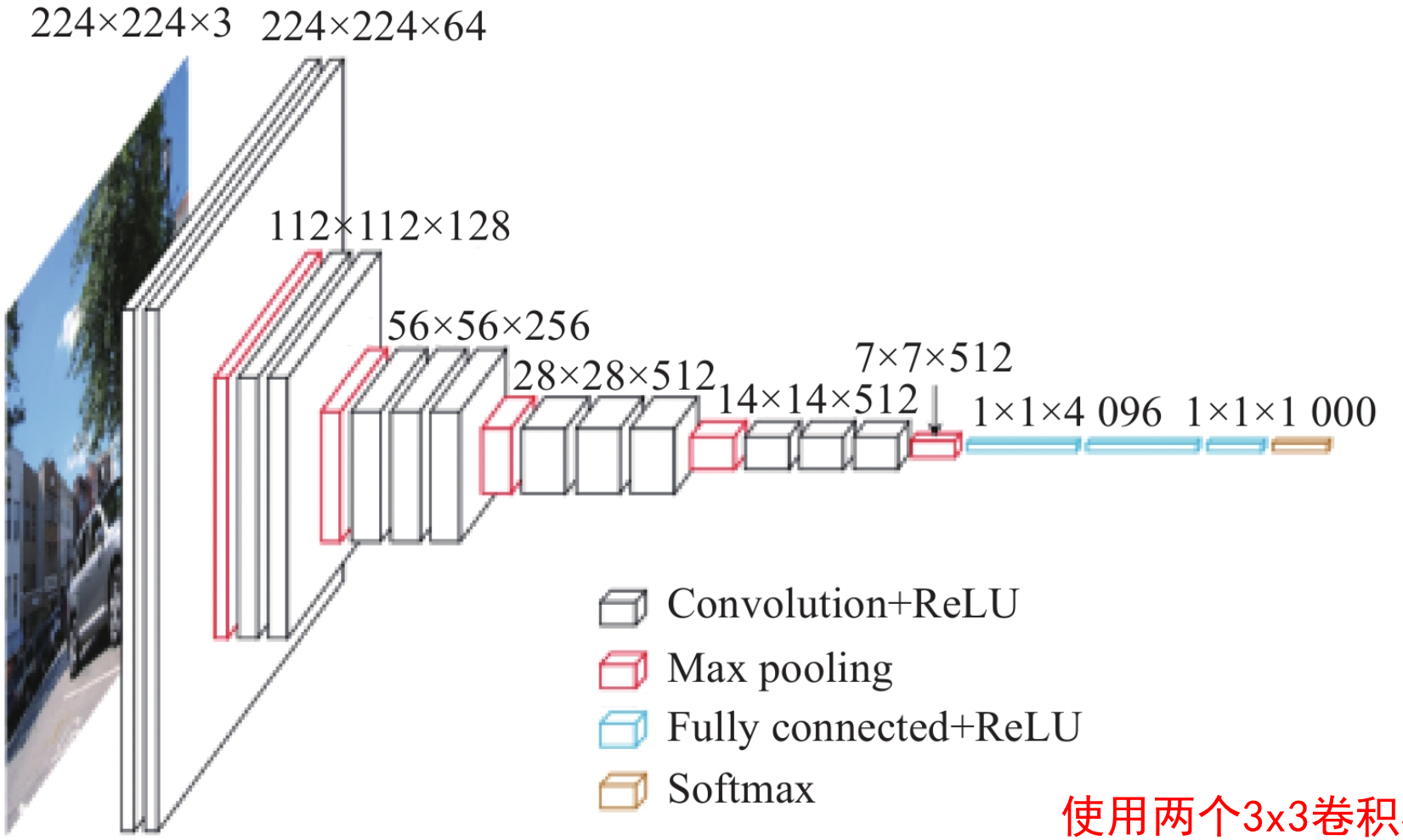
经典卷积神经网络介绍——LeNet



经典卷积神经网络介绍——AlexNet



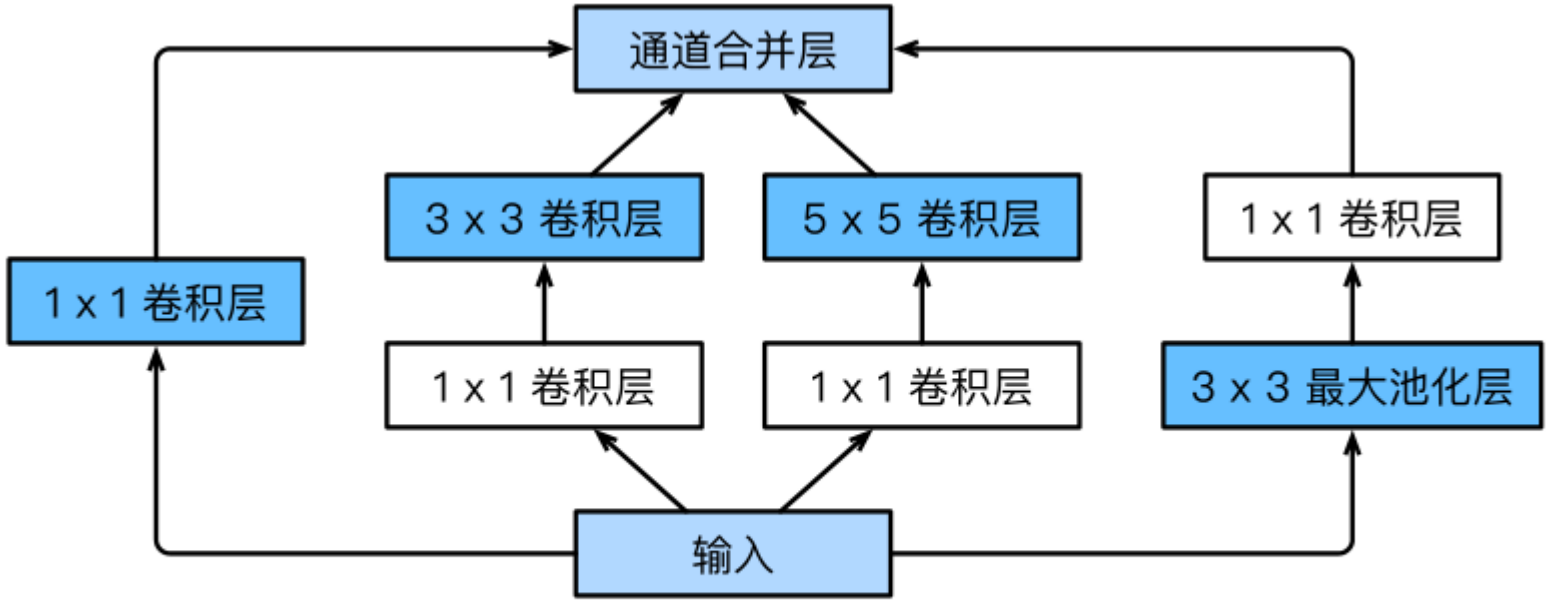
经典卷积神经网络介绍——VGG



使用两个3x3卷积核代替5x5卷积核

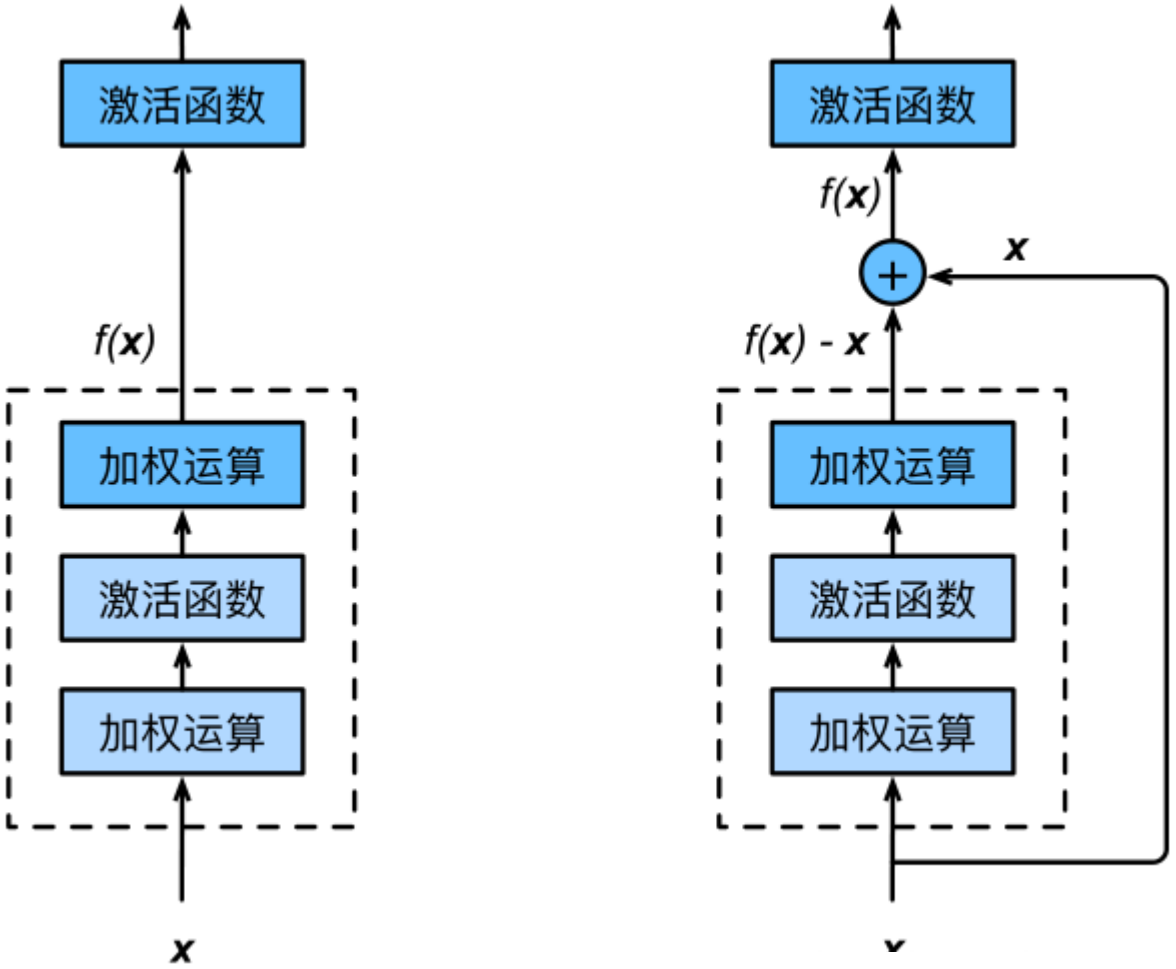
经典卷积神经网络介绍——GoogleNet

GoogLeNet中的基础卷积块叫作Inception块，得名于同名电影《盗梦空间》（Inception）



Inception块的结构

经典卷积神经网络介绍——ResNet



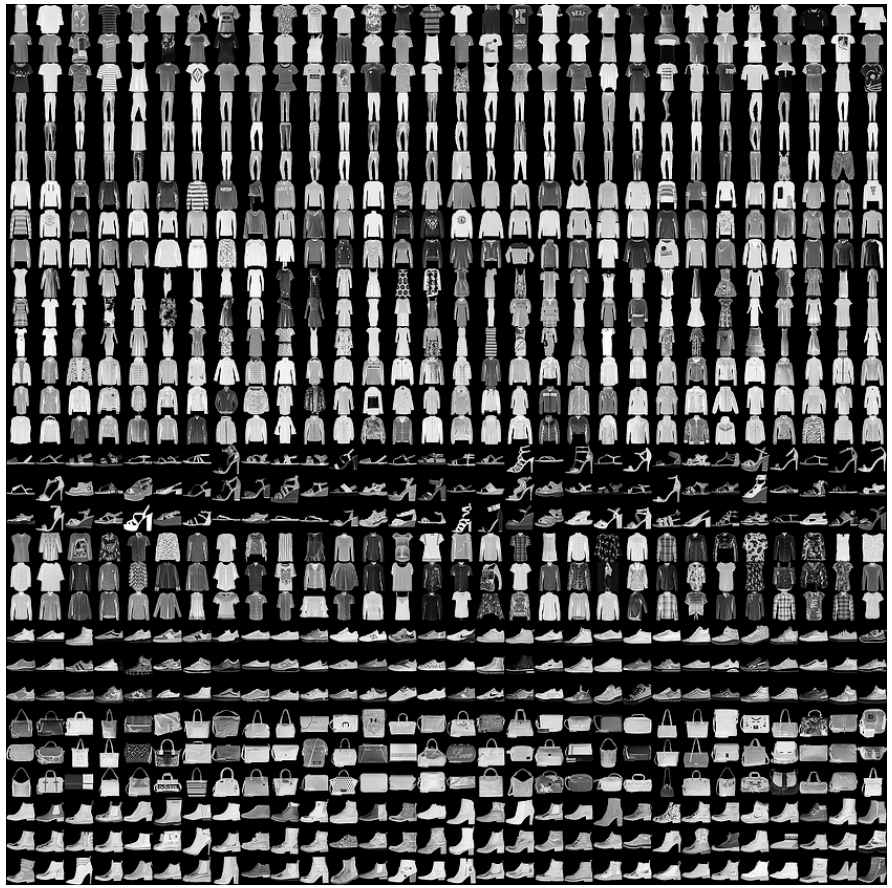
残差块（residual block）

总结

- 卷积神经网络通常由卷积层、池化层、全连接层组成
- 倾向于使用更小的尺寸的卷积核以及更深的网络架构

[(卷积层-ReLU)*N-池化层]*M-(全连接层-ReLU)*K,Softmax

作业五:使用卷积神经网络对Fashion-MNIST数据集进行分类



- 推荐编程环境:Anaconda+Jupyter notebook+pytorch

  安装教程： 点这

- 仿照经典网络搭建自己的模型
- 可以使用 BatchNorm、Dropout等技巧提升分类准确性

参考文献：

1. A.Krizhevsky, I.Sutskever, G.E.Hinton. ImageNet Classification with Deep Convolutional Neural Networks[C]. In Advances in Neural Information Processing Systems (NeurIPS),2012: 1106-1114 AlexNet
2. Y.Lecun, L.Bottou, Y.Bengio and et al. Gradient-Based Learning Applied to Document Recognition. LeNet
3. K.Simonyan, A.Zisserman. Very Deep Convolutional Networks for Large-scale Image Recognition[C]. Proceedings of the International Conference on Learning Representations (ICLR),2015 VGG
4. P.Sermanet, S.Reed, D.Anguelov and et al. Going deeper with convolutions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1-9 GoogleNet
5. K.He, X.Zhang, S.Ren and et al. Deep Residual Learning for Image Recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778 ResNet

Sources

- https://www.youtube.com/watch?v=YRhxdVk_sIs
- Convolutional Neural Networks (CNNs) explained 8mins
- https://www.youtube.com/watch?v=QzY57FaENXg
- What are Convolutional Neural Networks (CNNs)? 6mins
- https://www.youtube.com/watch?v=m8pOnJxOcqY
- Convolution Neural Networks - EXPLAINED 19mins
- https://www.youtube.com/watch?v=d12ra3b_M-0
- Intro to Feature Engineering with TensorFlow 7mins