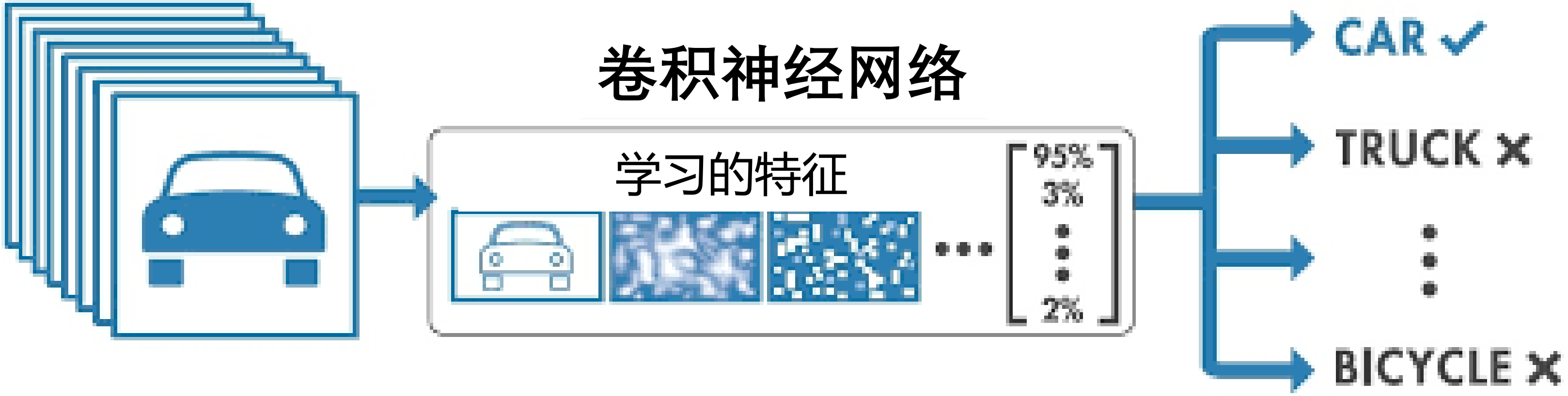
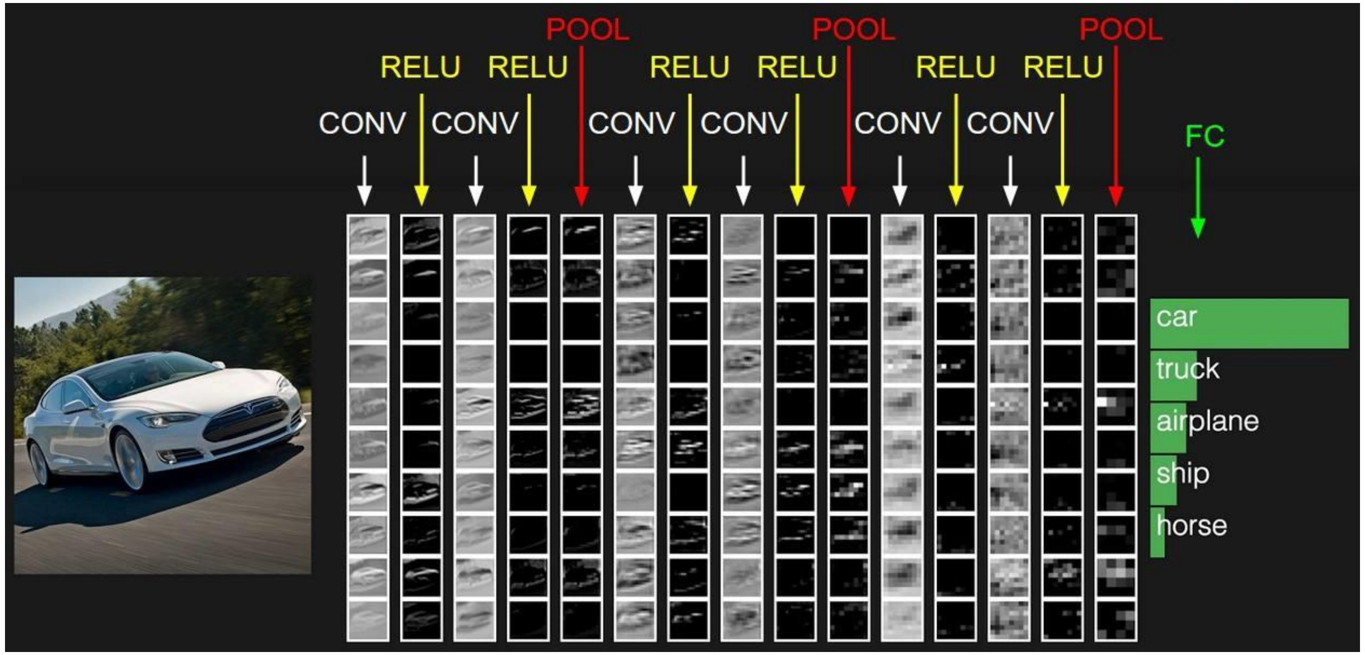
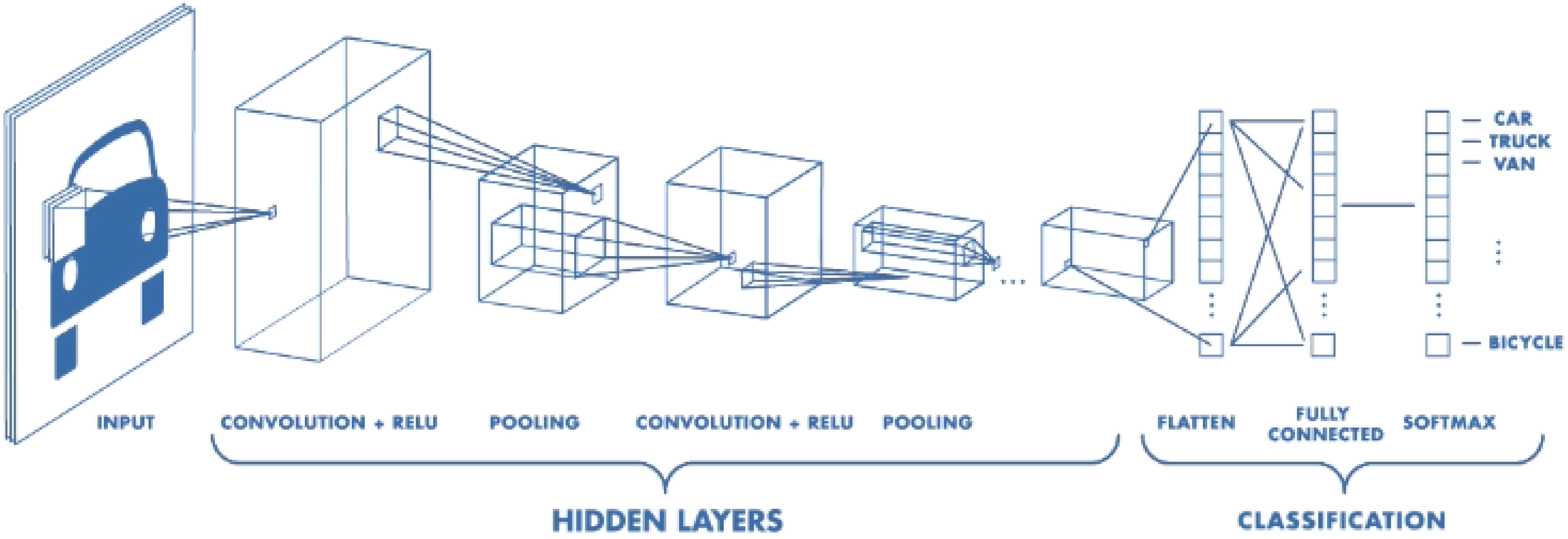
卷积神经网络
CNN (Convolutional neural network)





卷积神经网络架构

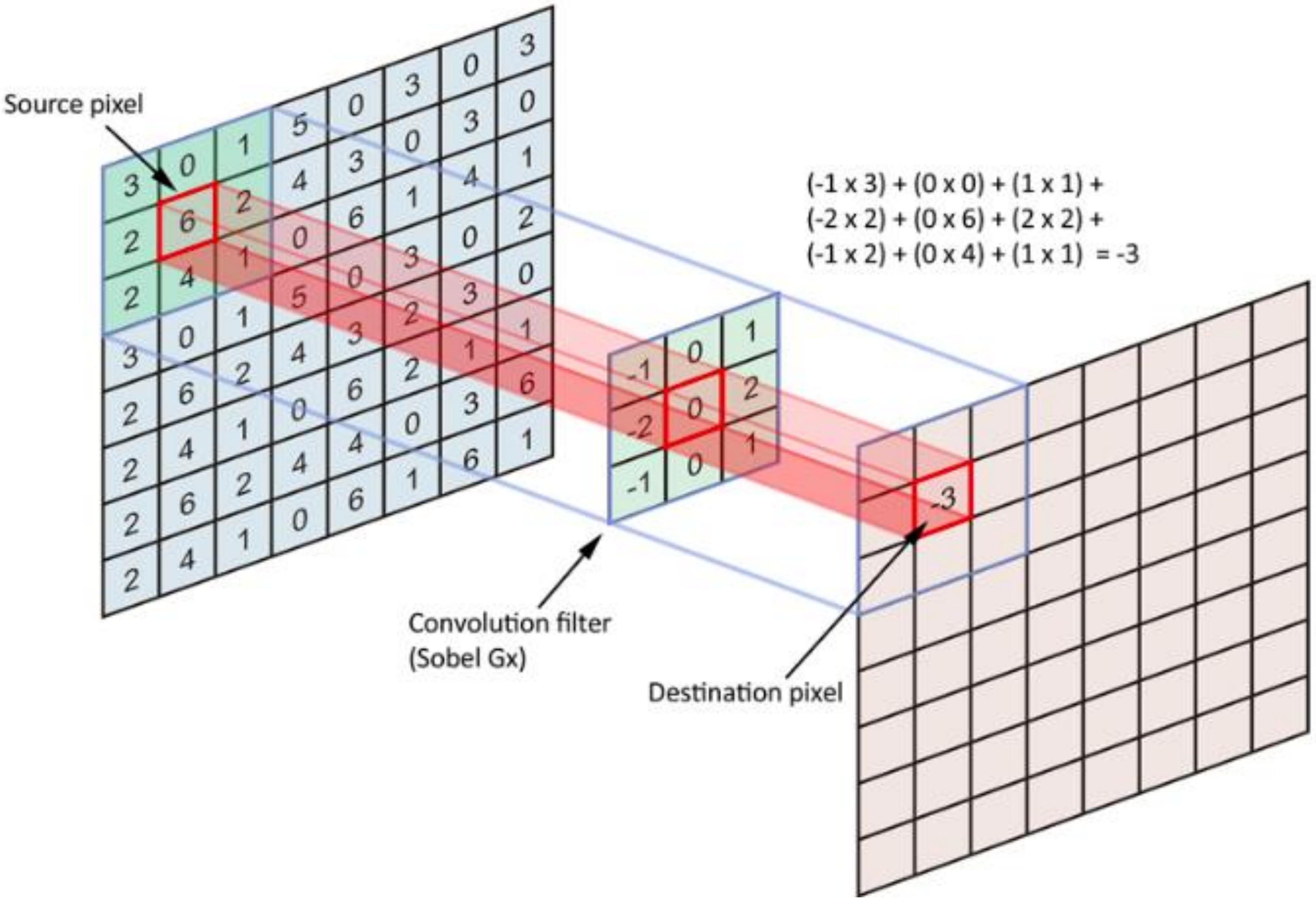


CAR
TRUCK
VAN
BICYCLE
INPUT
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FLATTEN
FULLY CONNECTED
SOFTMAX
HIDDEN LAYERS
CLASSIFICATION

典型卷积神经网络的层结构

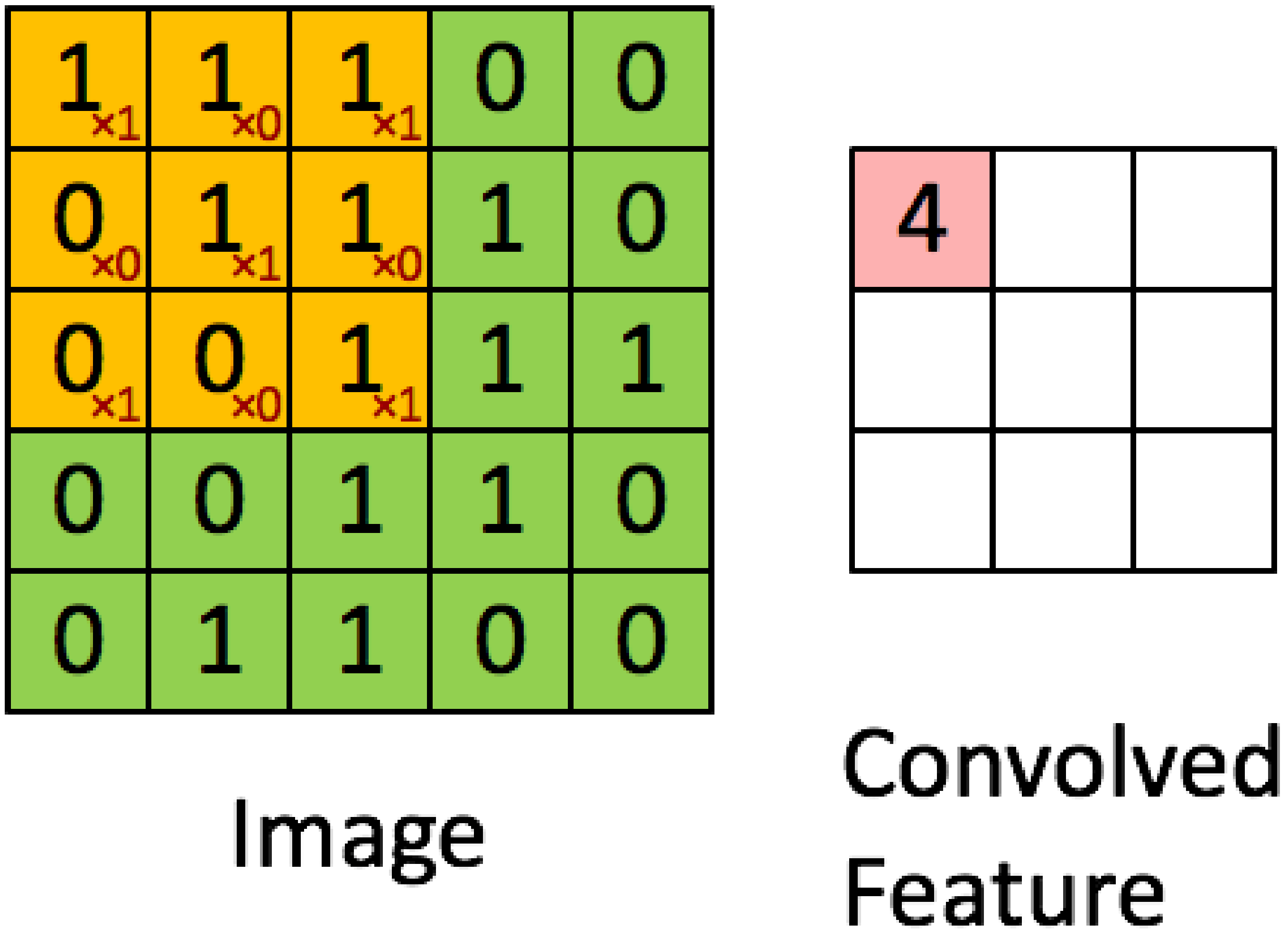
输入 → 卷积 → 非线性激活 → 池化 → 输出：特征图

卷积(Convolution)运算



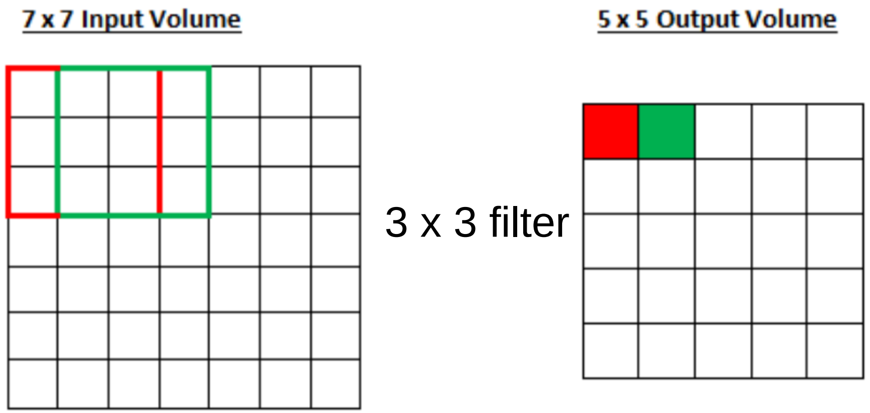
实际上是相关(Cross-Correlation)运算

卷积运算

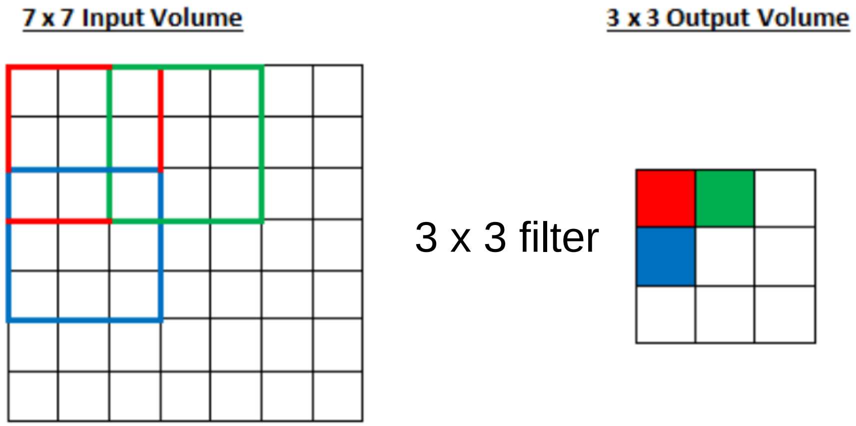


1 ×1
1 ×0
1 ×1
0
0
0 ×0
1 ×1
1 ×0
1
0
0 ×1
0 ×0
1 ×1
1
1
0
0
1
1
0
0
1
1
0
0
4
Image
Convolved Feature

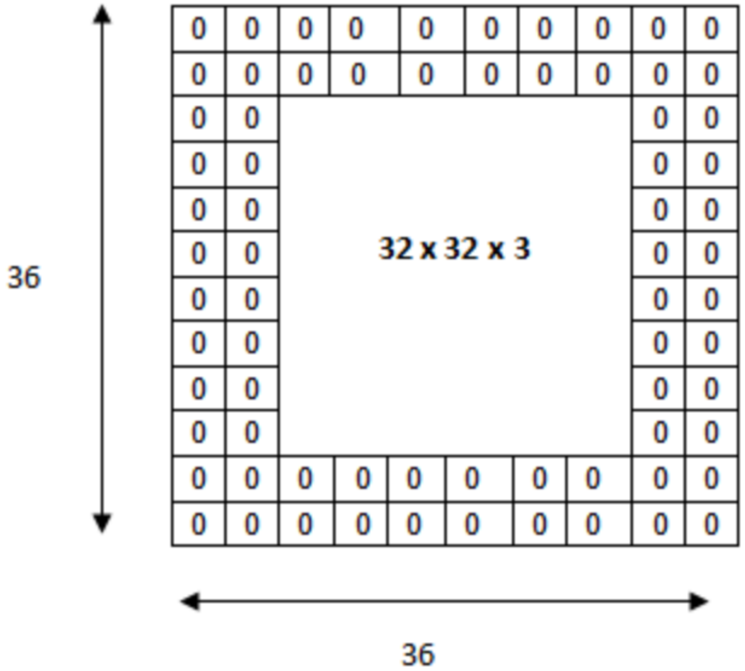
Stride (步幅)和Padding (填充)



stride = 1



stride = 2

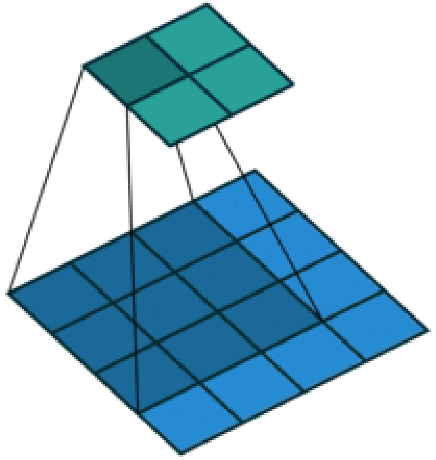


No Zero Padding → Valid: 输入和输出尺寸不同

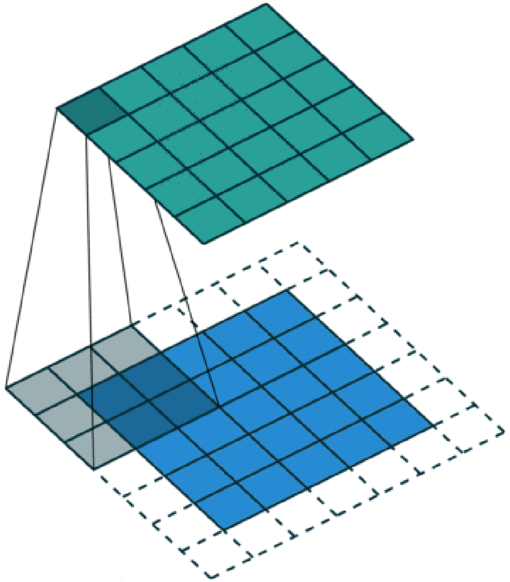
$Zero\ Padding = \frac{K-1}{2}$ → Same: 输入和输出尺寸相同

$$O = \frac{W-K+2P}{S} + 1$$

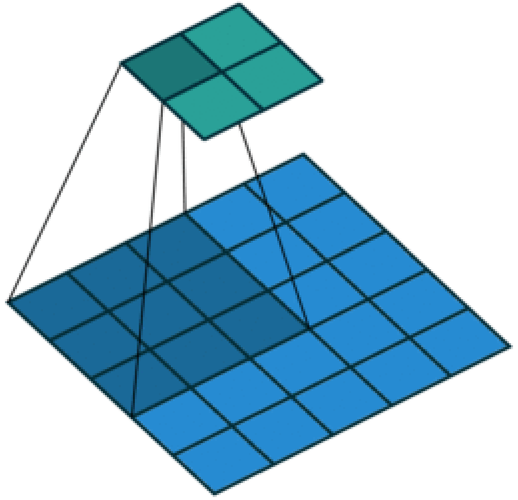
O is the output height/length, W is the input height/length, K is the filter size, P is the padding, and S is the stride



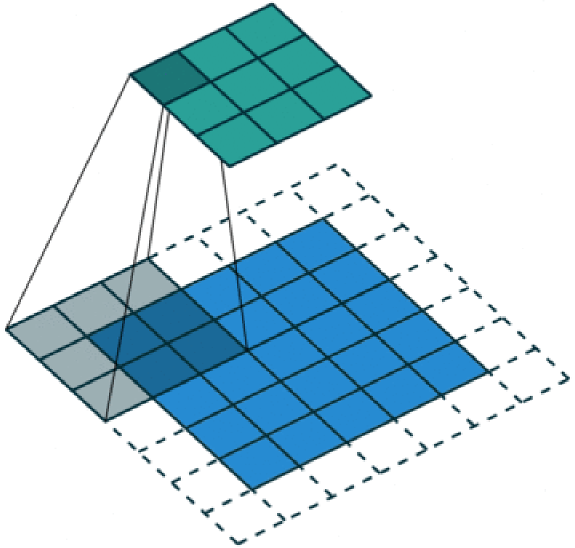
No padding, no strides



Half padding, no strides



No padding, strides



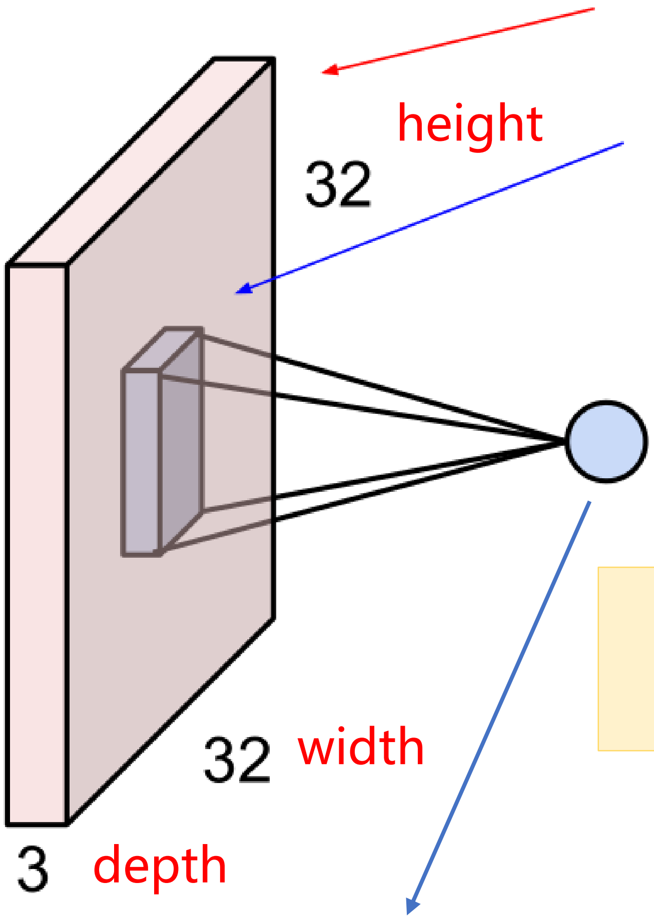
Padding, strides

卷积层特点一：局部连接

feature map
特征图

32x32x3 image

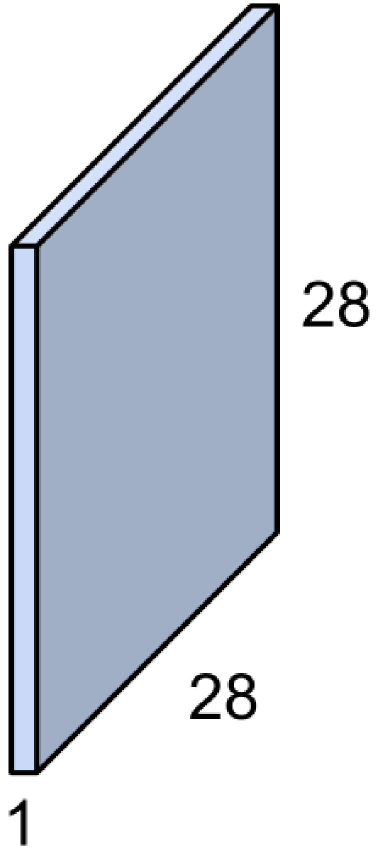
5x5x3 filter



$$\sigma(w^T x + b)$$

对所有空间位置做卷积

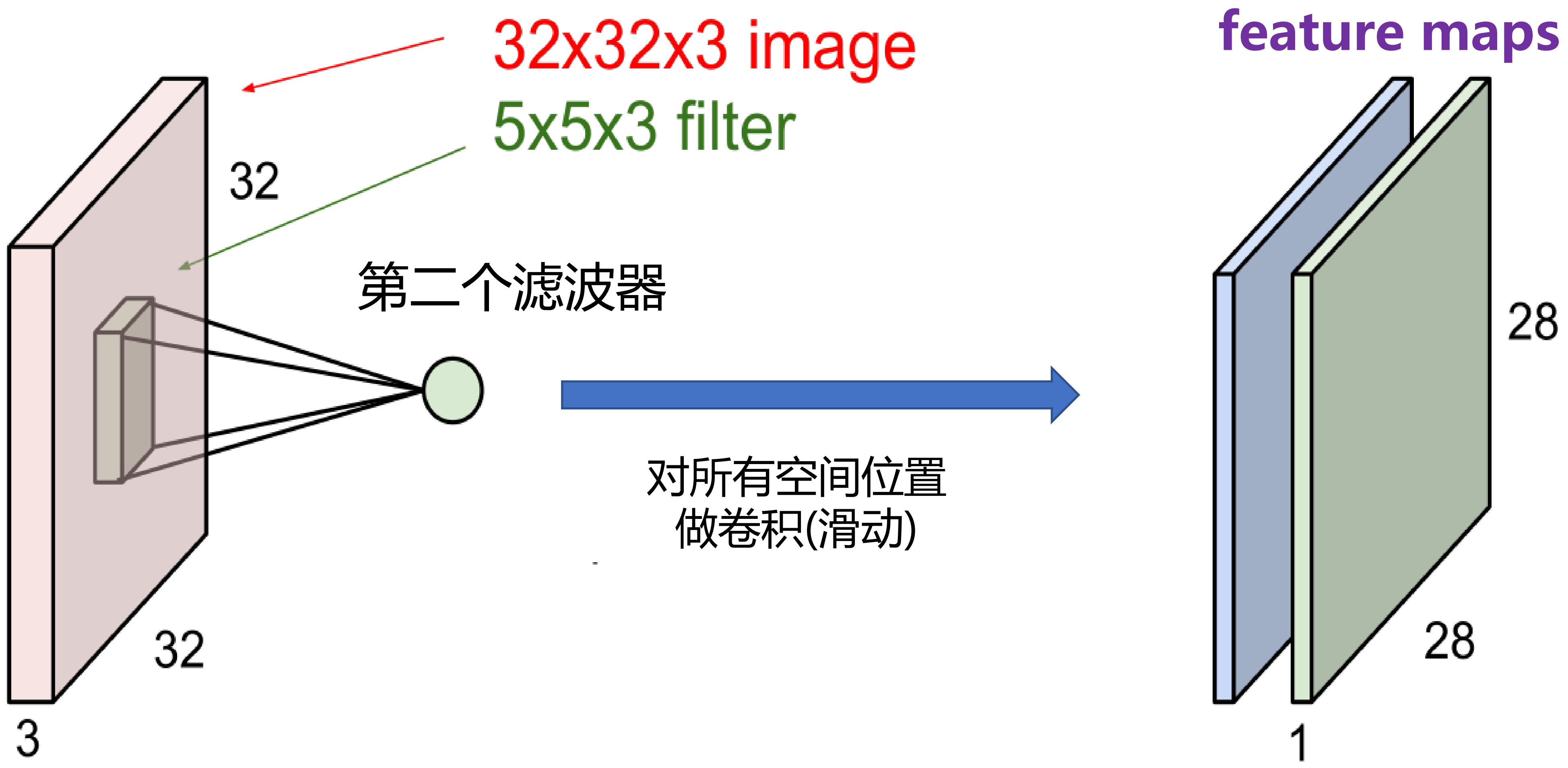
局部连接：每一个输出连接到输入图像的一个小区域



一个数值：滤波器和部分图像像素的点积
(5×5×3 = 75-维点积+bias)

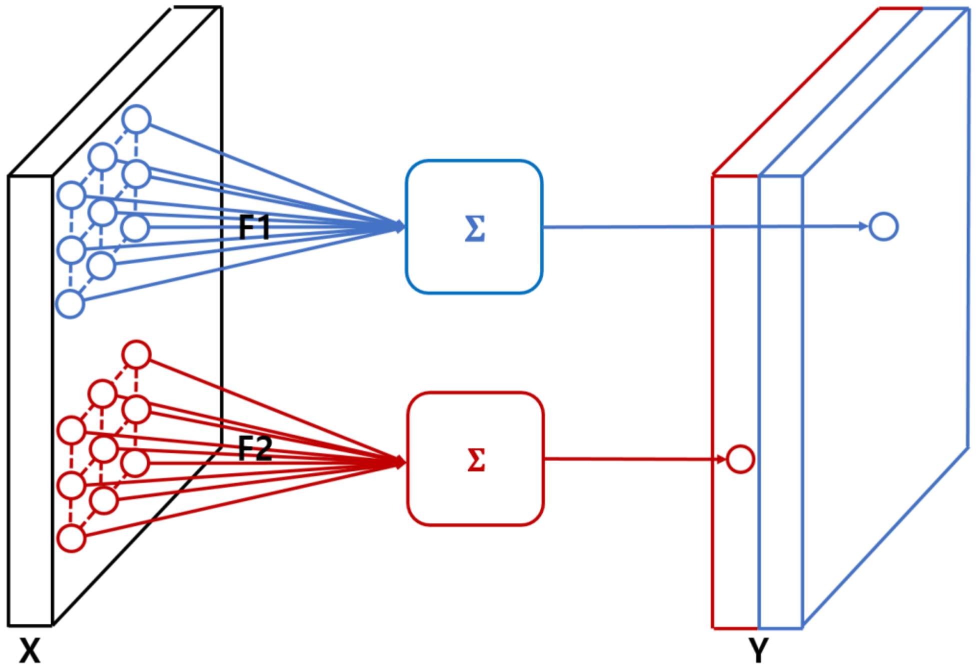
5x5 filter ⟶ 每个神经元的感受野是5x5

特征图：通过使用线性滤波器对图像进行卷积，添加偏置项并应用非线性函数来获得

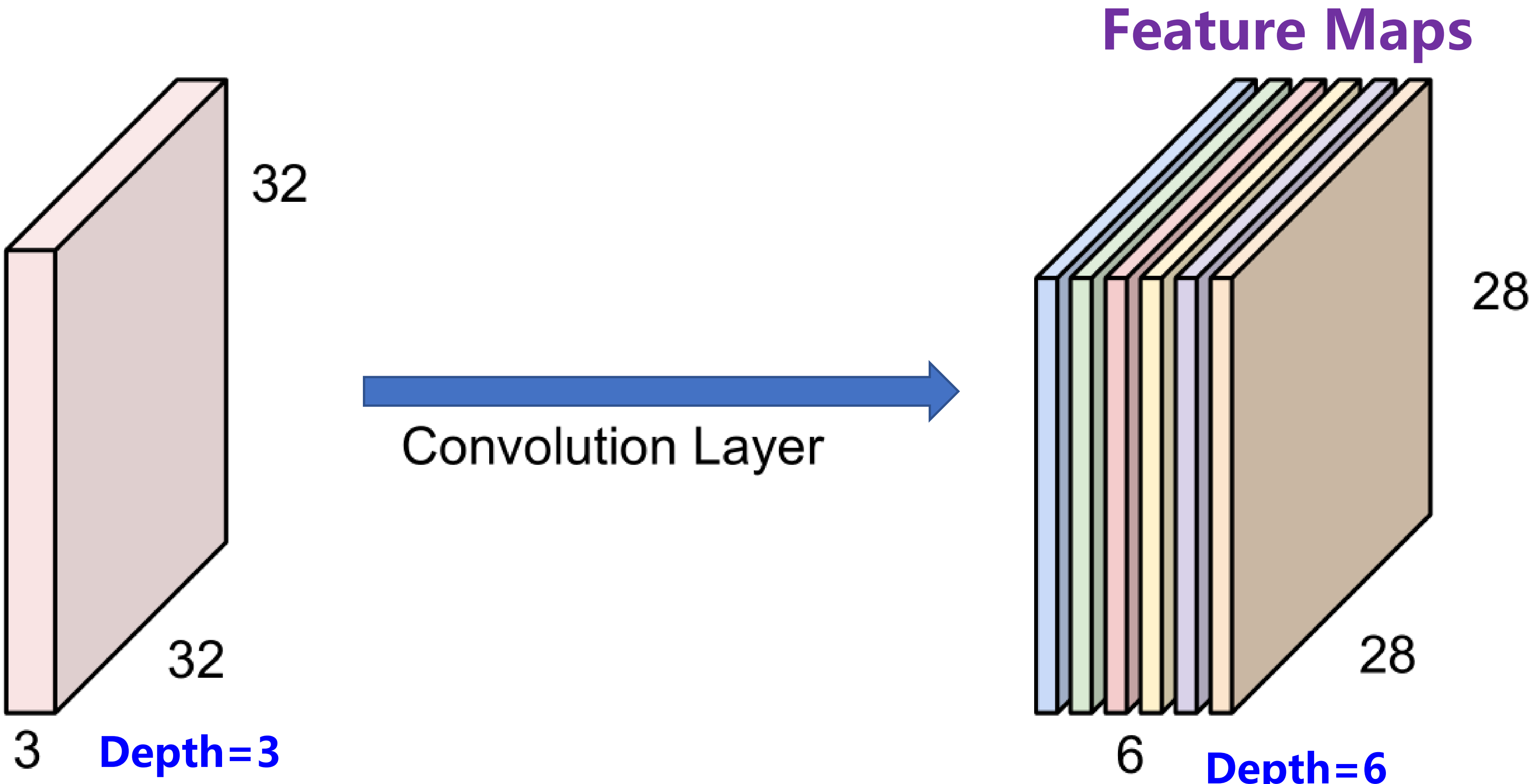


32x32x3 image
5x5x3 filter
32
32
3
第二个滤波器
对所有空间位置
做卷积(滑动)
feature maps
28
28
1

卷积层特点二：参数共享



- 卷积层在空间上共享权重；对不同位置的所有输入权重是相同的
- 导致平移不变性（translation-invariant）：平移将在目标相应平移的位置产生相同的响应。



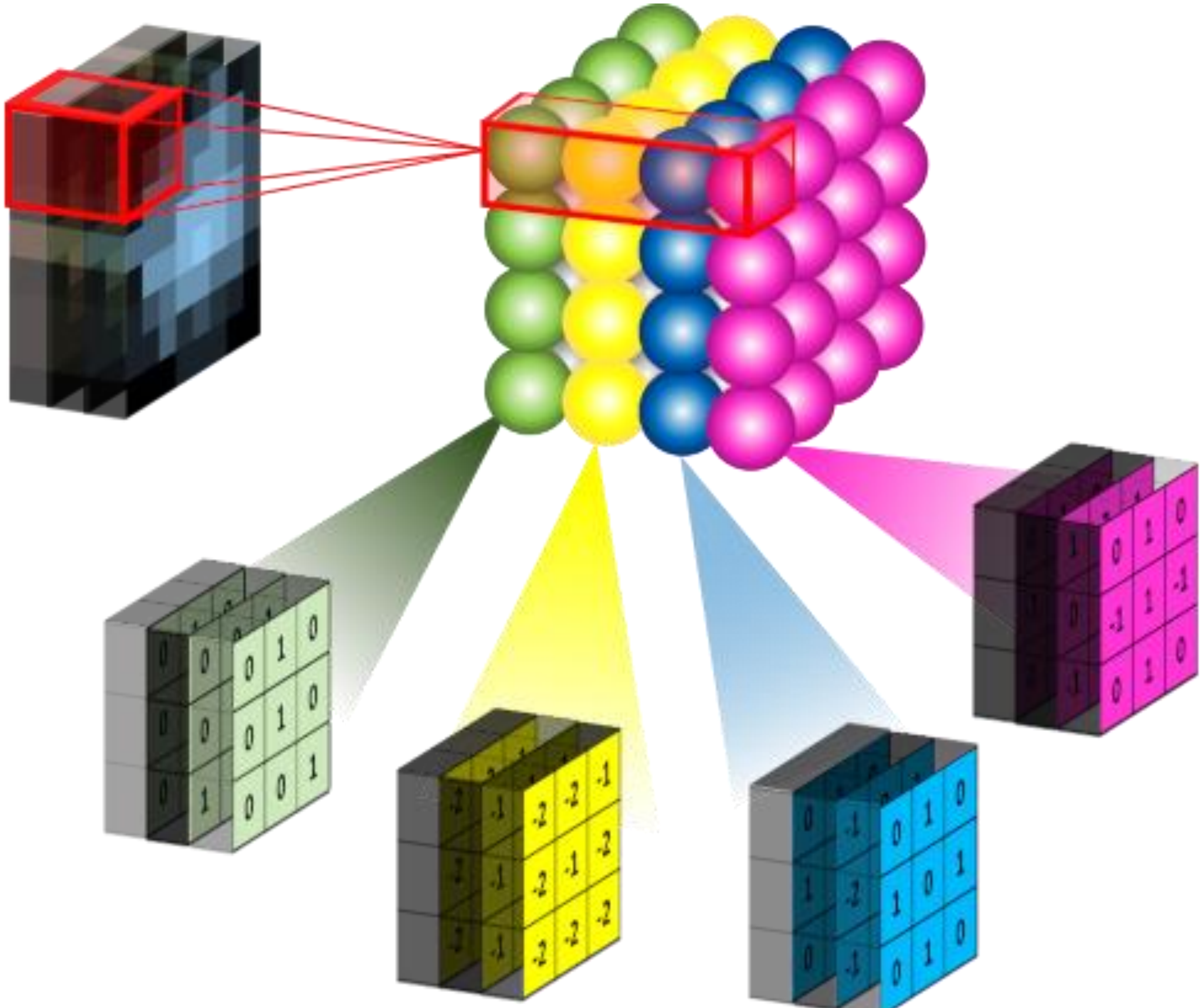
卷积运算的滤波器个数

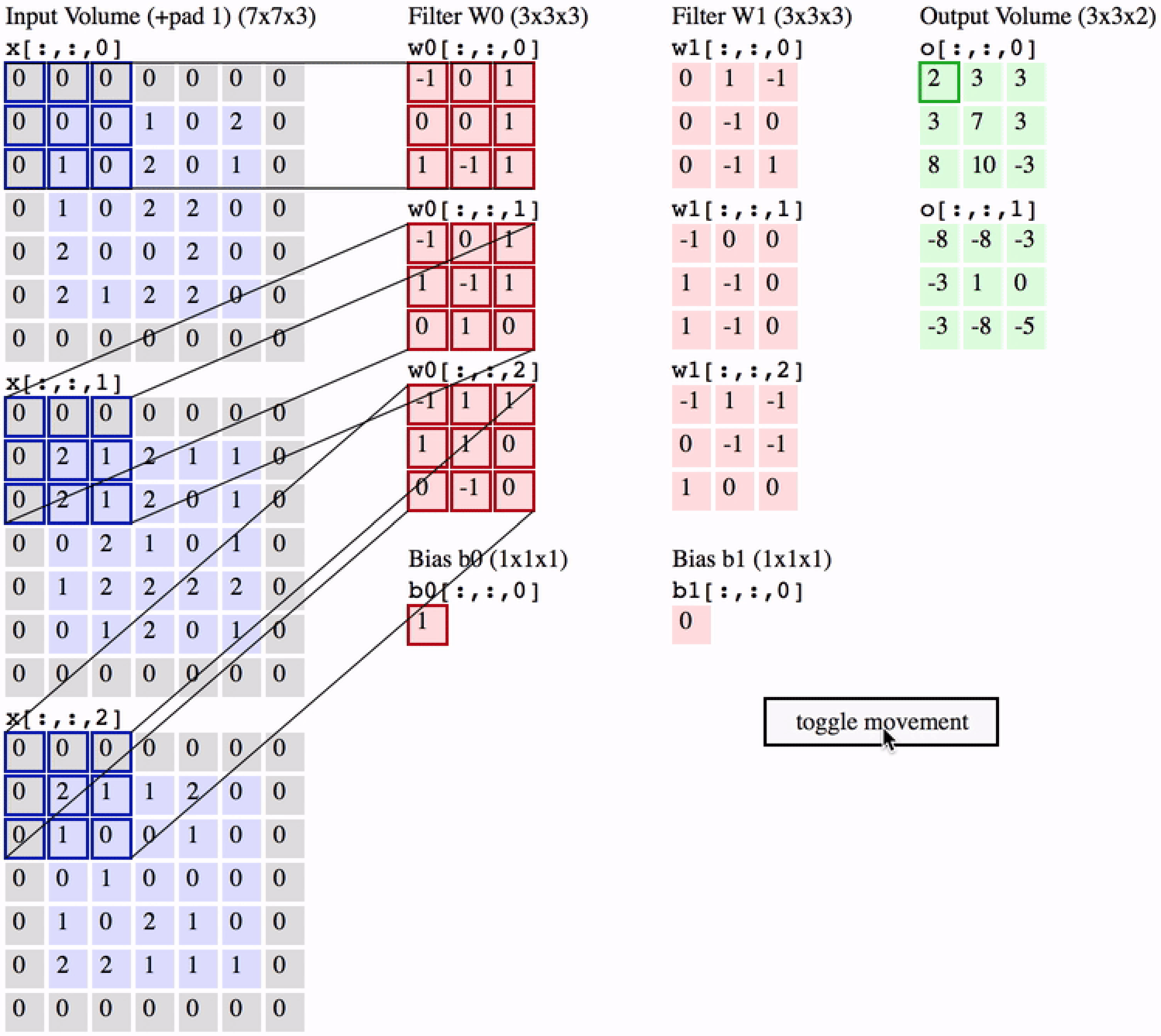
6个5x5滤波器，将得到6个单独的特征图；

堆叠后得到一个28x28x6的“新图像”

每个隐藏层就是由多个feature maps组成

多通道图像+多卷积核的卷积





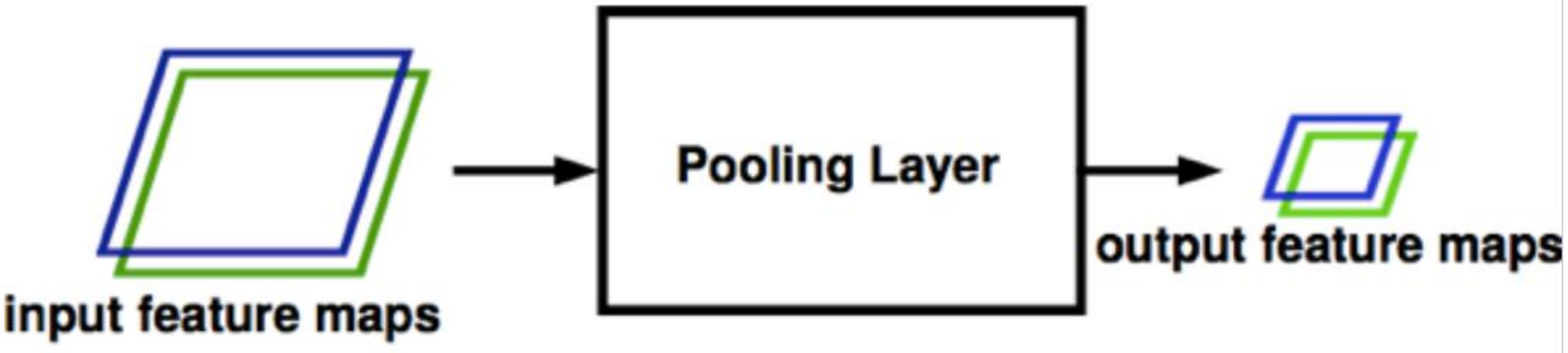
多通道图像+多卷积核的卷积

输入有3个通道，同时有2个卷积核。对于每个卷积核，先在输入3个通道分别作卷积，再将3个通道结果加起来得到卷积输出。

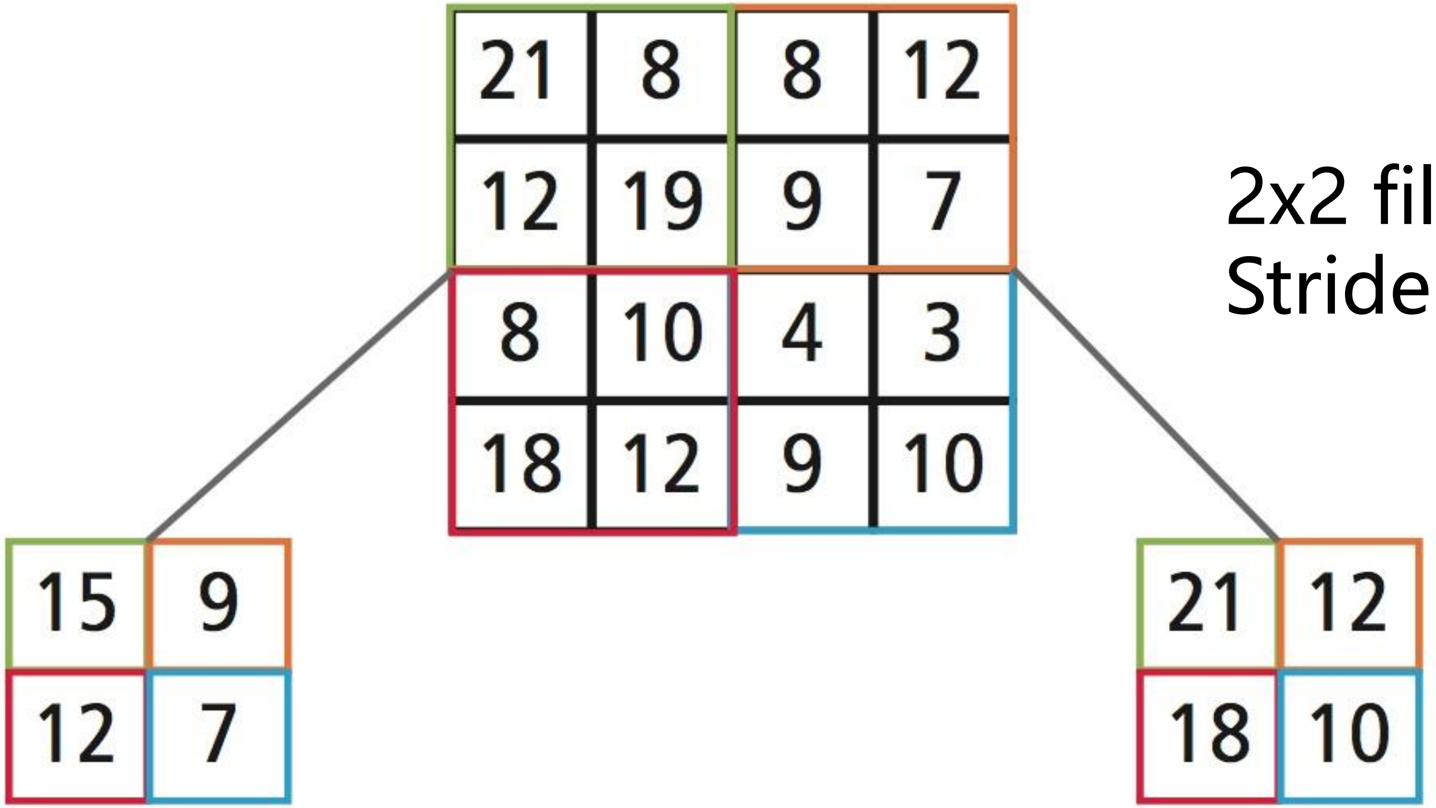
所以对于某个卷积层，无论输入图像有多少个通道，输出图像通道数总是等于卷积核数量！

池化 (Pooling)

- 非线性下采样(down-sampling)来减小特征图尺寸



池化 (Pooling)



2x2 filters
Stride=2

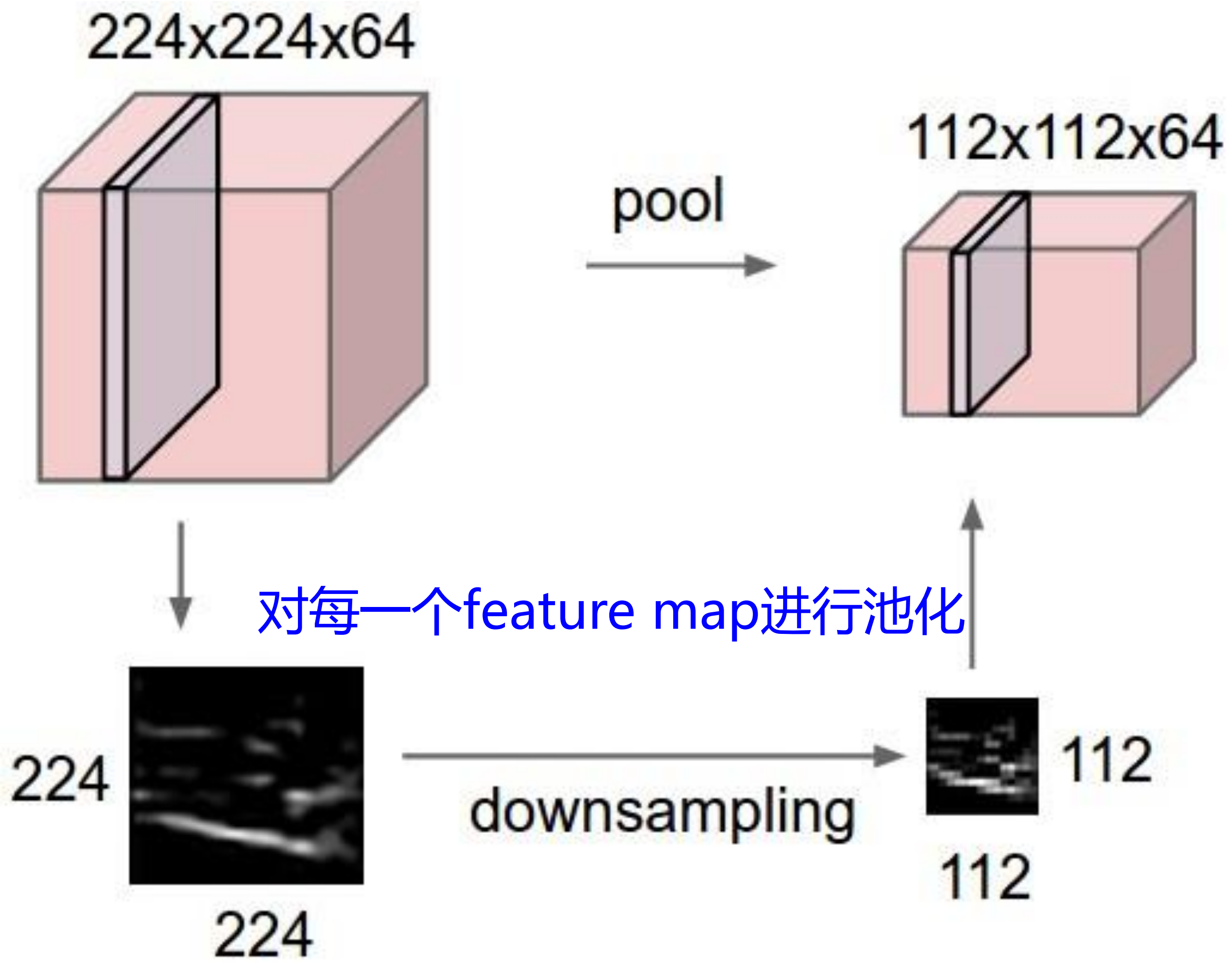
Average Pooling

平均池化

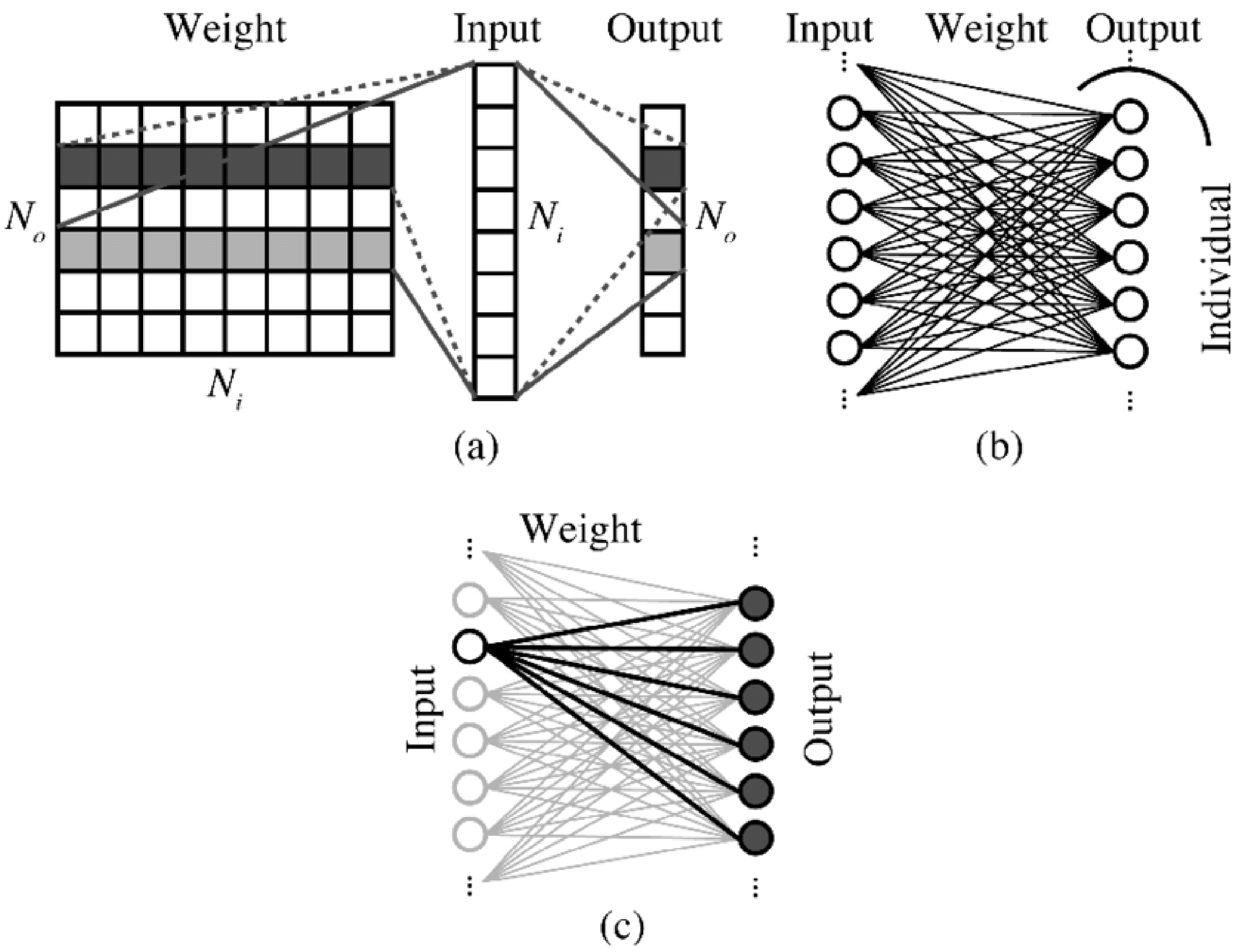
Max Pooling

最大池化

池化层


224x224x64
pool
112x112x64
对每一个feature map进行池化
224
224
downsampling
112
112

全连接层(Fully Connected Layer)



将学习到的特征表示映射到样本的标记空间

全连接层(Fully Connected Layer)

32x32x3 image → 展平成 3072 x 1