

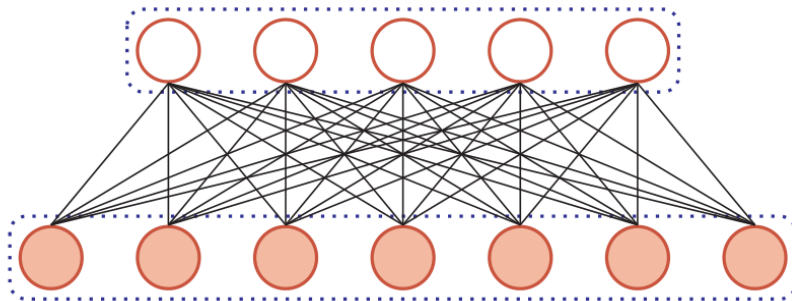
《机器学习基础》



卷积神经网络

全连接前馈神经网络

► 权重矩阵的参数非常多



► 局部不变性特征

- 自然图像中的物体都具有局部不变性特征
 - 尺度缩放、平移、旋转等操作不影响其语义信息。
- 全连接前馈网络很难提取这些局部不变特征

卷积神经网络

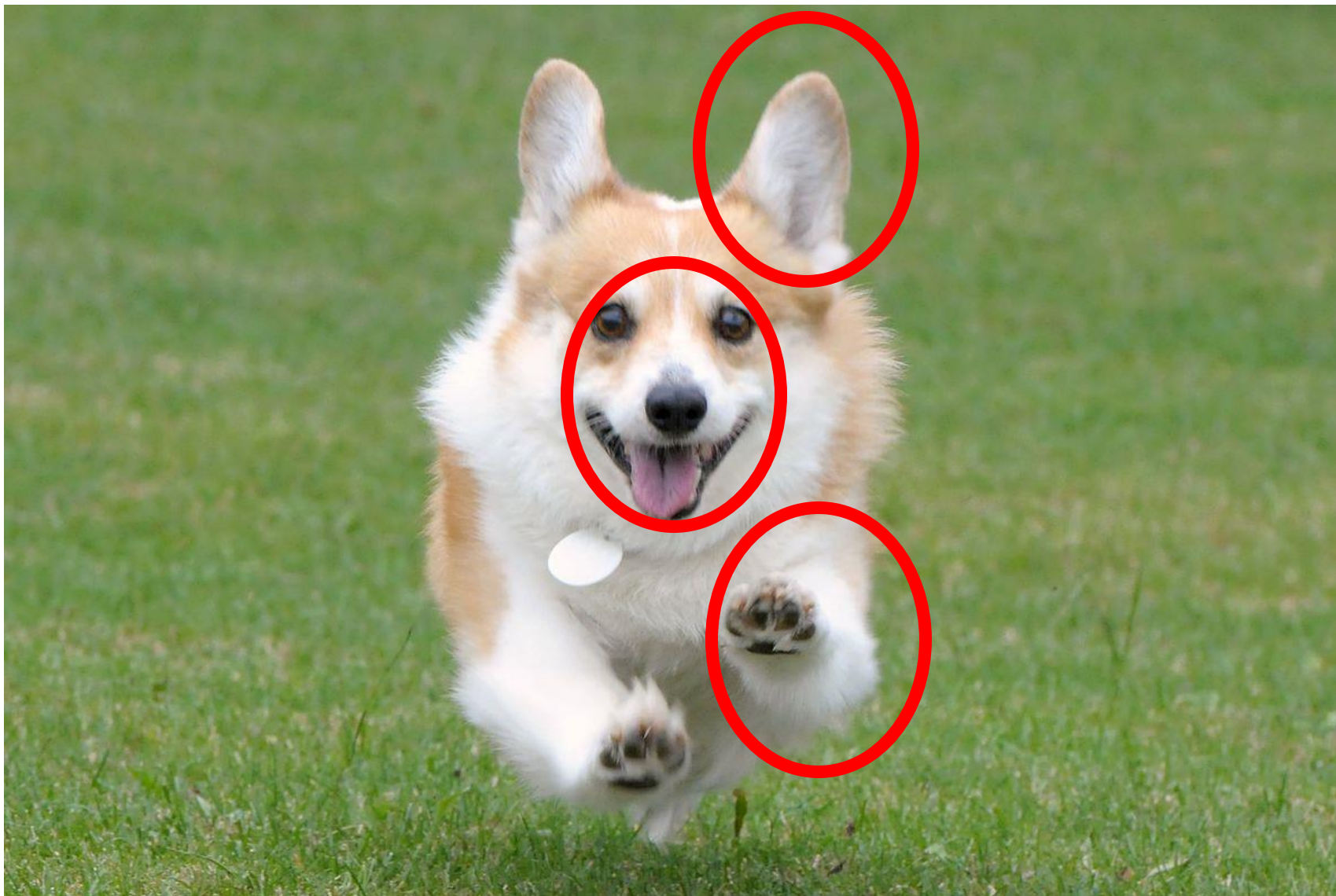
▶ 卷积神经网络 (Convolutional Neural Networks, CNN)

- ▶ 一种前馈神经网络
- ▶ 受生物学上感受野 (Receptive Field) 的机制而提出的
 - ▶ 在视觉神经系统中，一个神经元的感受野是指视网膜上的特定区域，只有这个区域内的刺激才能够激活该神经元。
- ▶ 特别适用于处理具有网格拓扑结构的数据，如图像 (2D网格) 和时间序列数据 (1D网格)
- ▶ 卷积神经网络有三个结构上的特性：
 - ▶ 局部连接
 - ▶ 权重共享
 - ▶ 空间或时间上的次采样

卷积



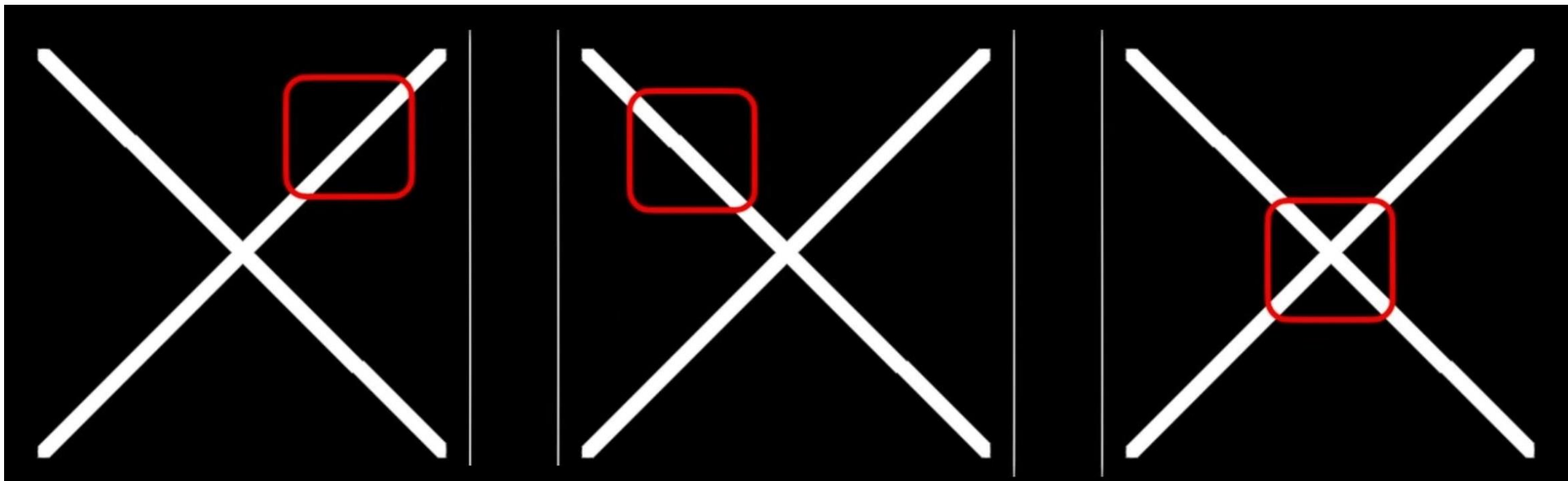
卷积



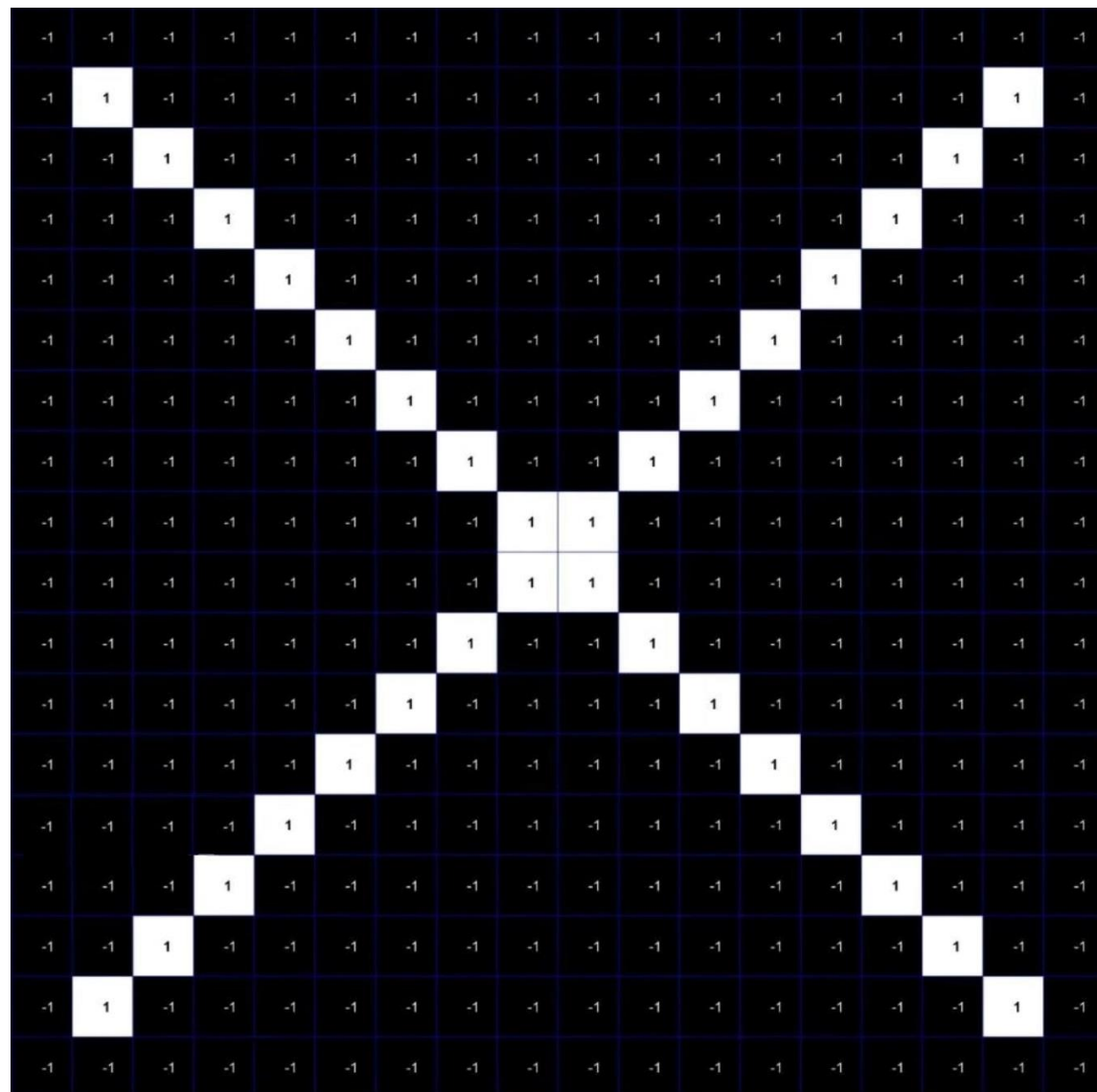
卷积



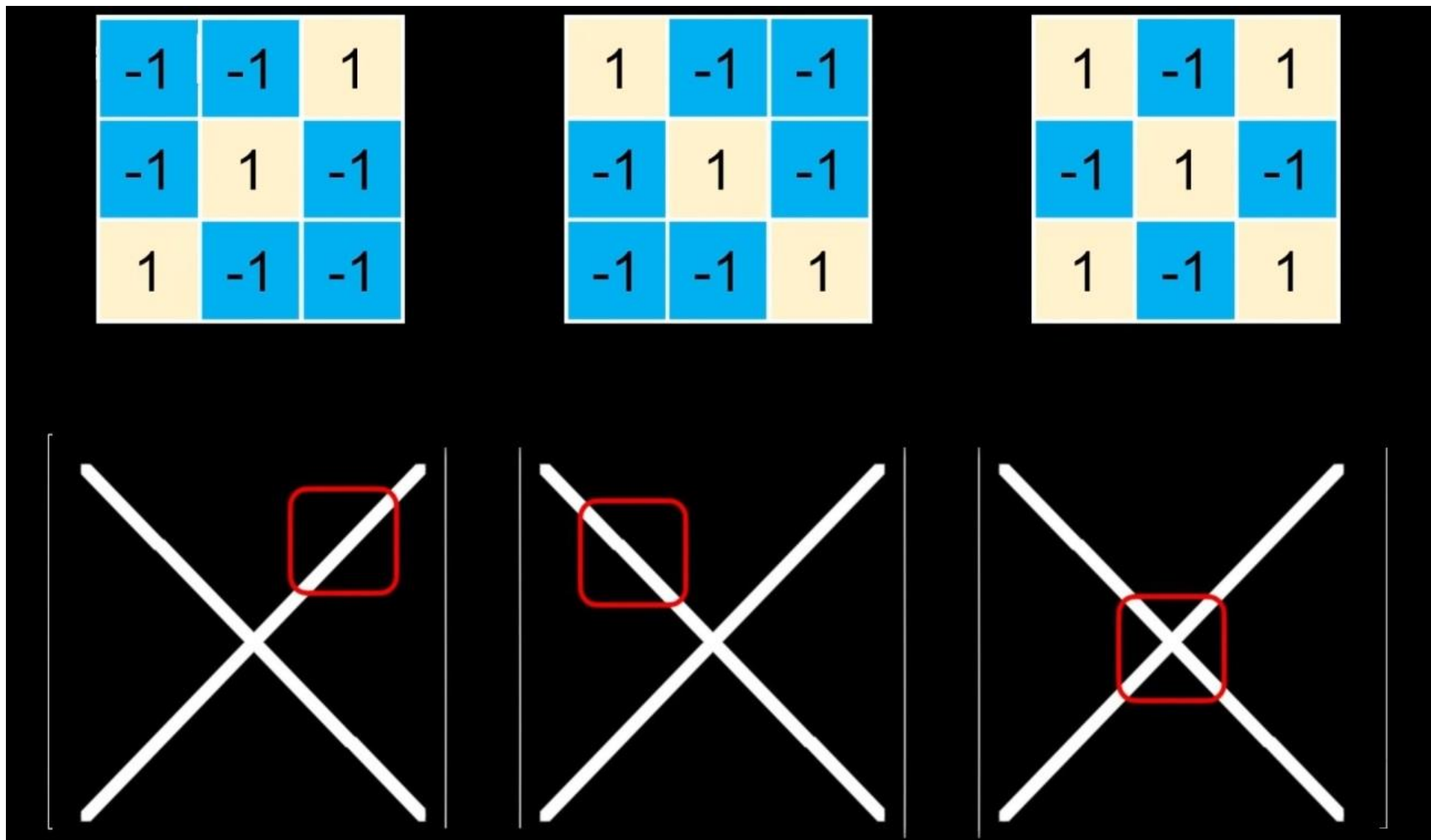
卷积



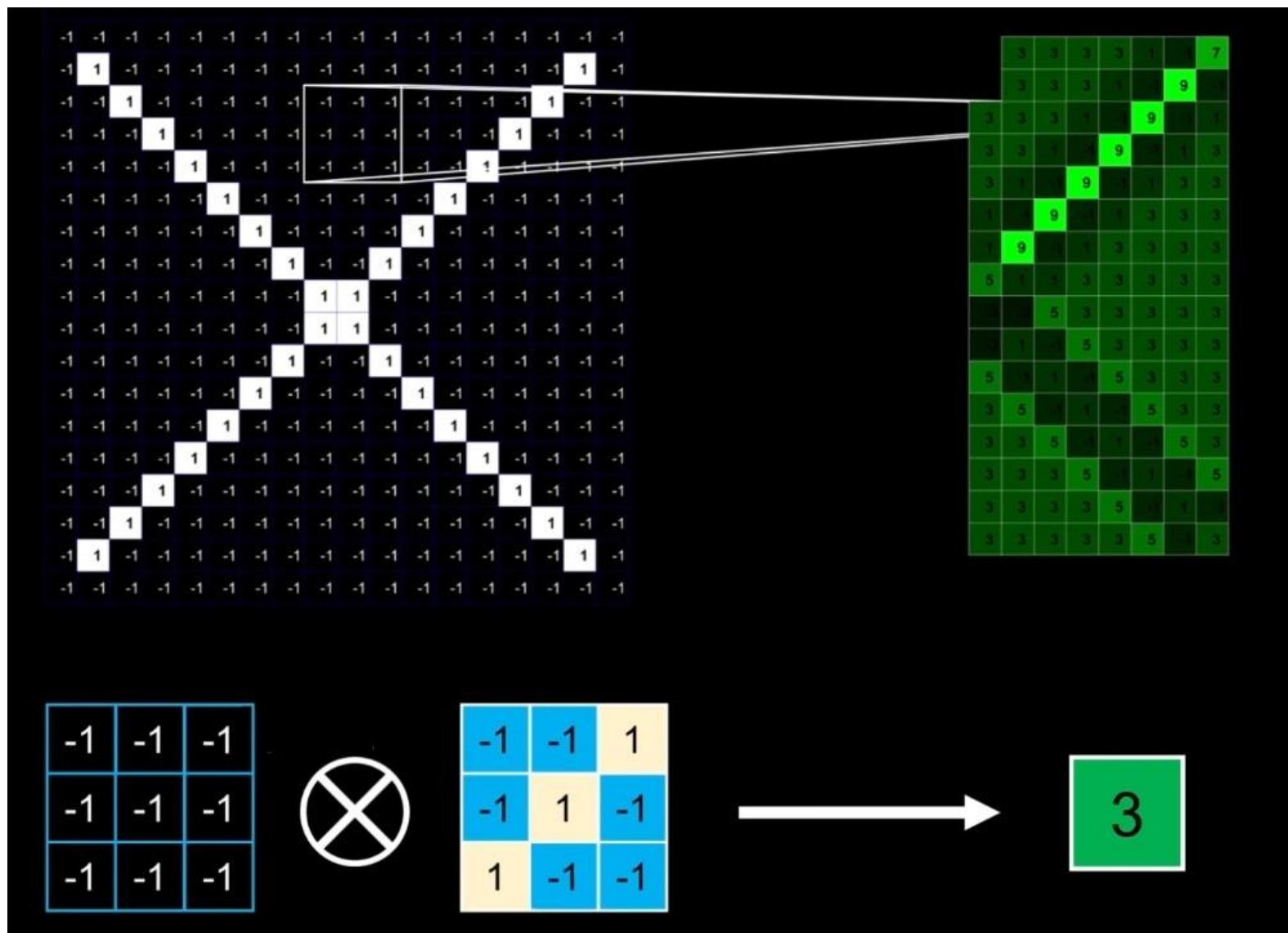
卷积



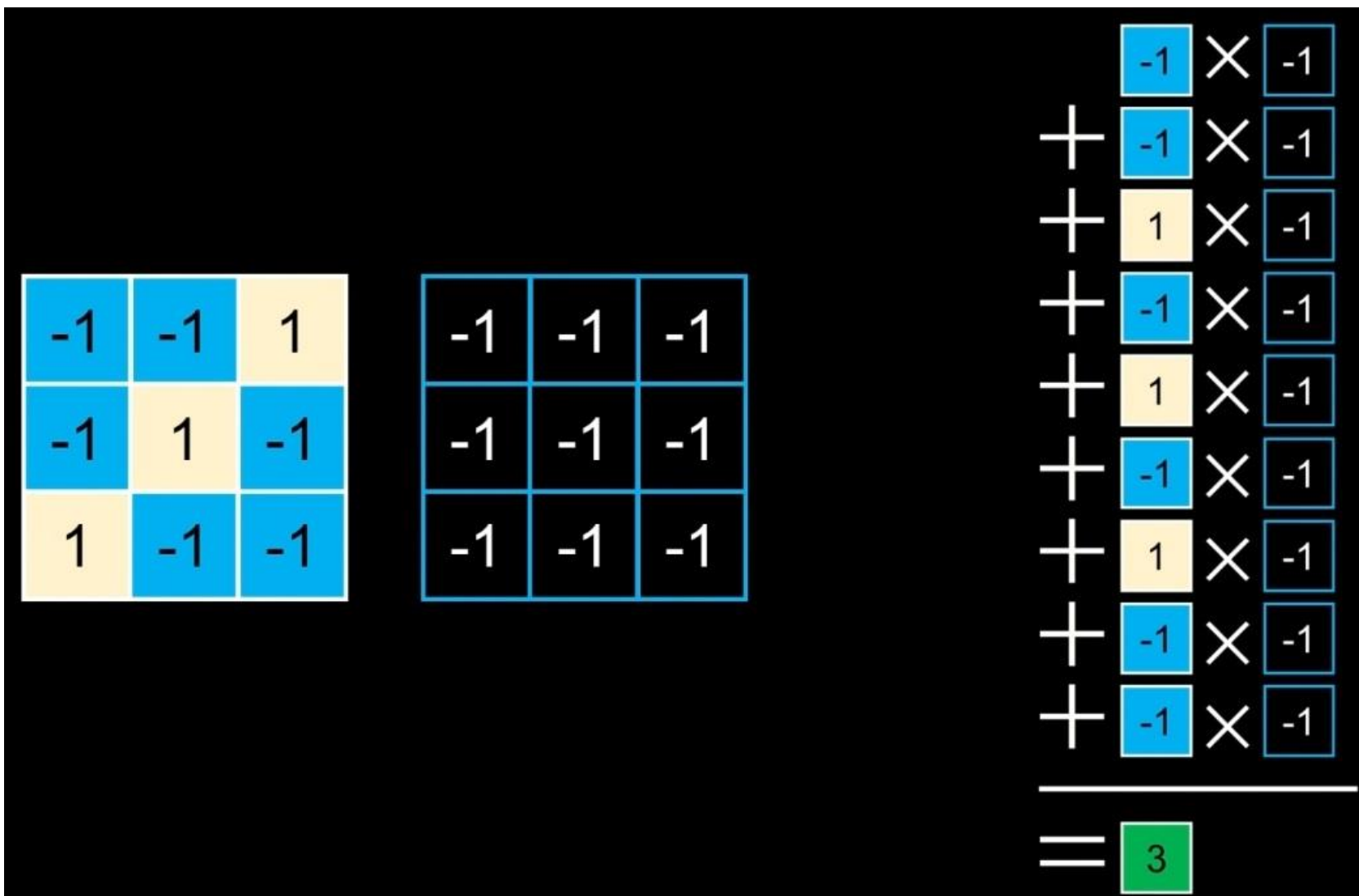
卷积核作为特征提取器



二维卷积



二维卷积



二维卷积

- 在图像处理中，图像是以二维矩阵的形式输入到神经网络中，因此我们需要二维卷积。

一个输入信息 \mathbf{X} 和滤波器 \mathbf{W} 的二维卷积定义为

$$\mathbf{Y} = \mathbf{W} * \mathbf{X},$$

$$y_{ij} = \sum_{u=1}^U \sum_{v=1}^V w_{uv} x_{i-u+1, j-v+1}.$$

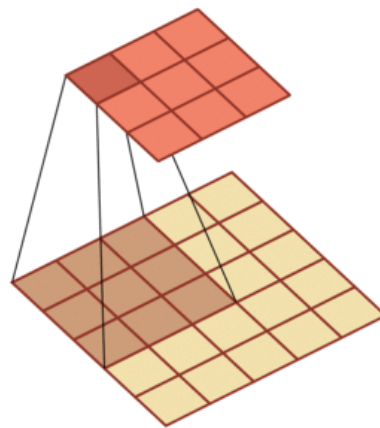
1	1	1	1	1
-1	0	-3	0	1
2	1	1	-1	0
0	-1	1	2	1
1	2	1	1	1

$*$

1	0	0
0	0	0
0	0	-1

$=$

0	-2	-1
2	2	4
-1	0	0

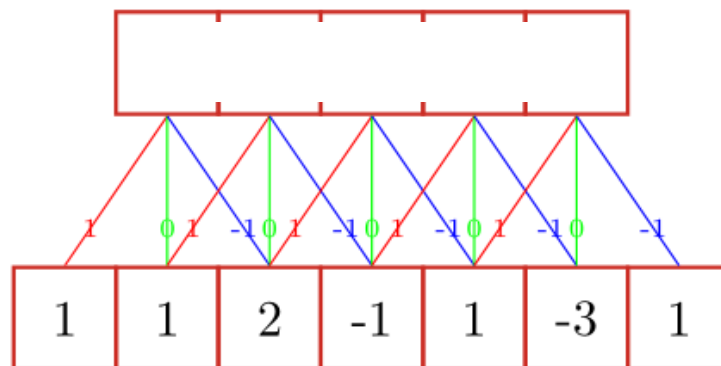


一维卷积

► 给定一个输入信号序列 x 和滤波器 w ,卷积的输出为:

$$y_t = \sum_{k=1}^K w_k x_{t-k+1}$$

Filter: [-1,0,1]

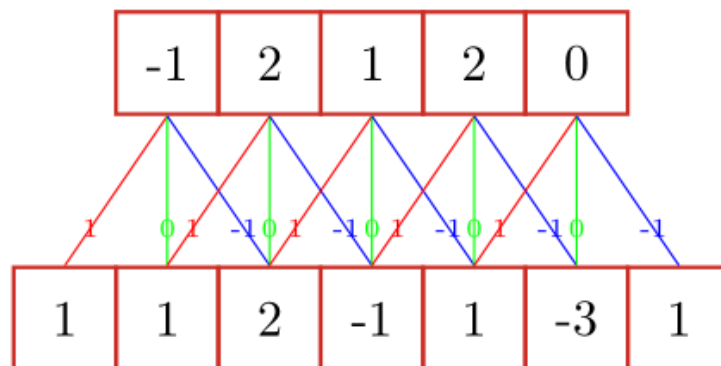


一维卷积

► 给定一个输入信号序列 x 和滤波器 w ,卷积的输出为:

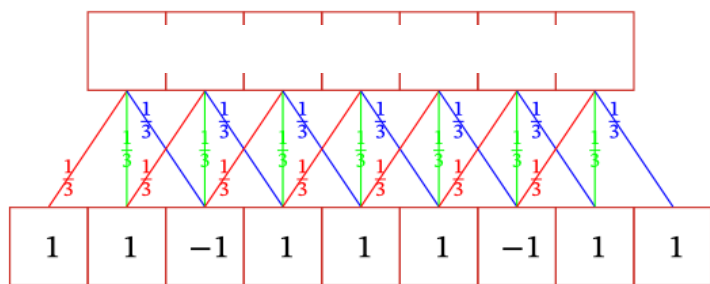
$$y_t = \sum_{k=1}^K w_k x_{t-k+1}$$

Filter: [-1,0,1]



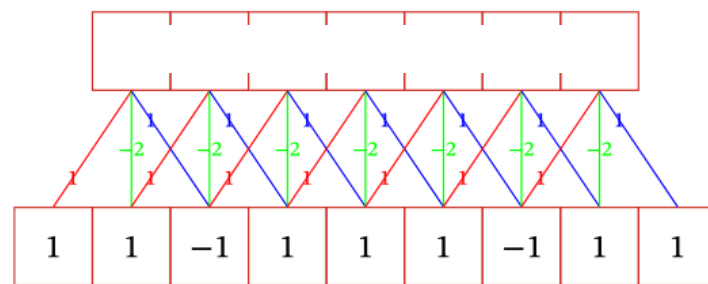
一维卷积

► 不同的滤波器来提取信号序列中的不同特征



(a) 滤波器 $[1/3, 1/3, 1/3]$

低频信息



(b) 滤波器 $[1, -2, 1]$

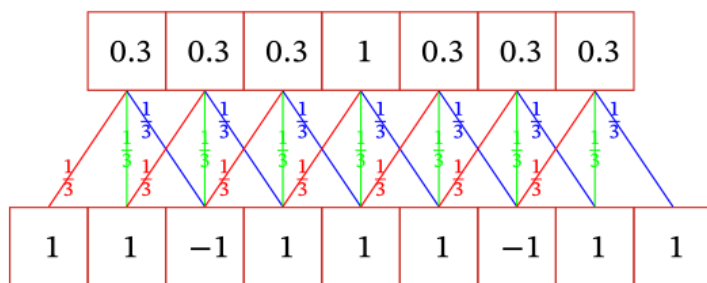
高频信息

$$y''(u) = y(u + 1) + y(u - 1) - 2y(u)$$

二阶微分

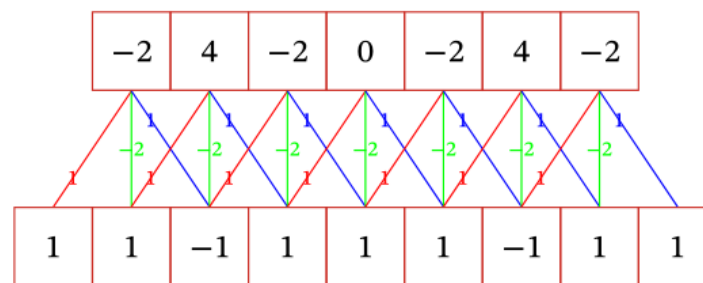
一维卷积

► 不同的滤波器来提取信号序列中的不同特征



(a) 滤波器 $[1/3, 1/3, 1/3]$

低频信息



(b) 滤波器 $[1, -2, 1]$

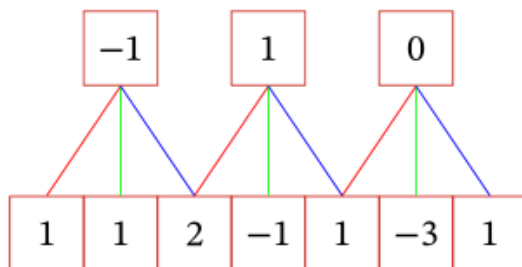
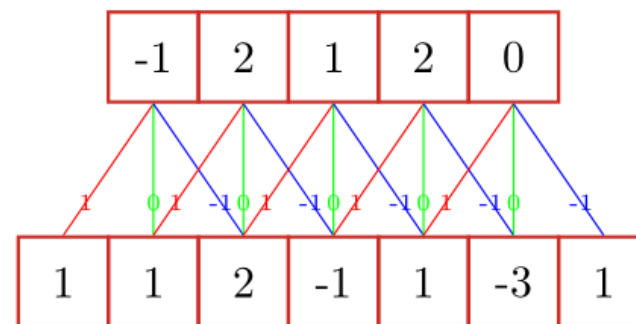
高频信息

$$y''(u) = y(u + 1) + y(u - 1) - 2y(u)$$

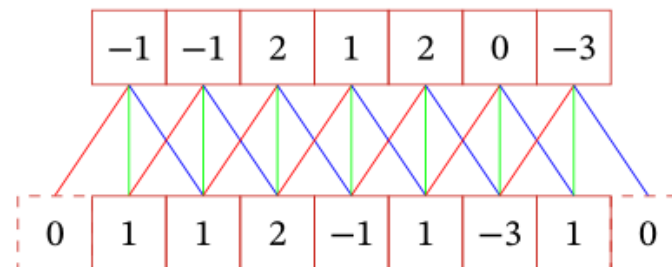
二阶微分

卷积扩展

► 引入滤波器的滑动步长 S 和零填充 P



(a) 步长 $S = 2$



(b) 零填充 $P = 1$

卷积类型

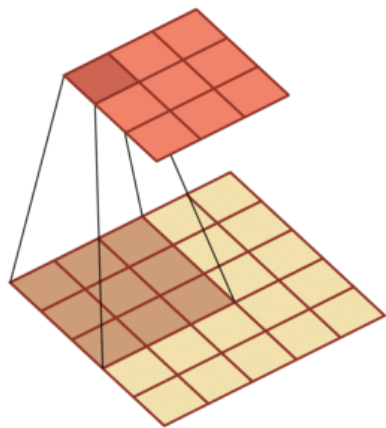
► 卷积的结果按输出长度不同可以分为三类：

► 窄卷积：步长 $T = 1$ ，两端不补零 $P = 0$ ，卷积后输出长度为 $M - K + 1$

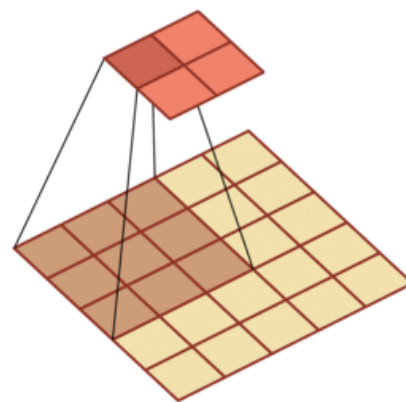
► 宽卷积：步长 $T = 1$ ，两端补零 $P = K - 1$ ，卷积后输出长度 $M + K - 1$

► 等宽卷积：步长 $T = 1$ ，两端补零 $P = (K - 1)/2$ ，卷积后输出长度 M

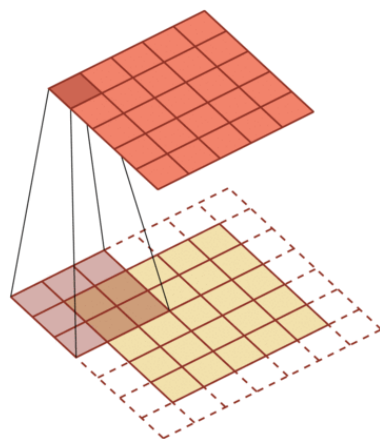
二维卷积



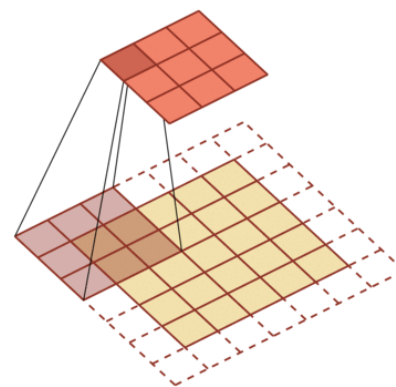
步长1，零填充0



步长2，零填充0

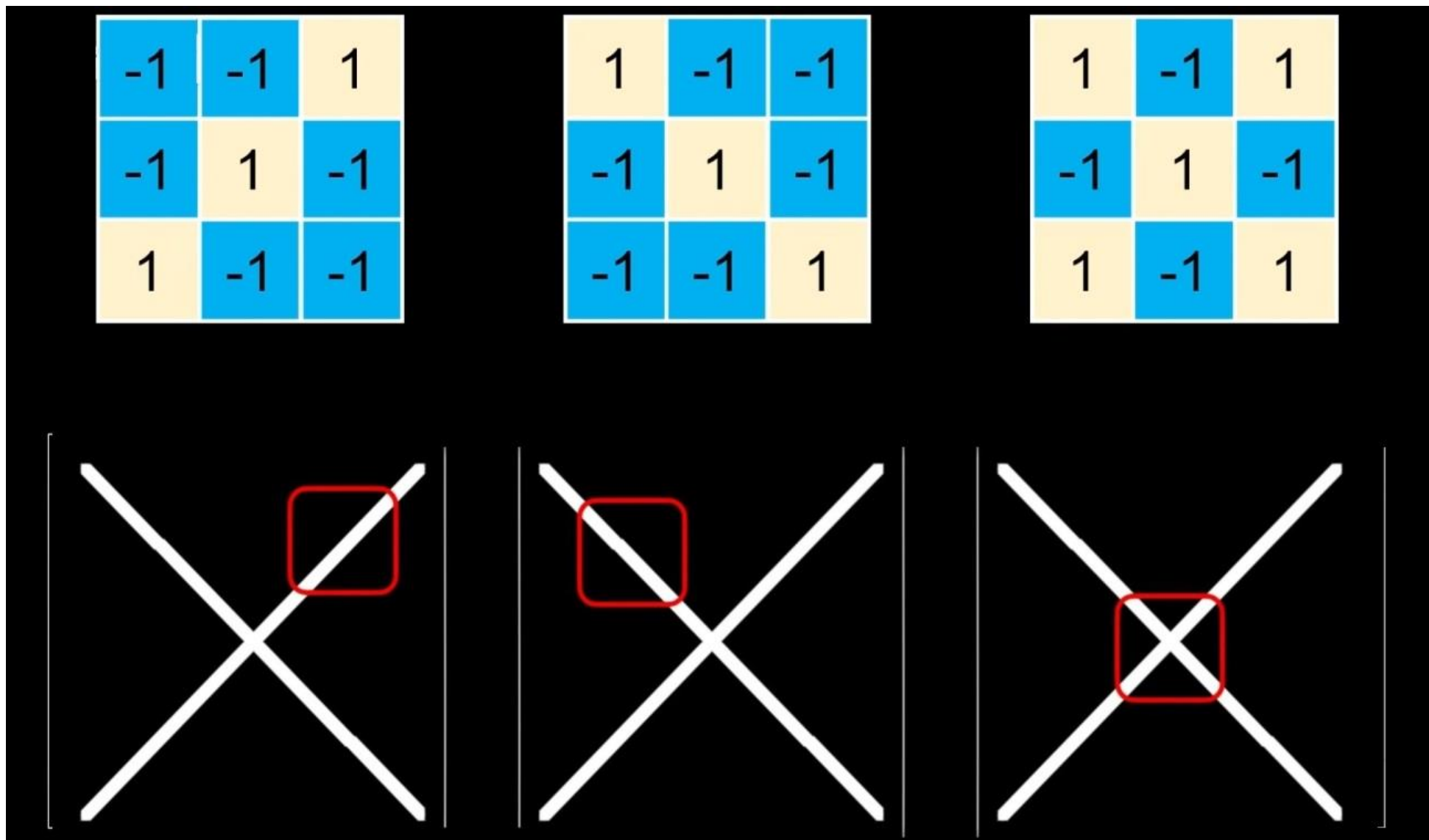


步长1，零填充1

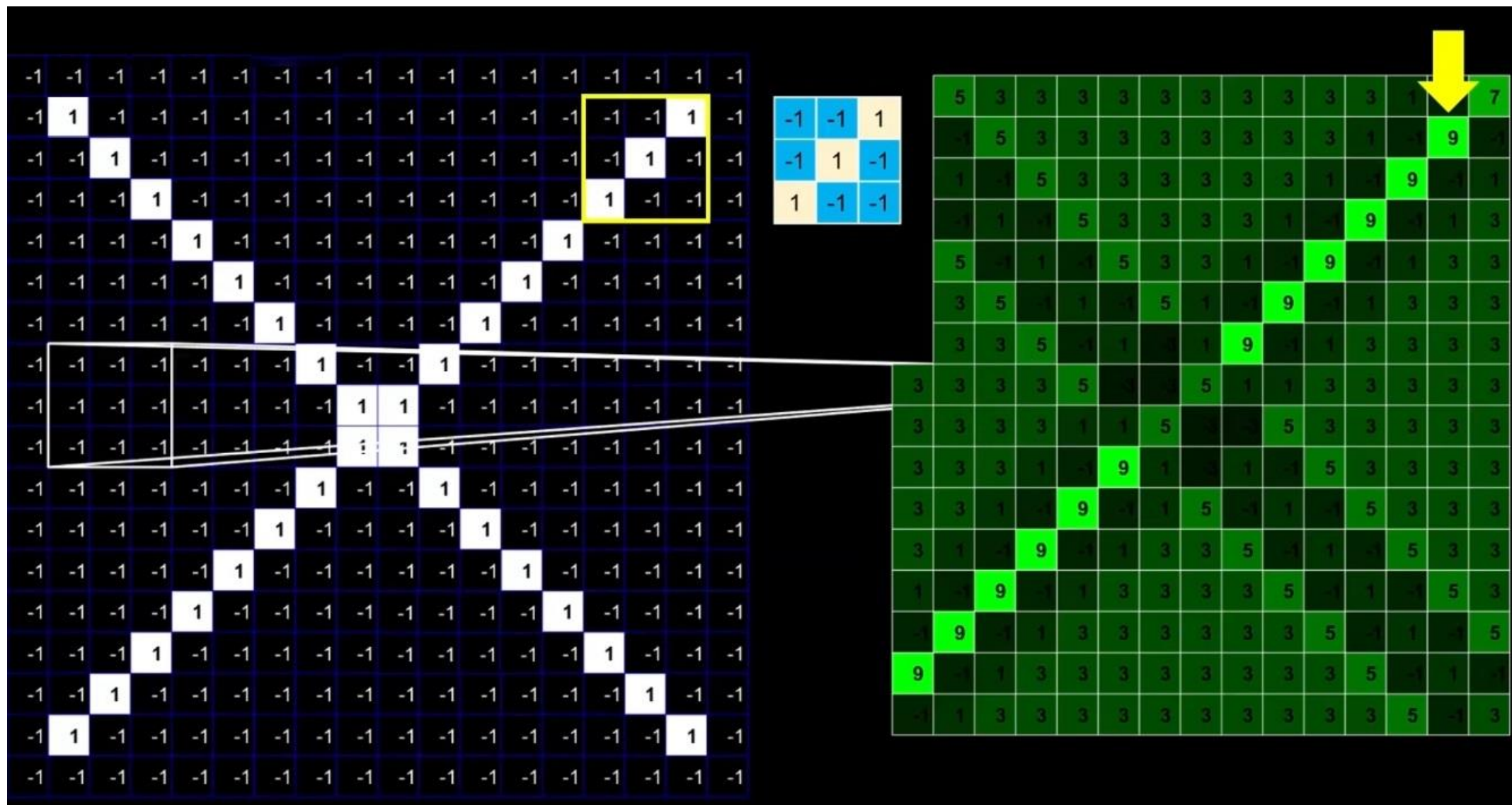


步长2，零填充1

卷积核作为特征提取器

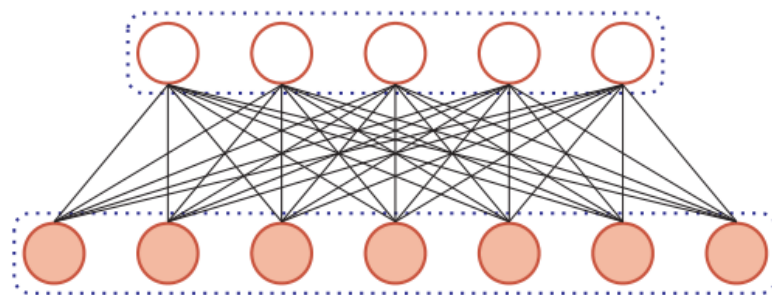


卷积核作为特征提取器

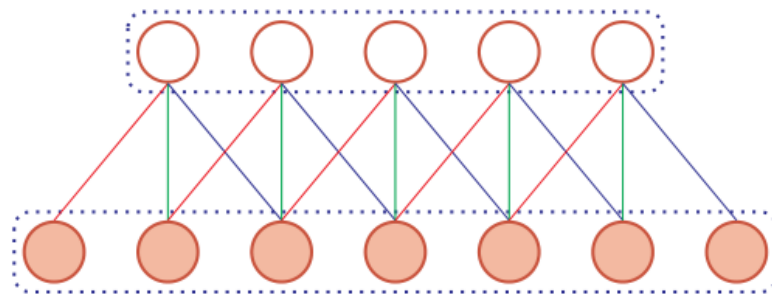


卷积神经网络

► 用卷积层代替全连接层



(a) 全连接层



(b) 卷积层

多个卷积核

▶ **特征映射 (Feature Map)** : 图像经过卷积后得到的特征。

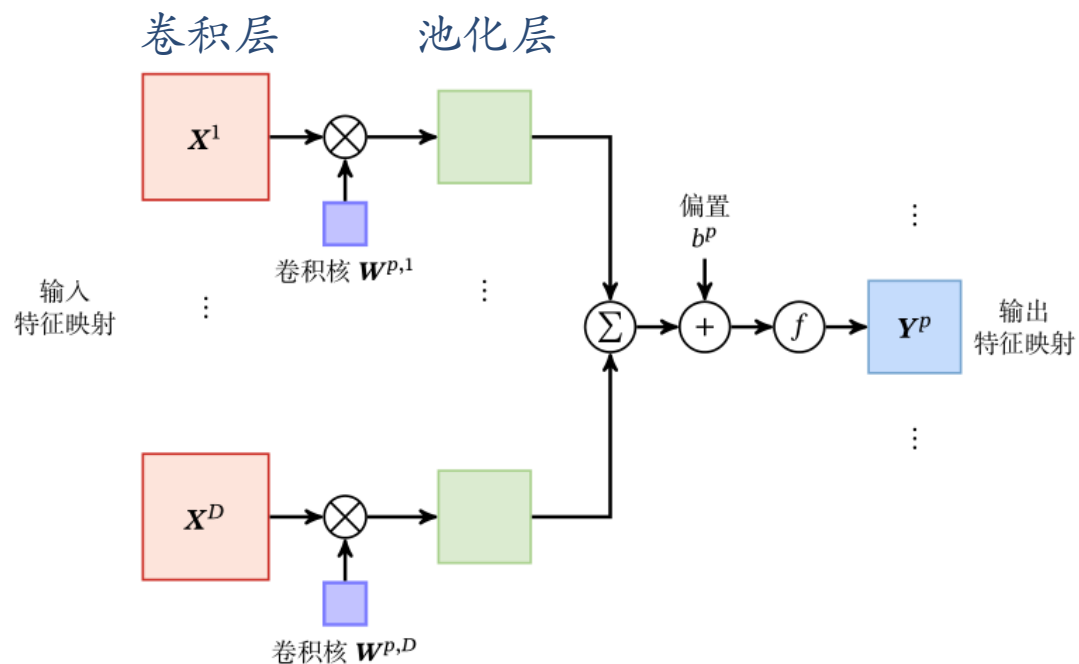
▶ 卷积核看成一个特征提取器

▶ **卷积层**

▶ 输入: D 个特征映射 $M \times N \times D$

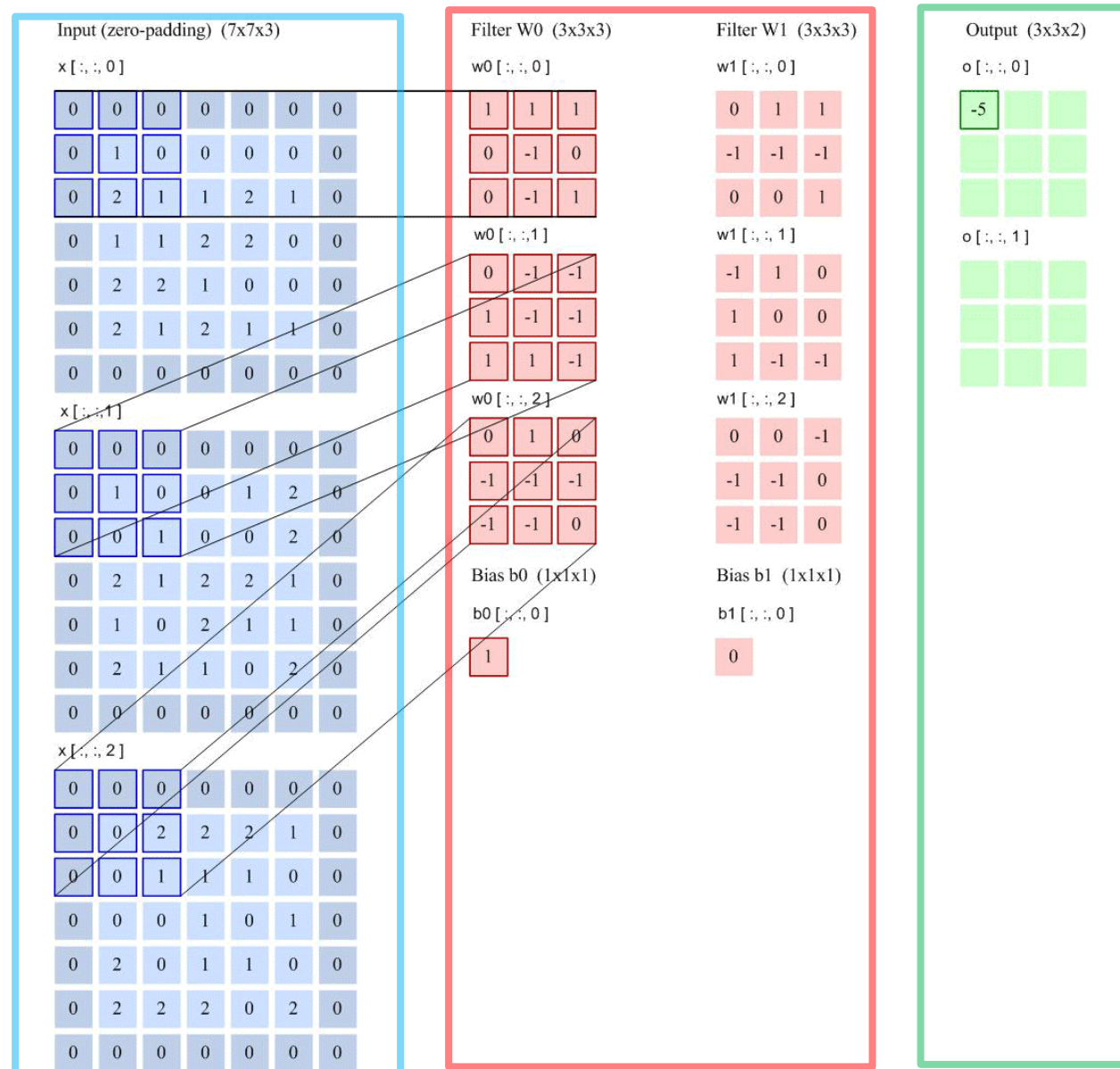
▶ 输出: P 个特征映射 $M' \times N' \times P$

卷积层的映射关系

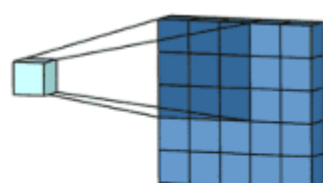
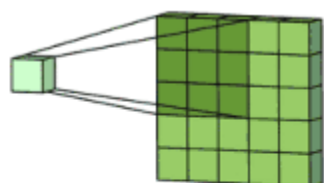
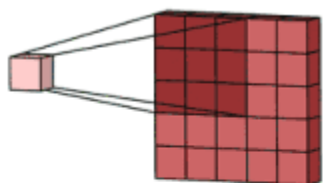


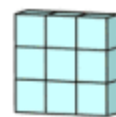
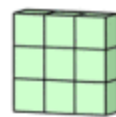
$$Z^p = W^p \otimes X + b^p = \sum_{d=1}^D W^{p,d} \otimes X^d + b^p,$$

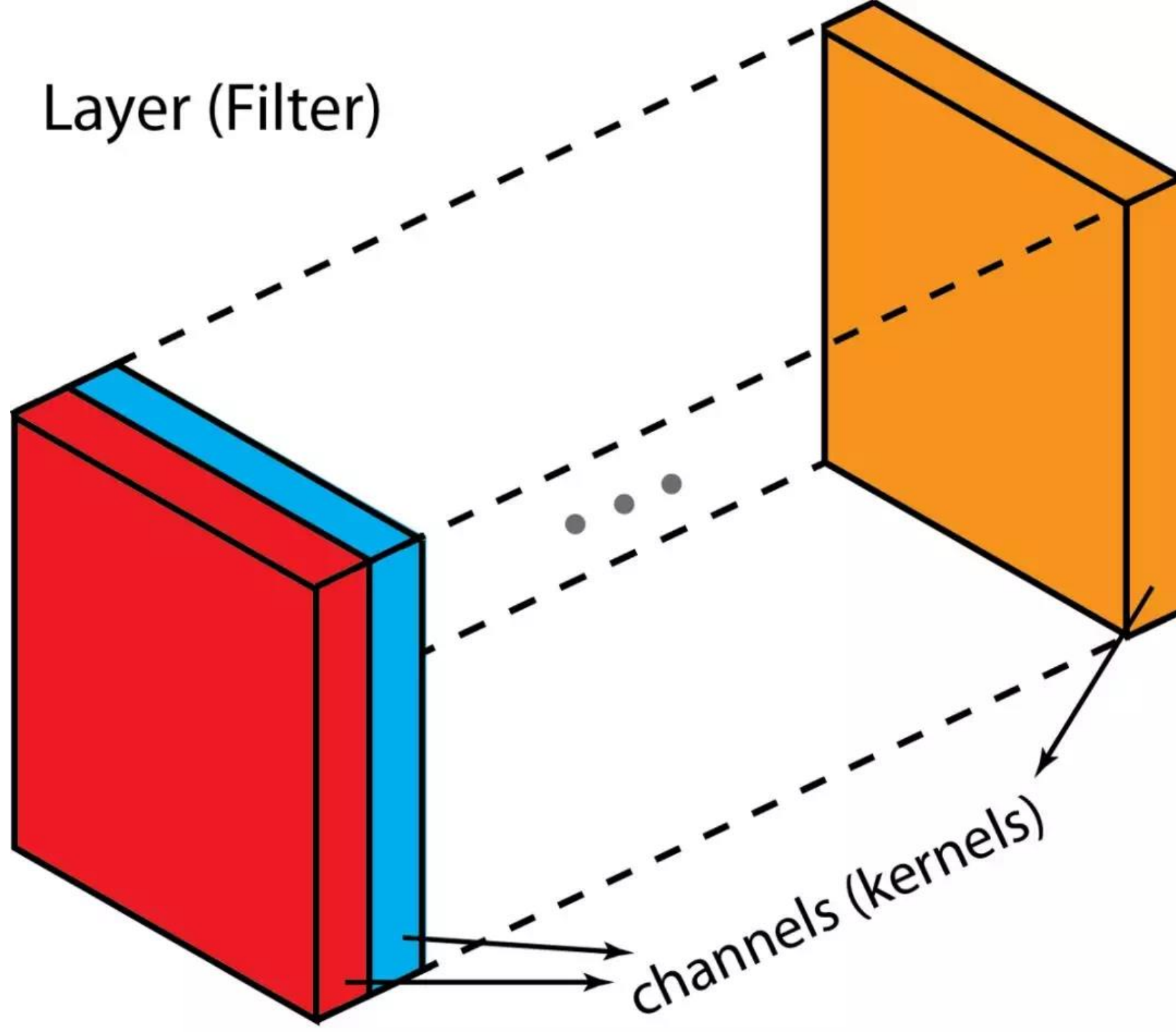
$$Y^p = f(Z^p).$$



步长2
filter 3*3
filter个数6
零填充 1

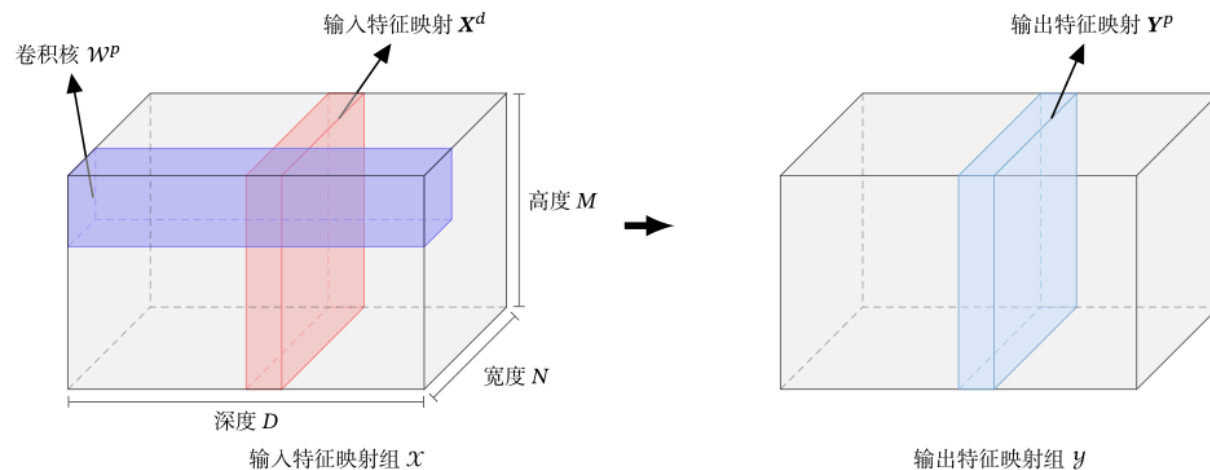






卷积层

► 典型的卷积层为3维结构

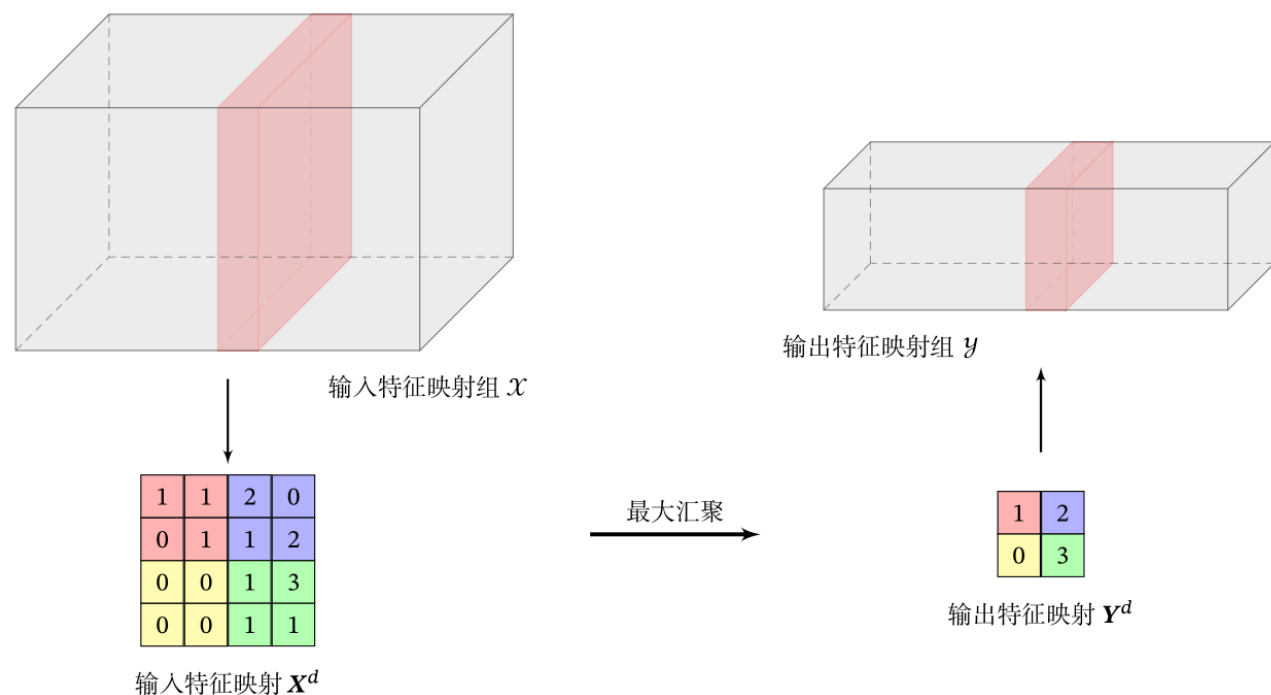


$$Z^p = W^p \otimes X + b^p = \sum_{d=1}^D W^{p,d} \otimes X^d + b^p,$$

$$Y^p = f(Z^p).$$

池化层/汇聚层

- 卷积层虽然可以显著减少连接的个数，但是每一个特征映射的神经元个数并没有显著减少。



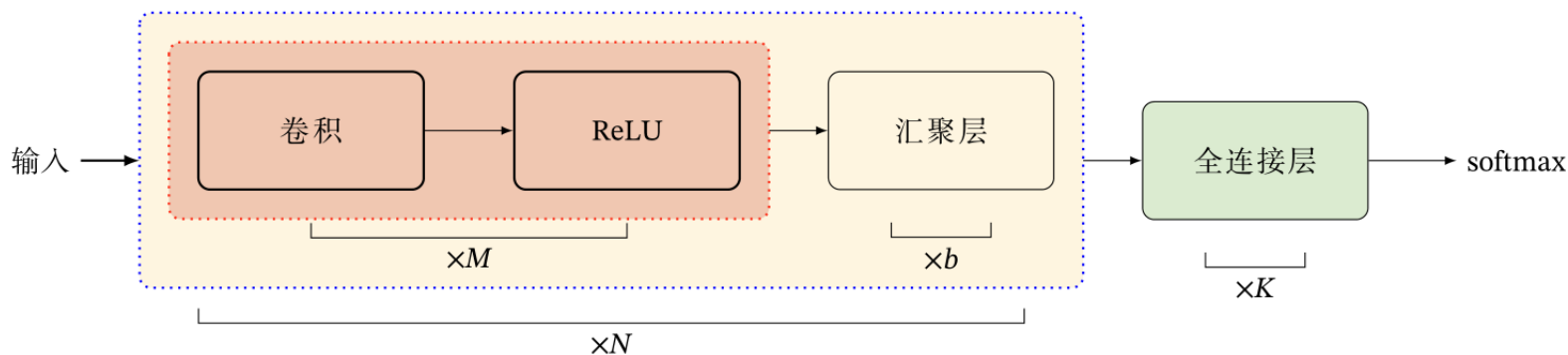
卷积网络结构

► 卷积网络是由卷积层、汇聚层(池化层)、全连接层交叉堆叠而成。

► 趋向于小卷积、大深度

► 趋向于全卷积

► 典型结构



► 一个卷积块为连续 M 个卷积层和 b 个汇聚层 (M 通常设置为 $2 \sim 5$, b 为 0 或 1)。一个卷积网络中可以堆叠 N 个连续的卷积块, 然后在接着 K 个全连接层 (N 的取值区间比较大, 比如 $1 \sim 100$ 或者更大; K 一般为 $0 \sim 2$)。

卷积神经网络

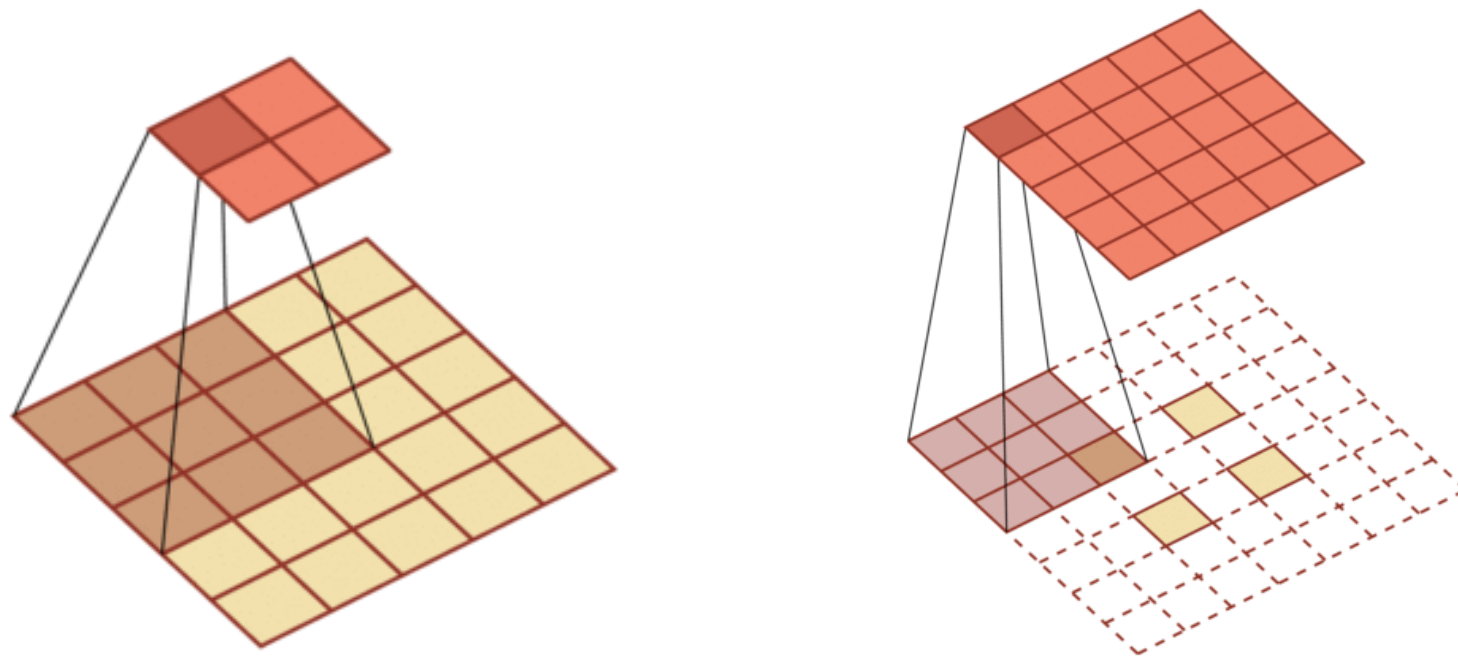
- ▶ 卷积层 (Convolutional Layer) : 卷积层通过一系列的卷积核 (或滤波器) 扫描输入数据, 计算卷积核与输入数据局部区域的点积, 从而提取特征。这些卷积核的参数在训练过程中学习得到, 能够捕捉到输入数据中的重要特征。
- ▶ 池化层 (Pooling Layer) : 池化层用于降低特征图的空间尺寸, 减少参数数量和计算量, 同时保持特征的不变性。常见的池化操作有最大池化 (Max Pooling) 和平均池化 (Average Pooling) 。
- ▶ 全连接层 (Fully Connected Layer) : 在多个卷积和池化层之后, 全连接层将学习到的特征映射到样本的标签空间, 用于分类或回归任务。
- ▶ 丢弃层 (Dropout Layer) : 丢弃层在训练过程中随机丢弃一部分神经元的激活值, 以防止过拟合, 增强模型的泛化能力。



其它卷积种类

转置卷积/微步卷积

► 低维特征映射到高维特征



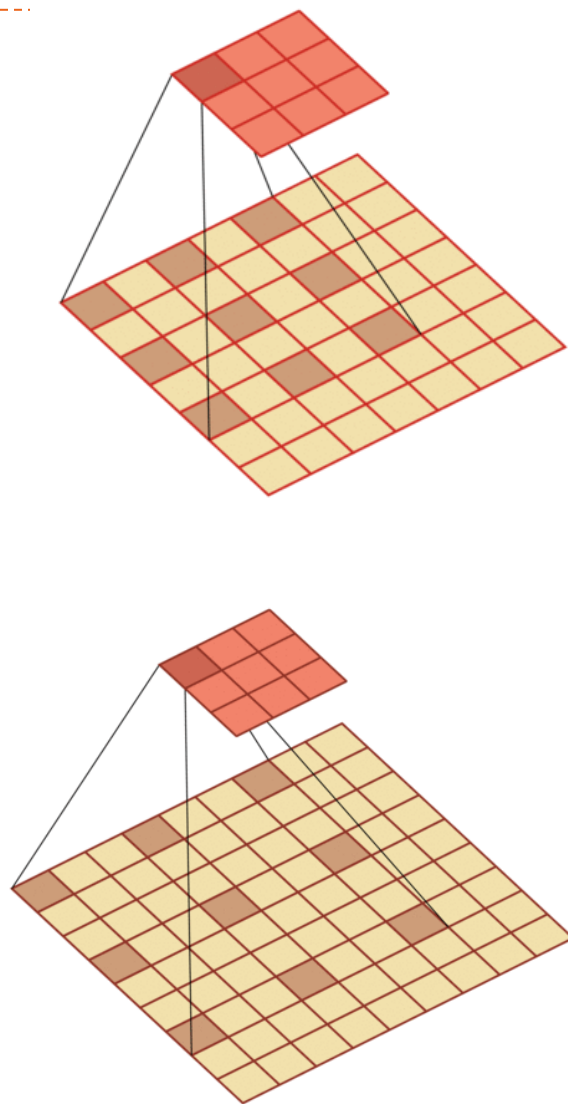
空洞卷积

► 如何增加输出单元的感受野

- 增加卷积核的大小
- 增加层数来实现
- 在卷积之前进行汇聚操作

► 空洞卷积

- 通过给卷积核插入“空洞”来变相地增加其大小。

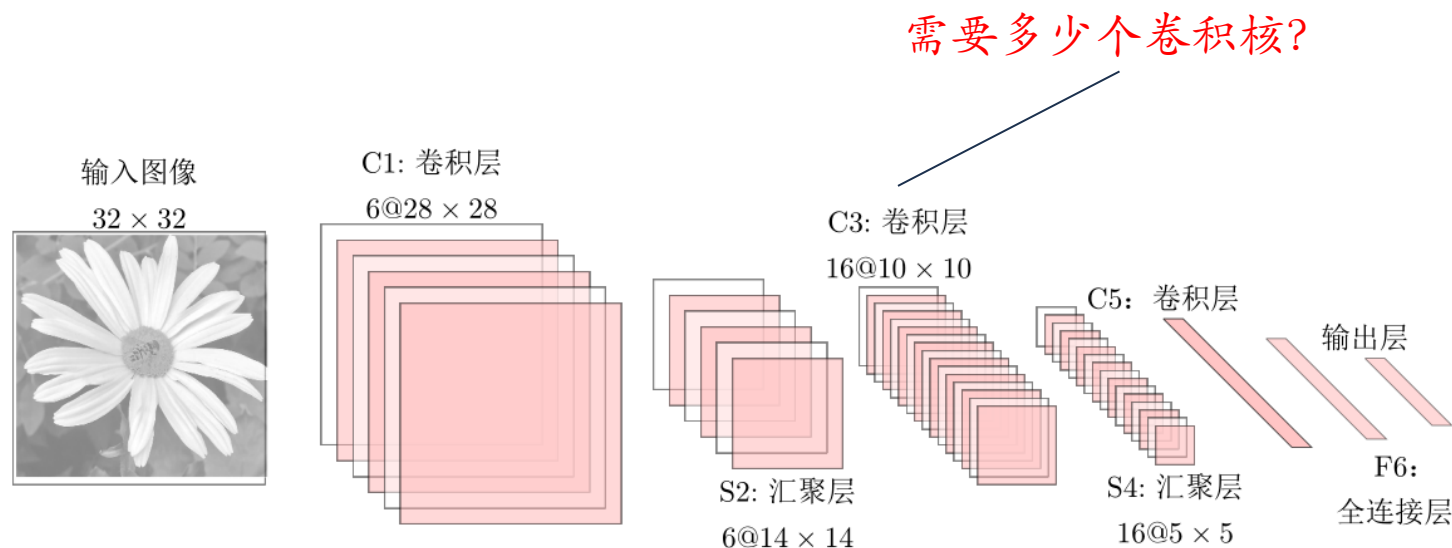




典型的卷积网络

LeNet-5

- ▶ **LeNet-5 是一个非常成功的神经网络模型。**
- ▶ 基于 LeNet-5 的手写数字识别系统在 90 年代被美国很多银行使用，用来识别支票上面的手写数字。
- ▶ LeNet-5 共有 7 层。



LeNet-5

▶1、输入层 (Input layer)

- ▶输入层接收大小为 的手写数字图像，其中包括灰度值 (0-255) 。
在实际应用中，我们通常会对输入图像进行预处理，例如对像素值进行归一化，以加快训练速度和提高模型的准确性。

▶2、卷积层C1 (Convolutional layer C1)

- ▶卷积层C1包括6个卷积核，每个卷积核的大小为 5×5 ，步长为1，填充为0。因此，每个卷积核会产生一个大小为 28×28 的特征图（输出通道数为6）。

▶3、采样层S2 (Subsampling layer S2)

- ▶采样层S2采用最大池化 (max-pooling) 操作，每个窗口的大小为 2×2 ，步长为2。因此，每个池化操作会从4个相邻的特征图中选择最大值，产生一个大小为 14×14 的特征图（输出通道数为6）。

LeNet-5

►4、卷积层C3 (Convolutional layer C3)

- 卷积层C3包括16个卷积核，每个卷积核的大小为 5×5 ，步长为1，填充为0。因此，每个卷积核会产生一个大小为 10×10 的特征图（输出通道数为16）。

►5、采样层S4 (Subsampling layer S4)

- 采样层S4采用最大池化操作，每个窗口的大小为 2×2 ，步长为2。因此，每个池化操作会从4个相邻的特征图中选择最大值，产生一个大小为 5×5 的特征图（输出通道数为16）。

LeNet-5

▶6、全连接层C5 (Fully connected layer C5)

- ▶C5将每个大小为 5×5 的特征图拉成一个长度为400的向量，并通过一个带有120个神经元的全连接层进行连接。120是由LeNet-5的设计者根据实验得到的最佳值。

▶7、全连接层F6 (Fully connected layer F6)

- ▶全连接层F6将120个神经元连接到84个神经元。

▶8、输出层 (Output layer)

- ▶输出层由10个神经元组成，每个神经元对应0-9中的一个数字，并输出最终的分类结果。在训练过程中，使用交叉熵损失函数计算输出层的误差，并通过反向传播算法更新卷积核和全连接层的权重参数。

AlexNet

► 2012 ILSVRC winner

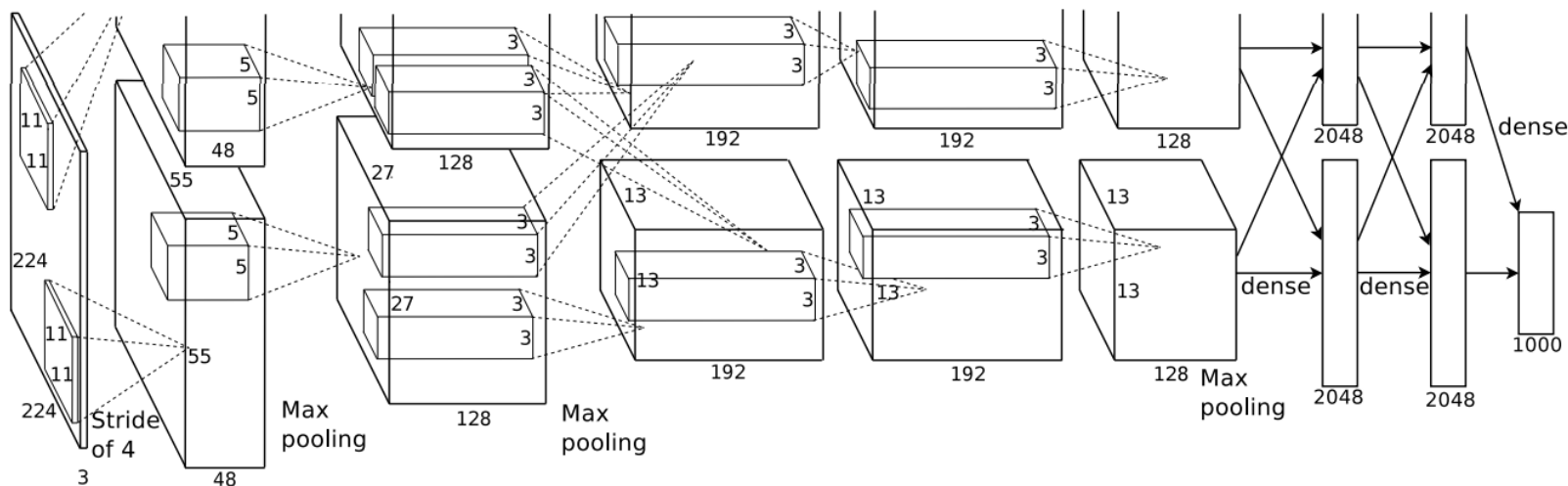
► (top 5 error of 16% compared to runner-up with 26% error)

► 第一个现代深度卷积网络模型

► 首次使用了很多现代深度卷积网络的一些技术方法

□ 使用GPU进行并行训练，采用了ReLU作为非线性激活函数，使用Dropout防止过拟合，使用数据增强

► 5个卷积层、3个汇聚层和3个全连接层



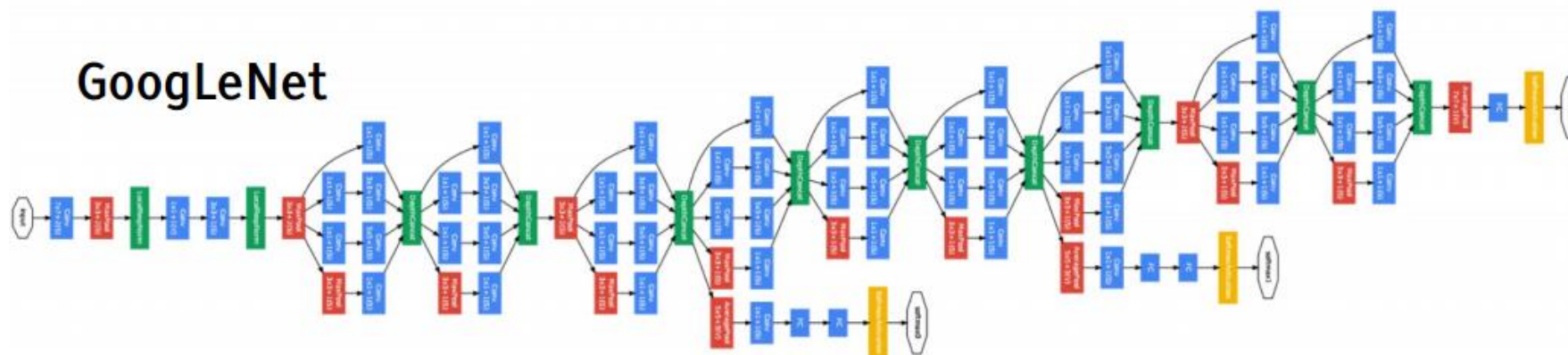
Inception网络

► 2014 ILSVRC winner (22层)

► 参数: GoogLeNet: 4M VS AlexNet: 60M

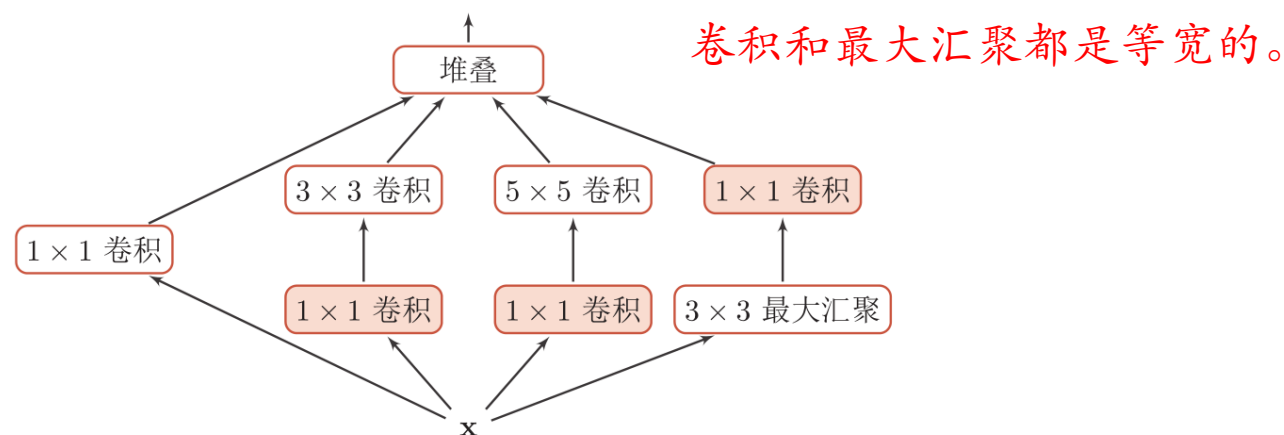
► 错误率: 6.7%

► Inception网络是由有多个inception模块和少量的汇聚层堆叠而成。



Inception模块 v1

- ▶ 在卷积网络中，如何设置卷积层的卷积核大小是一个十分关键的问题。
- ▶ 在Inception网络中，一个卷积层包含多个不同大小的卷积操作，称为Inception模块。
- ▶ Inception模块同时使用 1×1 、 3×3 、 5×5 等不同大小的卷积核，并将得到的特征映射在深度上拼接（堆叠）起来作为输出特征映射。



Inception模块 v3

- ▶ 用多层小卷积核替换大卷积核，以减少计算量和参数量。
- ▶ 使用两层3x3的卷积来替换v1中的5x5的卷积
- ▶ 使用连续的nx1和1xn来替换nxn的卷积。

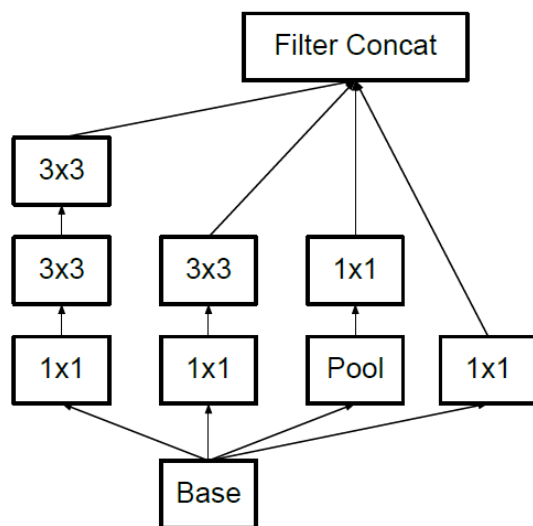


Figure 5. Inception modules where each 5×5 convolution is replaced by two 3×3 convolution, as suggested by principle [3] of Section 2.

<http://blog.csdn.net/xbinworld>

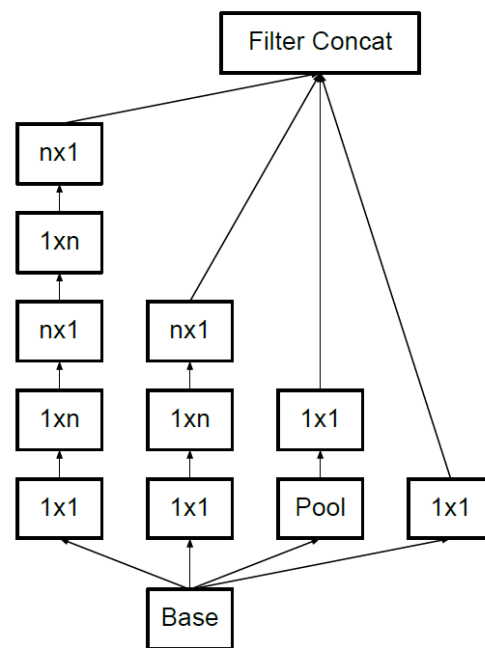


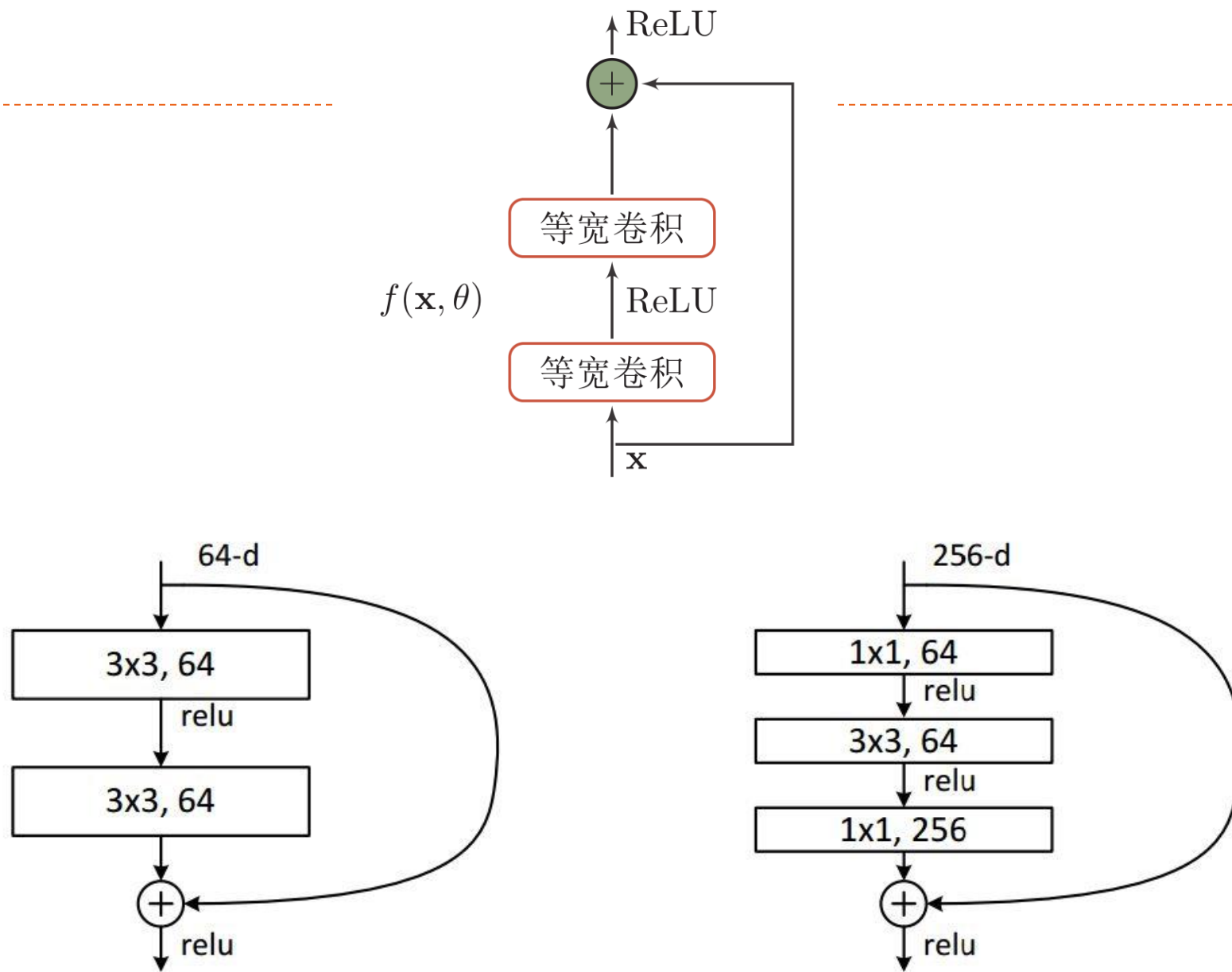
Figure 6. Inception modules after the factorization of the $n \times n$ convolutions. In our proposed architecture, we chose $n = 7$ for the 17×17 grid. (The filter sizes are picked using principle [3])

残差网络

- ▶ 残差网络 (Residual Network, ResNet) 是通过给非线性的卷积层增加**直连边**的方式来提高信息的传播效率。
- ▶ 假设在一个深度网络中, 我们期望一个非线性单元 (可以为一层或多层的卷积层) $f(\mathbf{x}, \theta)$ 去逼近一个目标函数为 $h(\mathbf{x})$ 。
- ▶ 将目标函数拆分成两部分: **恒等函数**和**残差函数**

$$h(\mathbf{x}) = \underbrace{\mathbf{x}}_{\text{恒等函数}} + \underbrace{(h(\mathbf{x}) - \mathbf{x})}_{\text{残差函数}} \rightarrow f(\mathbf{x}, \theta)$$

残差单元



ResNet

► 2015 ILSVRC winner (152层)

► 错误率: 3.57%

