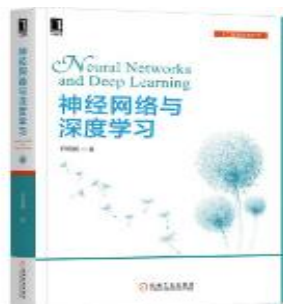


MACHINE LEARNING

机器学习

Neural Networks

卷积神经网络原理



参考：
《神经网络与深度学习》

Machine Learning Course
Copyright belongs to Wenting Tu.

动机

• 表示学习

为了提高机器学习系统的准确率，我们就需要将输入信息转换为有效的特征，或者更一般性地称为表示（Representation）。如果有一种算法可以自动地学习出有效的特征，并提高最终机器学习模型的性能，那么这种学习就可以叫作表示学习（Representation Learning）

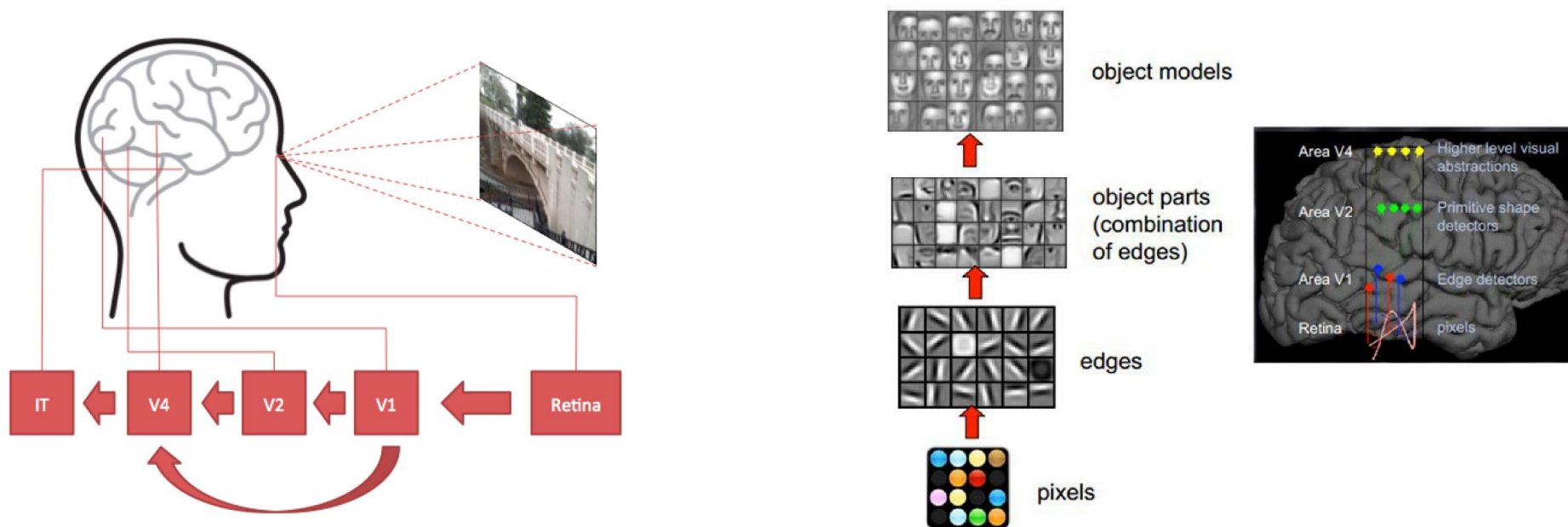
一般而言，一个好的表示具有以下几个优点：

- （1）一个好的表示应该具有很强的表示能力，即同样大小的向量可以表示更多信息。
- （2）一个好的表示应该使后续的学习任务变得简单，即需要包含更高层的语义信息。
- （3）一个好的表示应该具有一般性，是任务或领域独立的。虽然目前的大部分表示学习方法还是基于某个任务来学习，但我们期望其学到的表示可以比较容易地迁移到其他任务上。）

动机

• 表示学习

人类的可视皮层具有“深度结构/层次结构”(deep architecture)

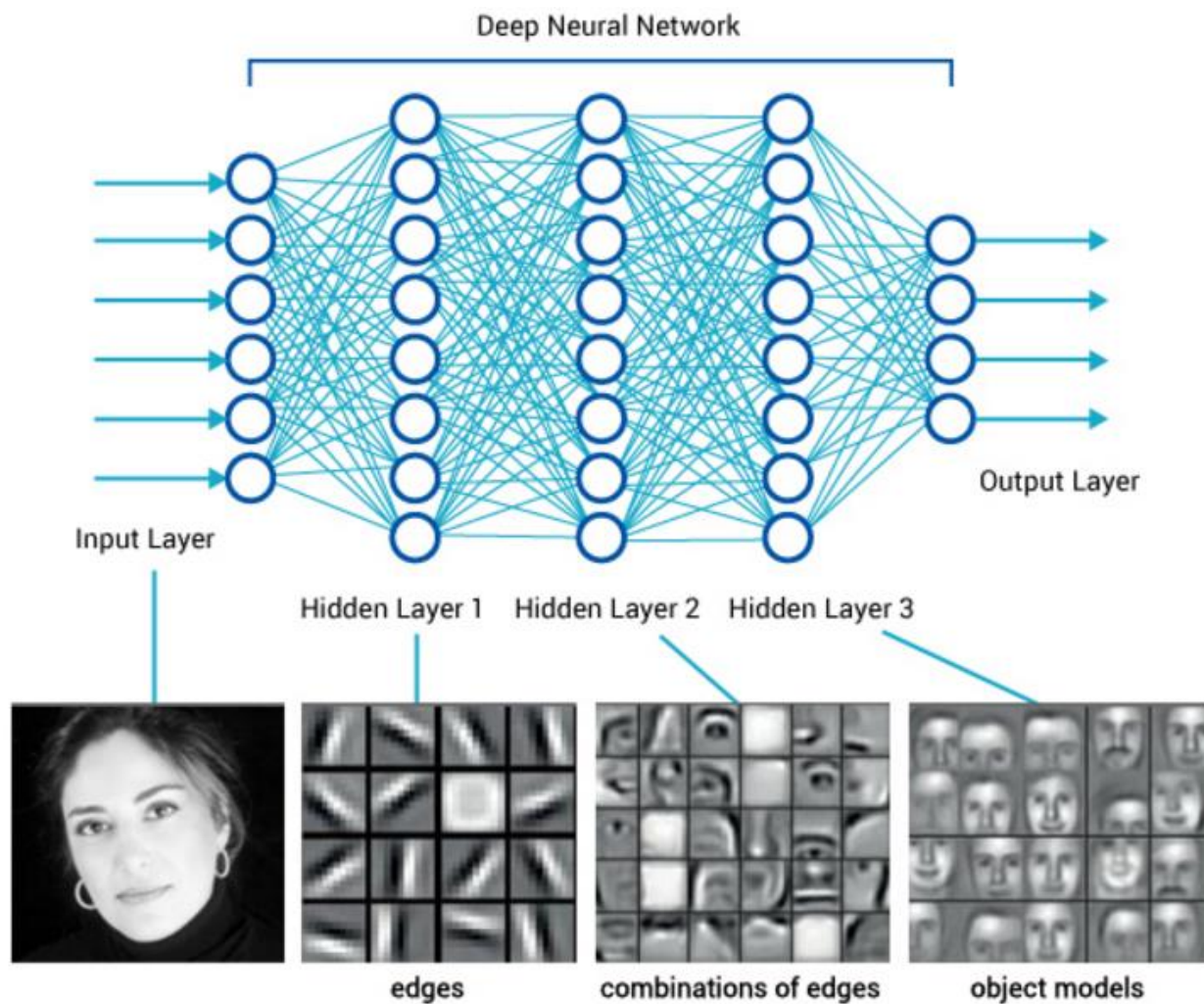


从低级的 V1 区提取边缘特征，再到 V2 区的形状或者目标的部分等，再到更高层，整个目标、目标的行为等。也就是说高层的特征是低层特征的组合，从低层到高层的特征表示越来越抽象，越来越能表现语义或者意图。而抽象层面越高，就越利于分类。

要学习到一种好的高层语义表示（一般为分布式表示），通常需要从底层特征开始，经过多步非线性转换才能得到。深层结构的优点是可以增加特征的重用性，从而指数级地增加表示能力。因此，表示学习的关键是构建具有一定深度的多层次特征表示 [Bengio et al., 2013]

动机

- 深度全连接前馈网络



动机

- 深度全连接前馈网络面临的难点
 - 参数过多
 - 梯度消失
 - 非凸优化
 - 解释性
 - 对数据和计算资源的需求

卷积神经网络 *Convolutional Neural Network (CNN or ConvNet)* 是一种具有局部连接、权重共享等特性的深层前馈神经网络。

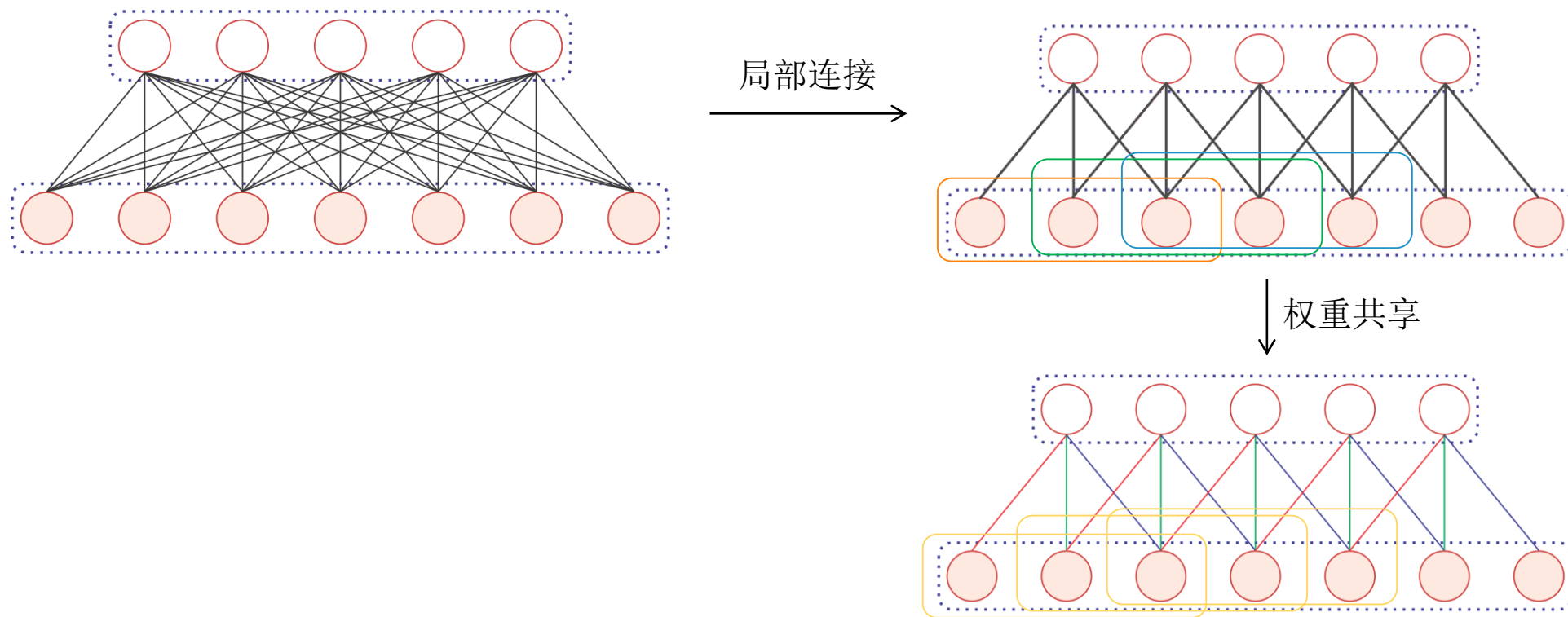
卷积神经网络

• 卷积层

卷积层通过**局部连接**、**权重共享**以及**汇聚**大幅度降低模型参数量。

局部连接: 在卷积层（假设是第 l 层）中的每一个神经元都只和下一层（第 $l-1$ 层）中某个局部窗口内的神经元相连，构成一个局部连接网络。所以比起全连接层，卷积层和下一层之间的连接数大大减少，由原来的 $m^l \times m^{l-1}$ 个连接变为 $m^l \times K$ 个连接（非权重共享）， K 为滤波器大小。

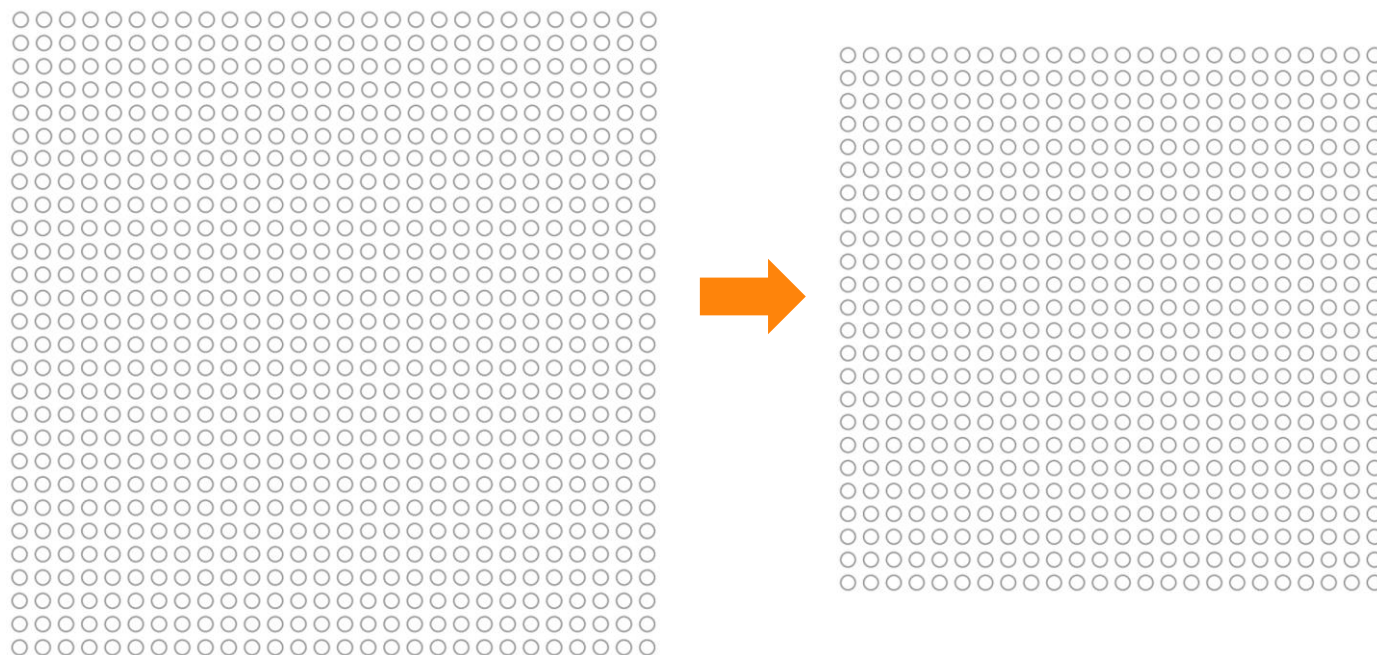
权重共享: 作为参数的卷积核 w^l 对于第 l 层的所有的神经元都是相同的（所有的同颜色连接上的权重是相同的）。权重共享可以理解为一个滤波器只捕捉输入数据中的一种特定的局部特征。因此，如果要提取多种特征就需要使用多个不同的滤波器。由原来的 $m^l \times K$ 个连接变为 $P \times K$ 个连接， P 为滤波器个数。



卷积神经网络

- 卷积层

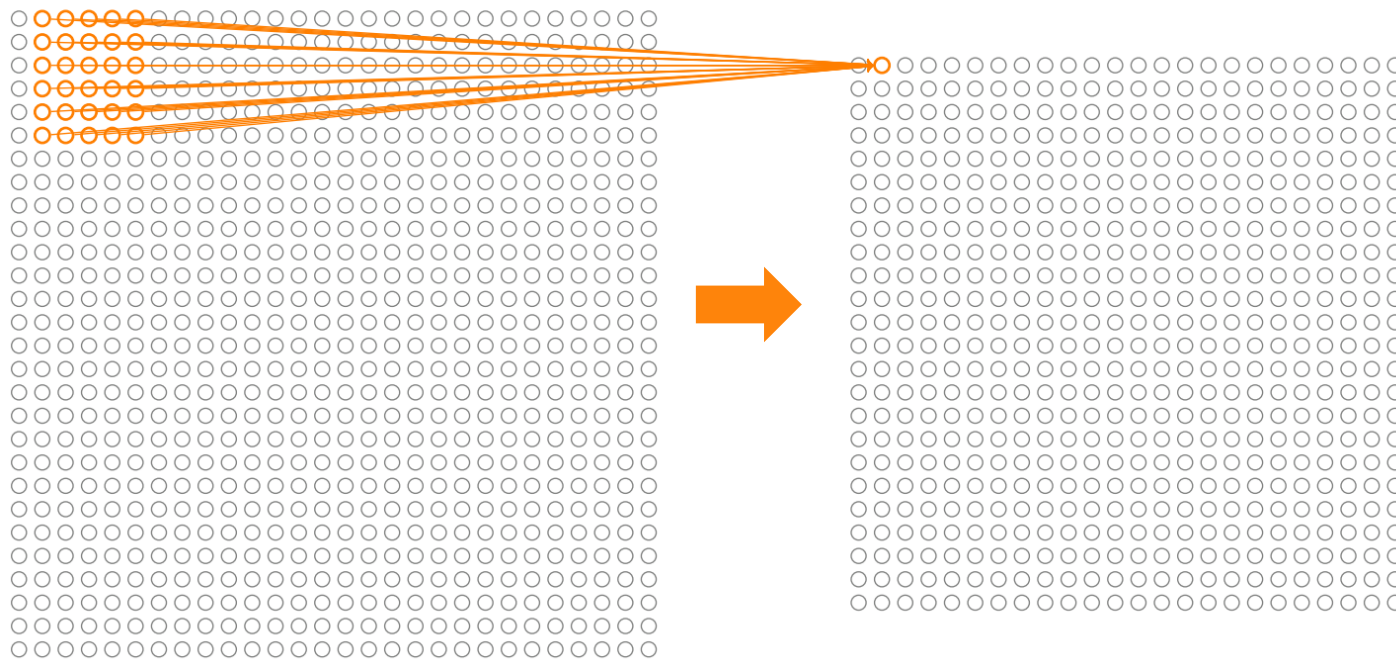
卷积层可处理矩阵数据



卷积神经网络

- 卷积层

卷积层可处理矩阵数据



卷积神经网络

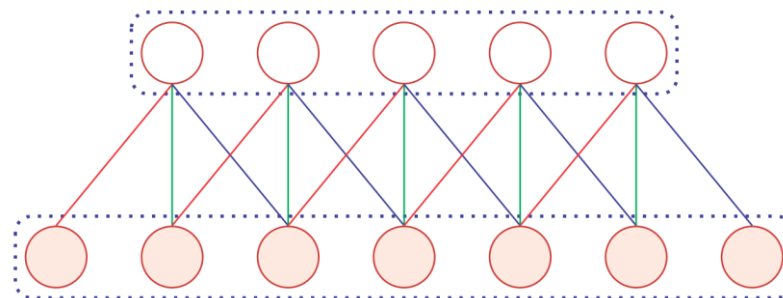
• 卷积层

卷积层由卷积（Convolution）运算来实现数据表示的转换

一维序列： x_1, x_2, \dots, x_N

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

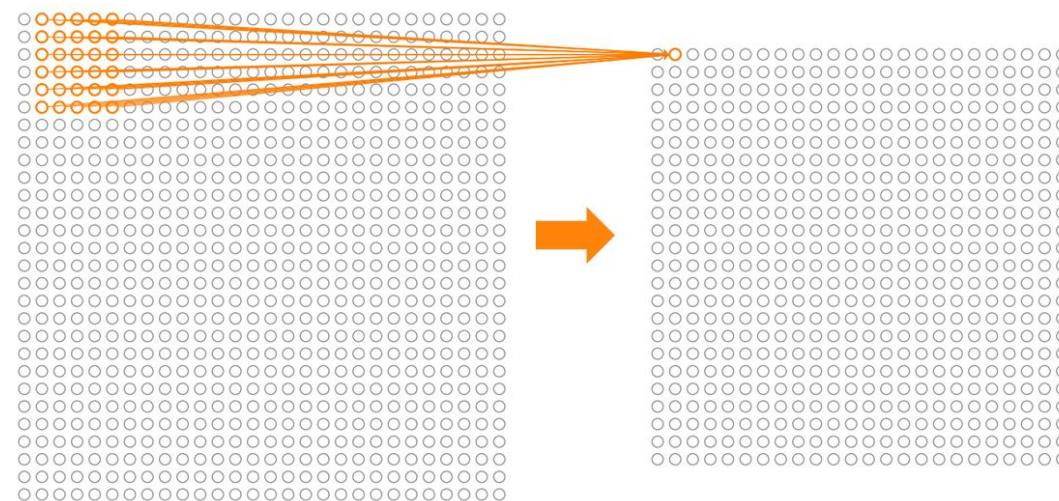
这里假设卷积的输出 y_t 的下标 t 从 K 开始



二维序列： $x_{1,1}, x_{1,2}, \dots, x_{M,N}$

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

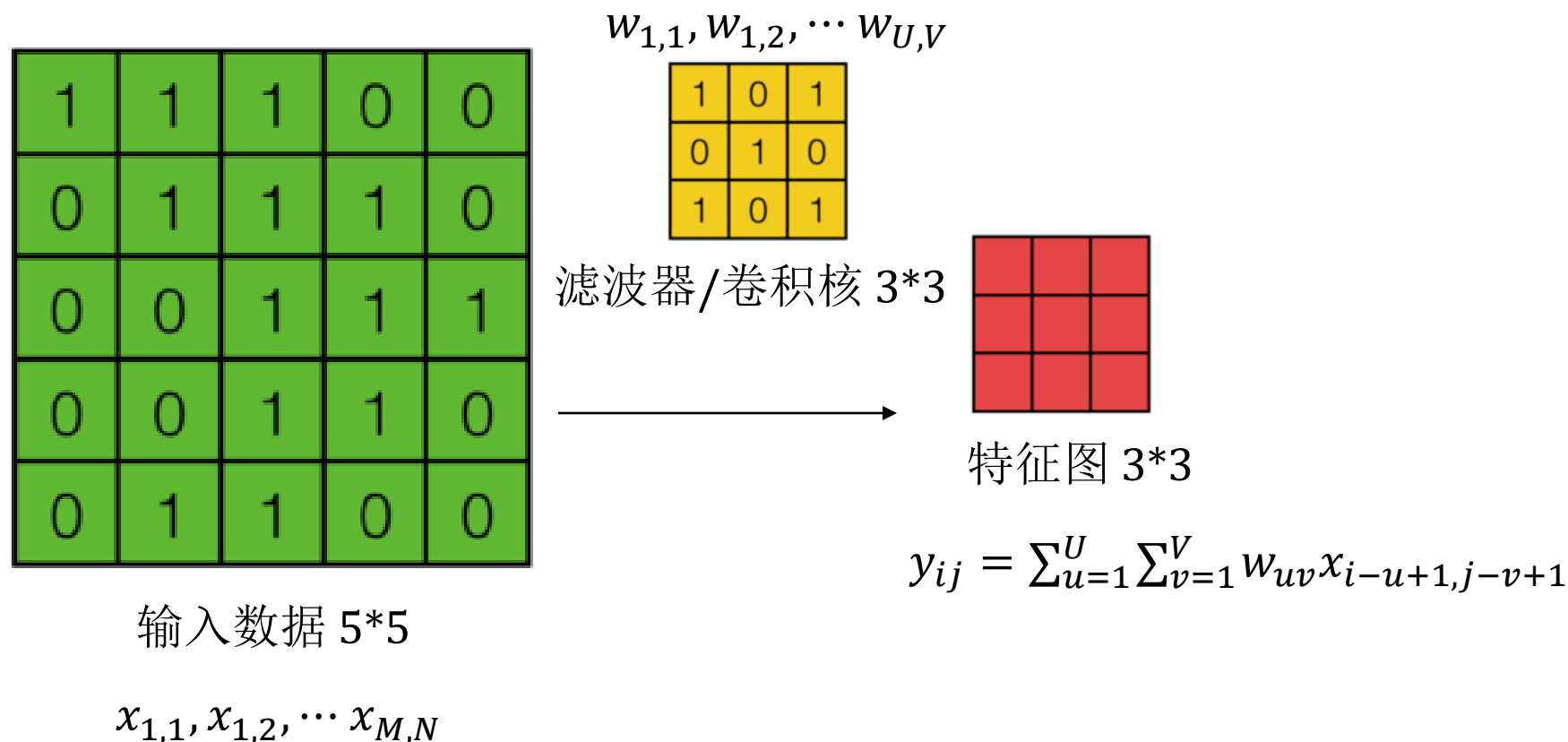
卷积的输出 y_t 的下标 (i, j) 从 (U, V) 开始



卷积神经网络

• 卷积层

卷积层由卷积（Convolution）运算来实现数据表示的转换



卷积神经网络

- 卷积层

卷积层由卷积（Convolution）运算来实现数据表示的转换

1 _{x1}	1 _{x0}	1 _{x1}	0	0
0 _{x0}	1 _{x1}	1 _{x0}	1	0
0 _{x1}	0 _{x0}	1 _{x1}	1	1
0	0	1	1	0
0	1	1	0	0

输入数据 5*5

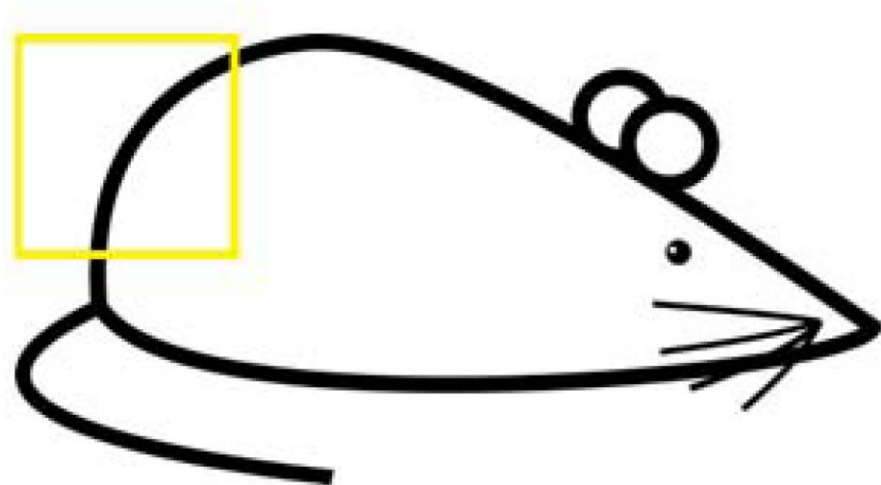
4		

特征图
3*3

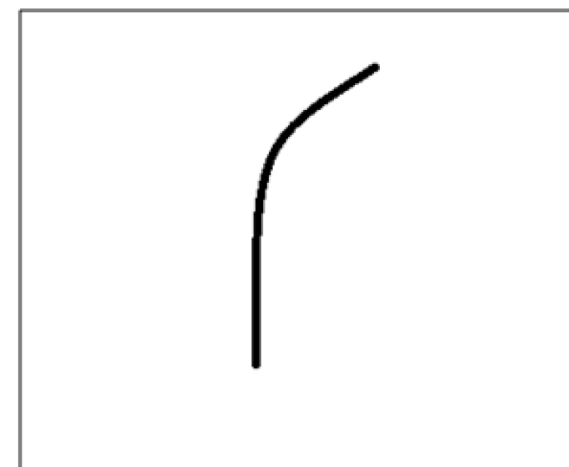
卷积神经网络

• 卷积层

卷积层由卷积（Convolution）运算来实现数据表示的转换



0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0



$$(50 \times 30) + (50 \times 30) + (50 \times 30) + (20 \times 30) + (50 \times 30) = 6600$$



0	0	0	0	0	0	30
0	0	0	0	50	50	50
0	0	0	20	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0

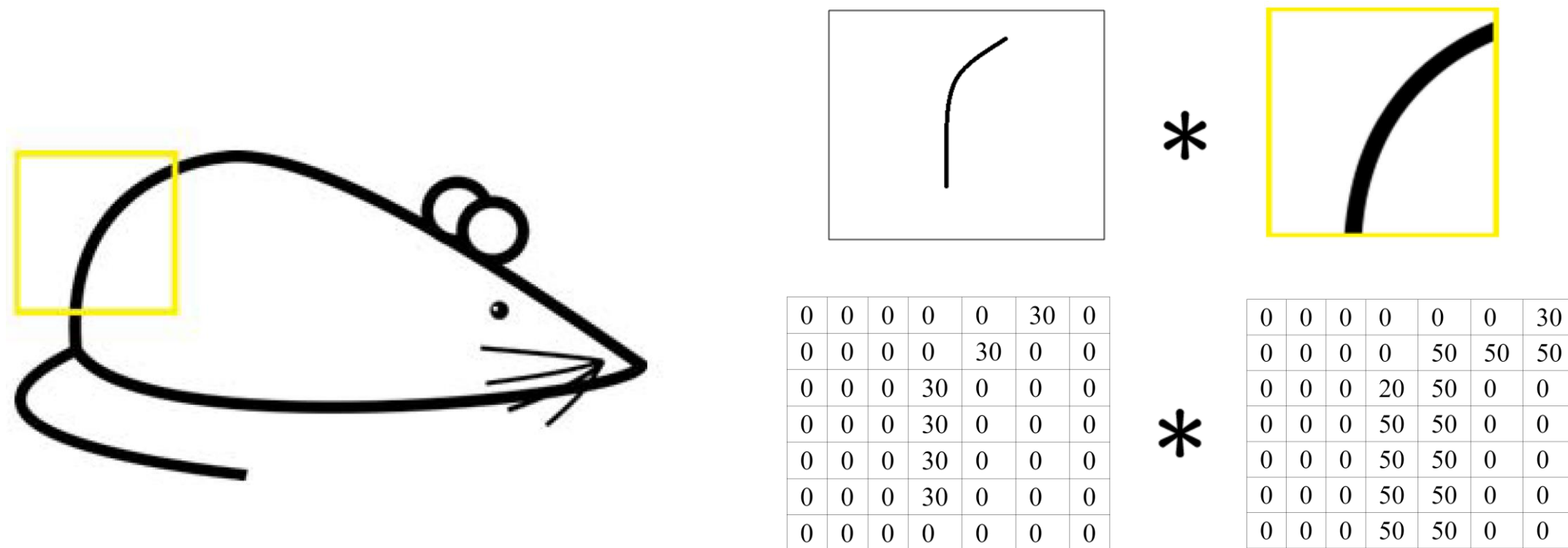
*

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

卷积神经网络

• 卷积层

卷积层由卷积（Convolution）运算来实现数据表示的转换

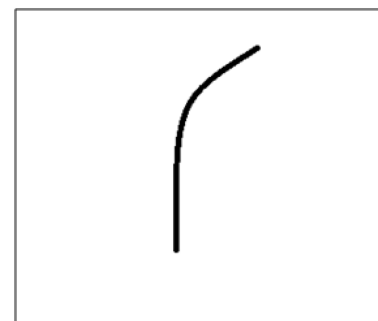


$$(50 \times 30) + (50 \times 30) + (50 \times 30) + (20 \times 30) + (50 \times 30) = 6600$$

卷积神经网络

• 卷积层

卷积层由卷积（Convolution）运算来实现数据表示的转换



*



*

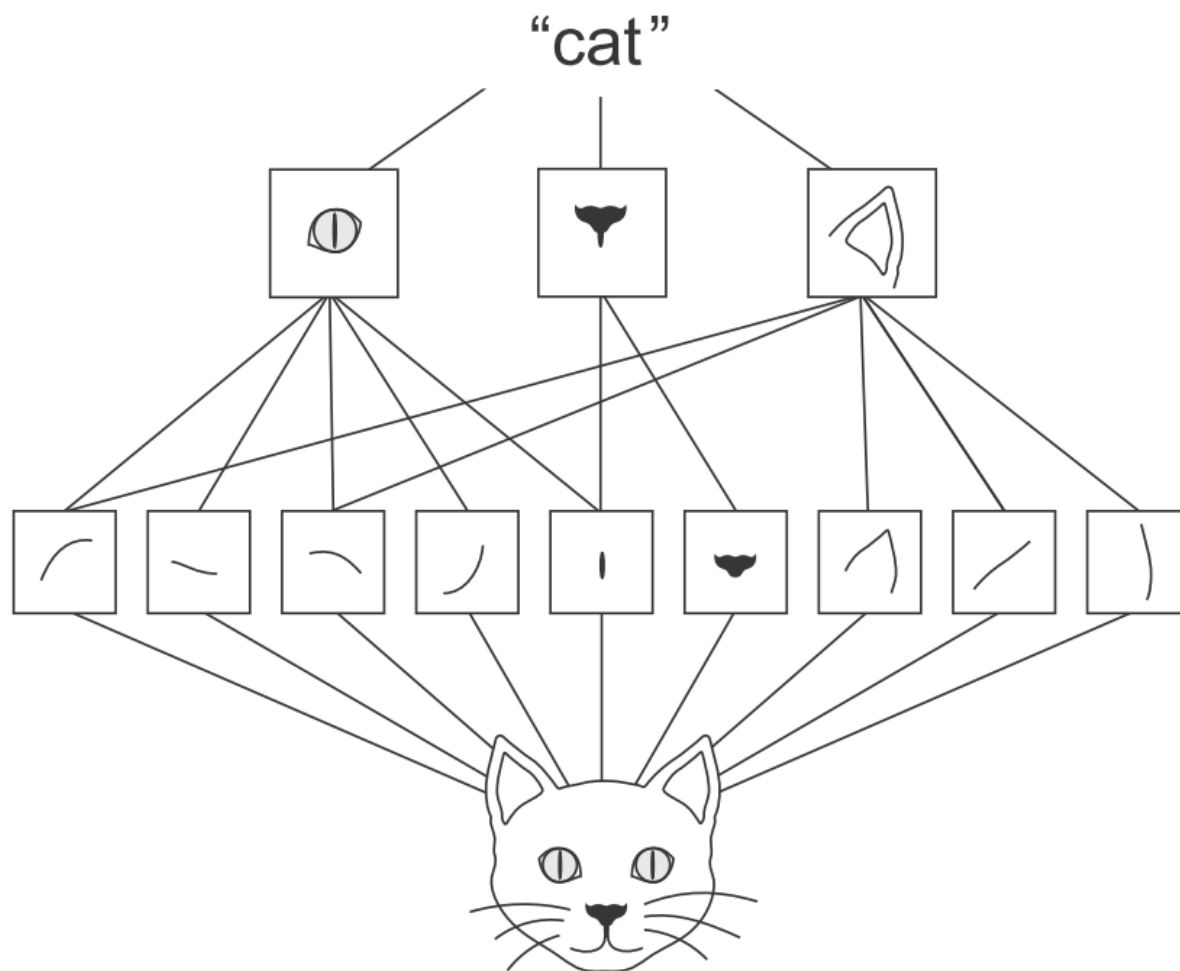
0	0	0	0	0	0	0
0	40	0	0	0	0	0
40	0	40	0	0	0	0
40	20	0	0	0	0	0
0	50	0	0	0	0	0
0	0	50	0	0	0	0
25	25	0	50	0	0	0

0

卷积神经网络

- 卷积层

卷积层可由多个卷积核共同作用实现提取数据不同角度的抽象特征



卷积神经网络

- 卷积层

可搭配激活函数使用

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

↓

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

$$a_{i,j} = f(z_{ij})$$

卷积神经网络

• 卷积层

卷积核的深度

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 black & white
picture image

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

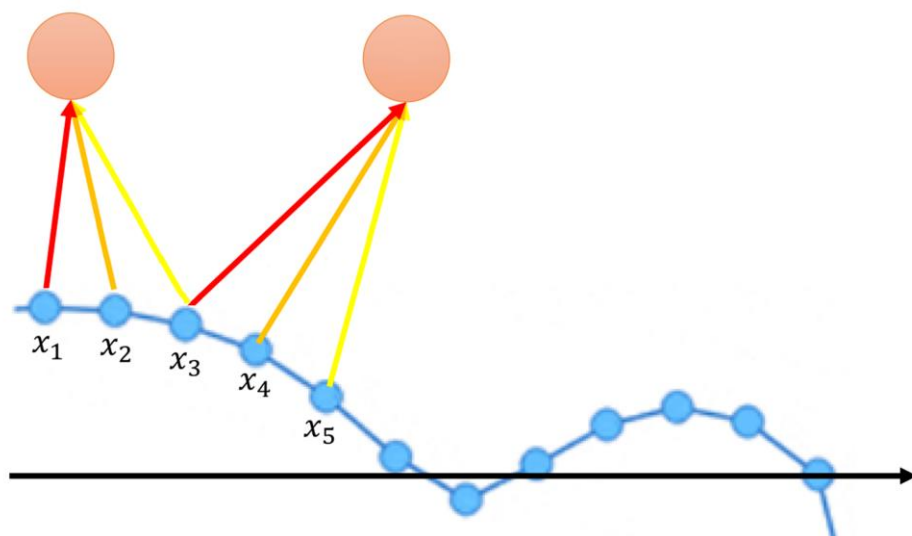
6 x 6 colorful image

$$a_{i,j} = f(\sum_{u=1}^U \sum_{v=1}^V w_{uv} x_{i-u+1, j-v+1} + w_b)$$

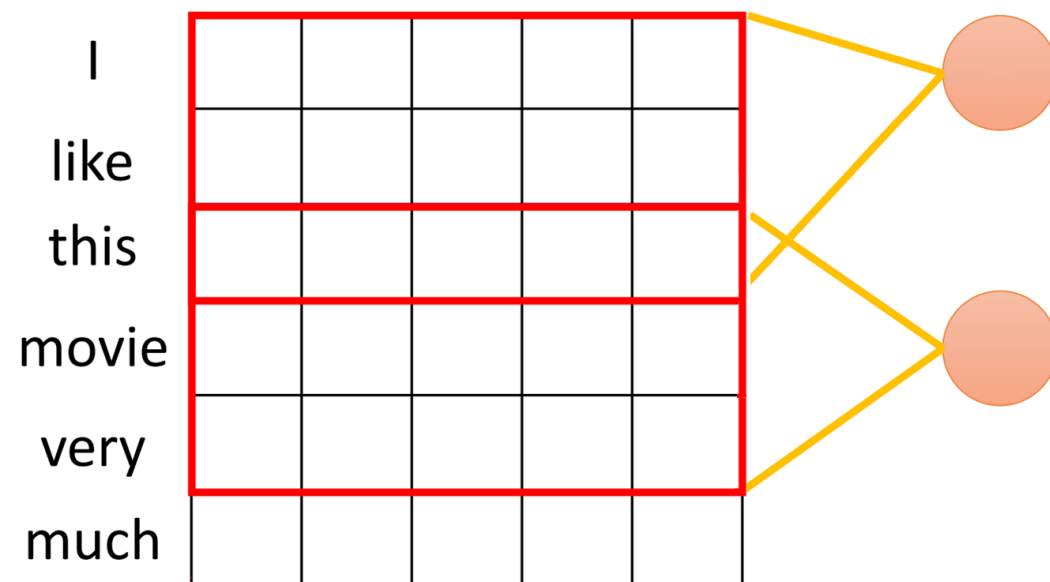
$$a_{i,j} = f(\sum_{d=1}^D \sum_{u=1}^U \sum_{v=1}^V w_{uvd} x_{i-u+1, j-v+1, d} + w_b)$$

卷积神经网络

- 卷积层
- 卷积核的维度
1DConv



$$y_t = w_1 \times x_t + w_2 \times x_{t-1} + w_3 \times x_{t-2}$$



$$y_t = \mathbf{w}_1 \times \mathbf{x}_t + \mathbf{w}_2 \times \mathbf{x}_{t-1} + \mathbf{w}_3 \times \mathbf{x}_{t-2}$$

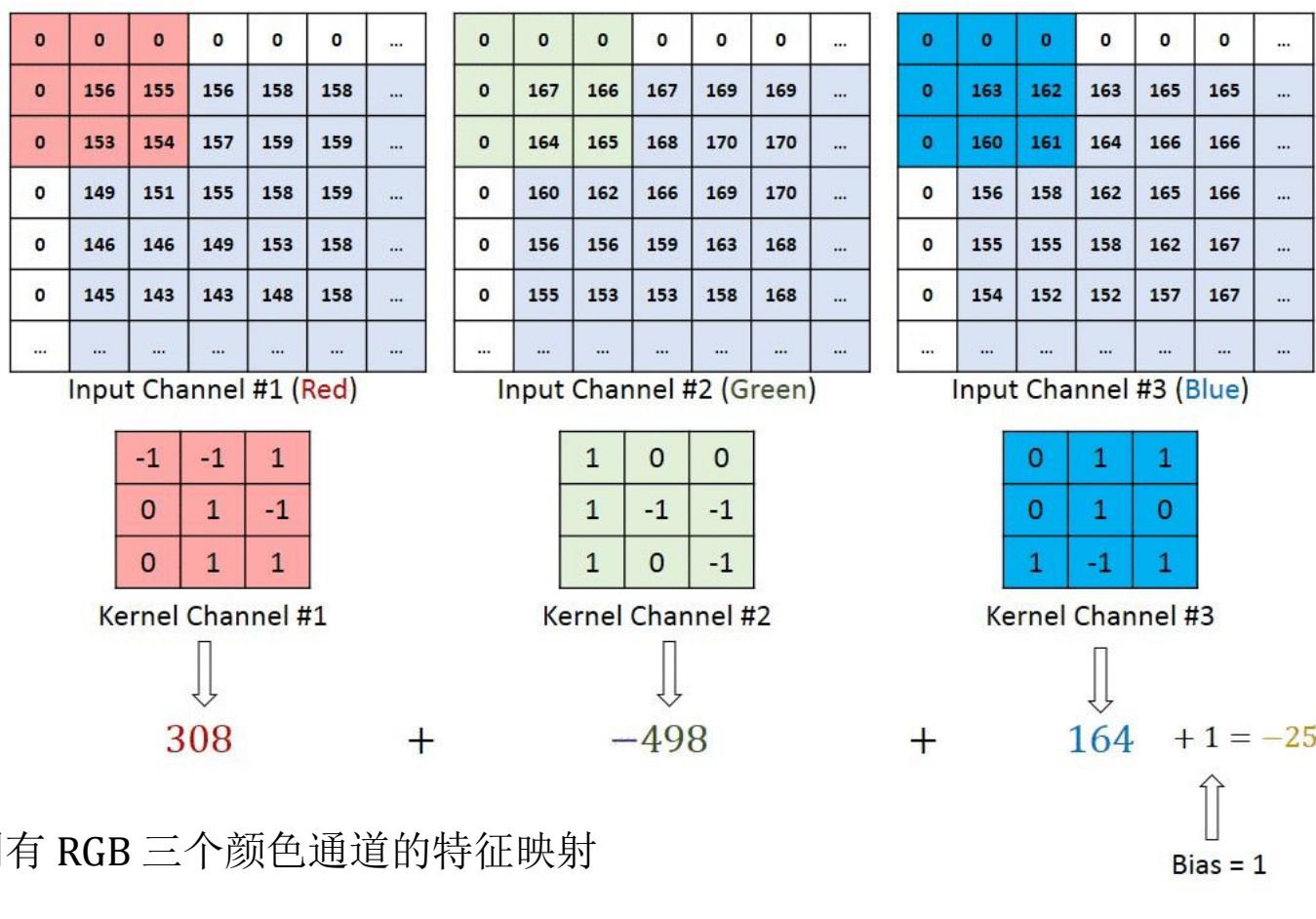
卷积神经网络

• 卷积层

通道的概念

深度有时候也称为数据的“通道” *channel*

卷积核通常与它所处理的数据/输入数据拥有同样的通道数



彩色图像分别有 RGB 三个颜色通道的特征映射

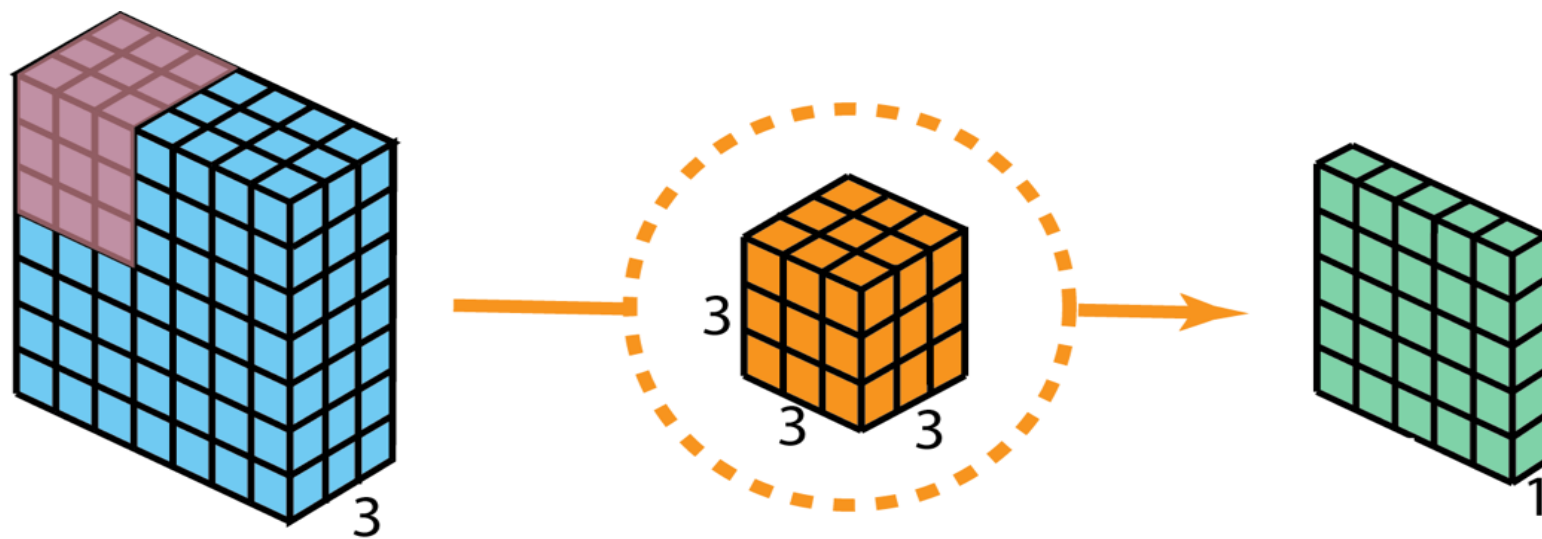
卷积神经网络

- 卷积层

通道的概念

深度有时候也称为数据的“通道” *channel*

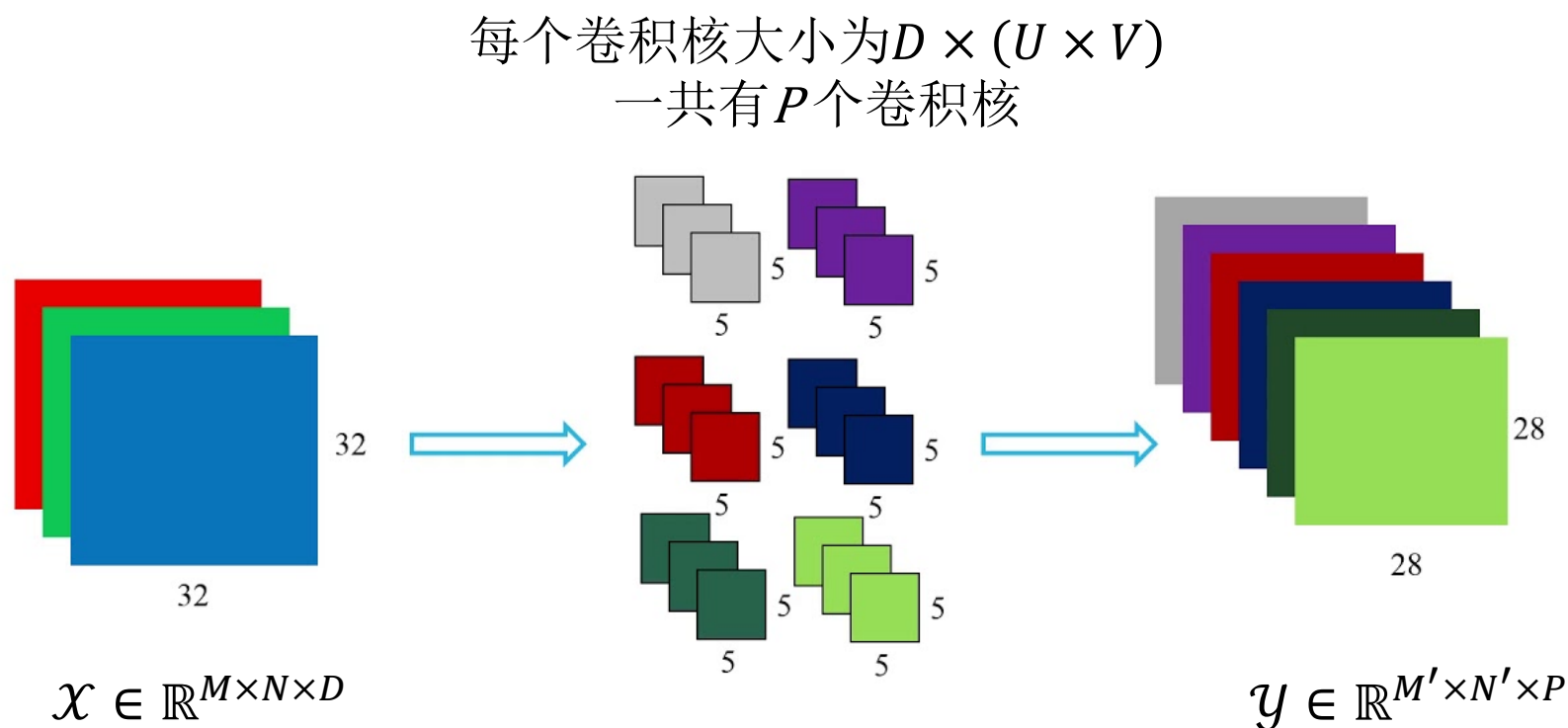
卷积核通常与它所处理的数据/输入数据拥有同样的通道数



卷积神经网络

• 卷积层

输出数据的通道数则是由卷积核个数所决定的

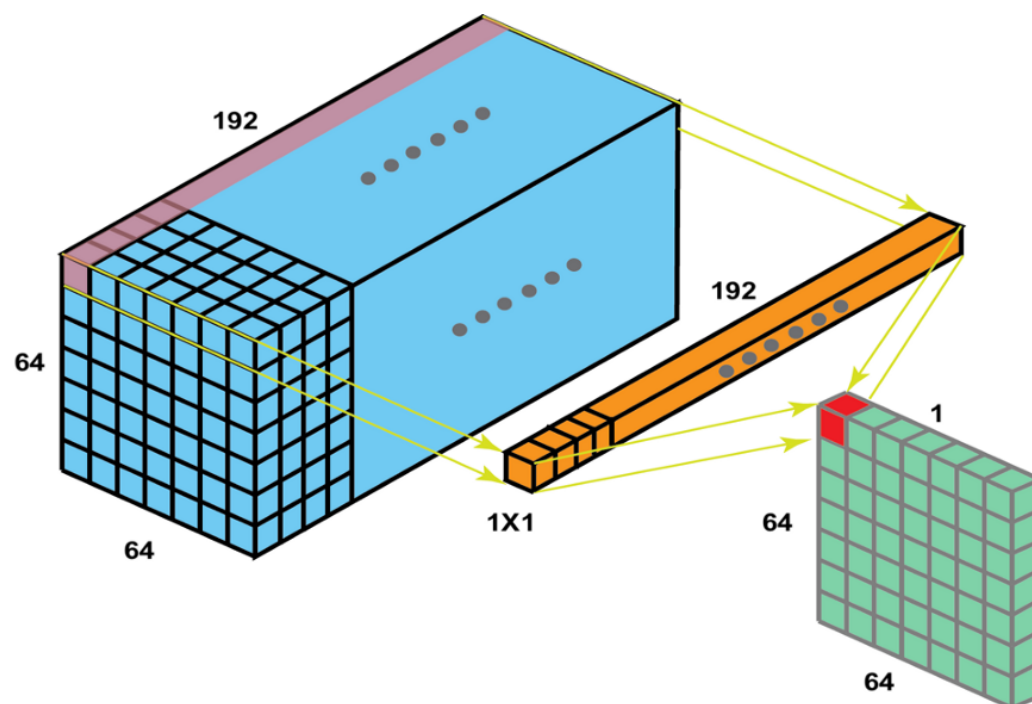


参数量: $P \times D \times (U \times V) + P$

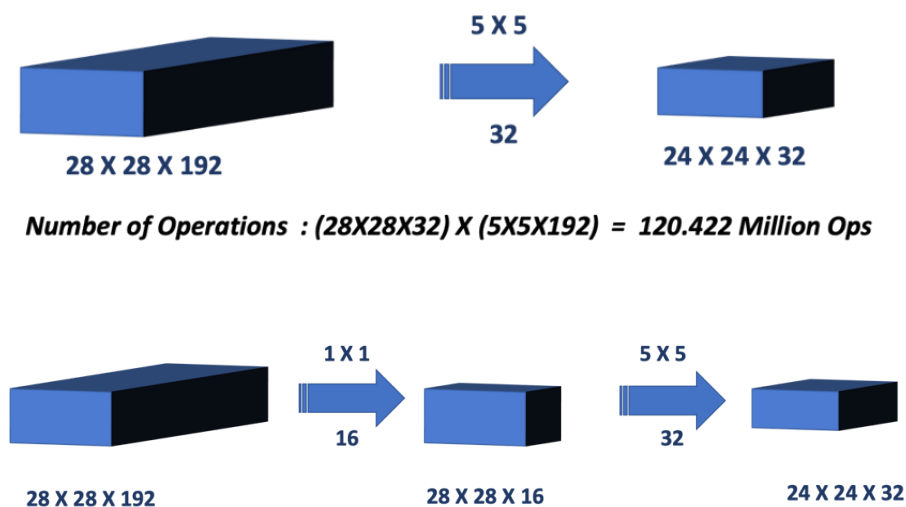
卷积神经网络

• 卷积层

1x1卷积核对输入数据的通道做约简



每个1x1卷积核相当于在输入数据的通道上做了一个降维（经过一个神经元个数为1的全连接层），从而相当于大幅度降低了特征图的数量，但不影响特征图的结构。



Number of Operations for 1 X 1 Conv Step : $(28 \times 28 \times 16) \times (1 \times 1 \times 192) = 2.4 \text{ Million Ops}$

Number of Operations for 5 X 5 Conv Step : $(28 \times 28 \times 32) \times (5 \times 5 \times 16) = 10 \text{ Million Ops}$

Total Number of Operations = 12.4 Million Ops

卷积神经网络

- 卷积层

卷积核滑动时的间隔，称为步长（Stride），是可以灵活设定的

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

image 5*5

1	0	1
0	1	0
1	0	1

bias = 0

filter 3*3

4	4
2	4

feature map 2*2

卷积神经网络

• 卷积层

卷积核可以在输入张量上进行零填充 *Zero Padding*

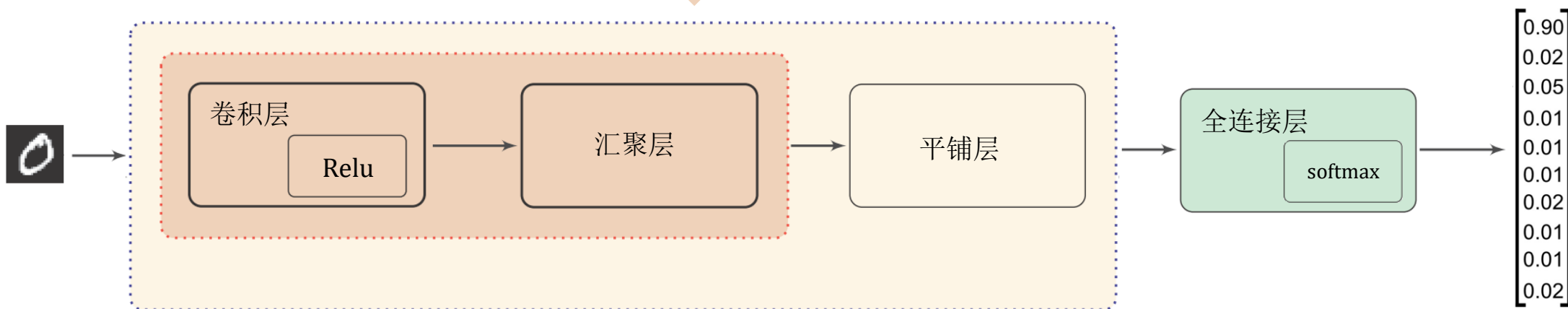
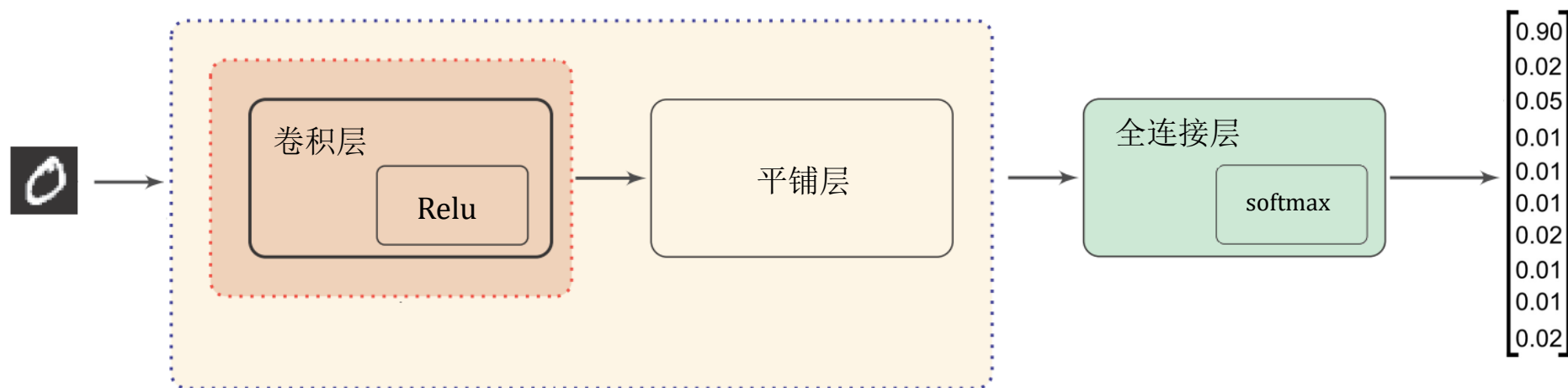
x1	x0	x1
x0	x1	x0
x1	x0	x1

0	0	0	0	0	0	0
0	1	1	1	0	0	0
0	0	1	1	1	0	0
0	0	0	1	1	1	0
0	0	1	1	1	0	0
0	0	1	1	0	0	0
0	0	0	0	0	0	0

2	2	3	1	1
1	4	3	4	1
2	2	4	3	3
1	2	3	4	1
1	2	3	1	1

卷积神经网络

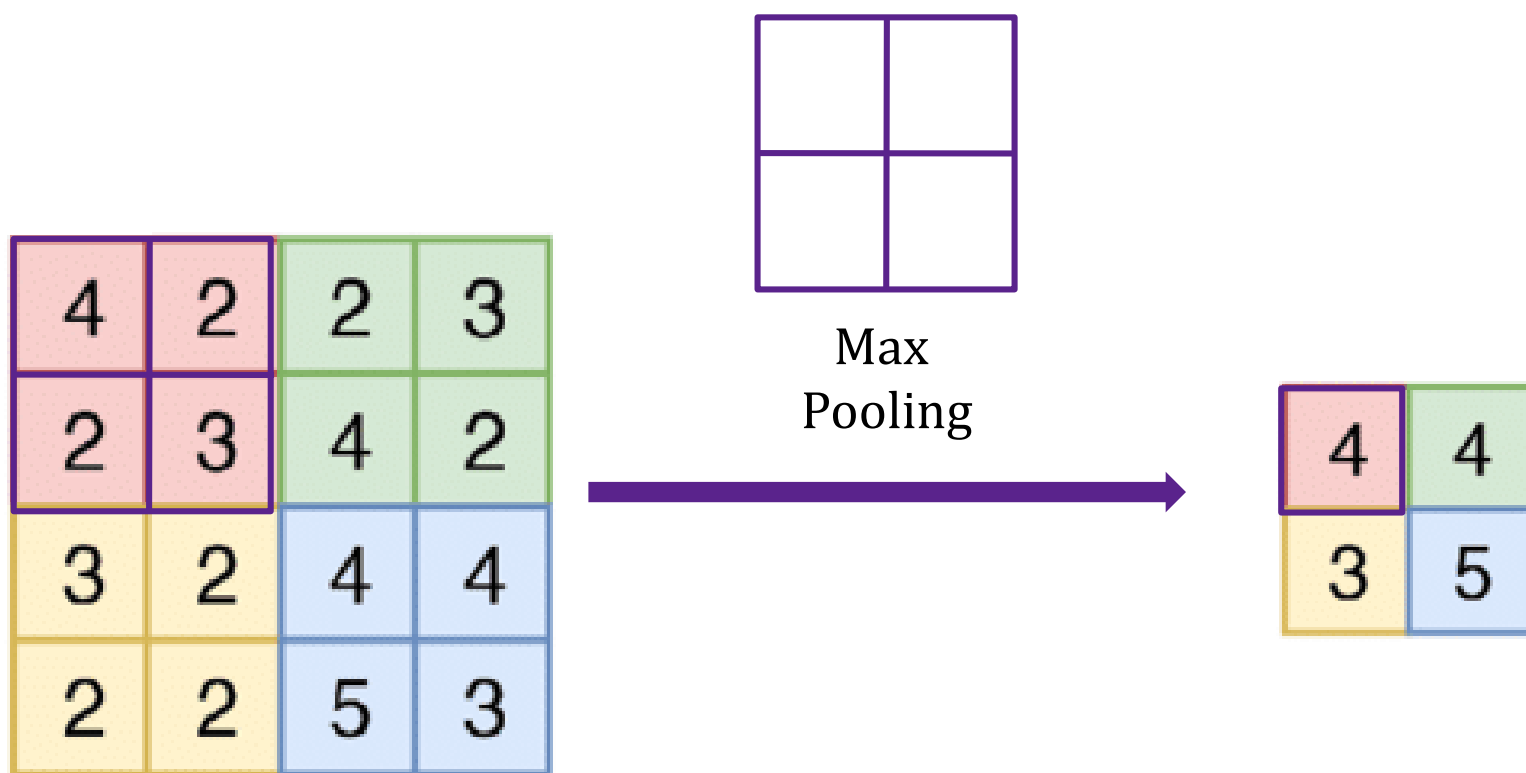
• 经典卷积神经网络



卷积神经网络

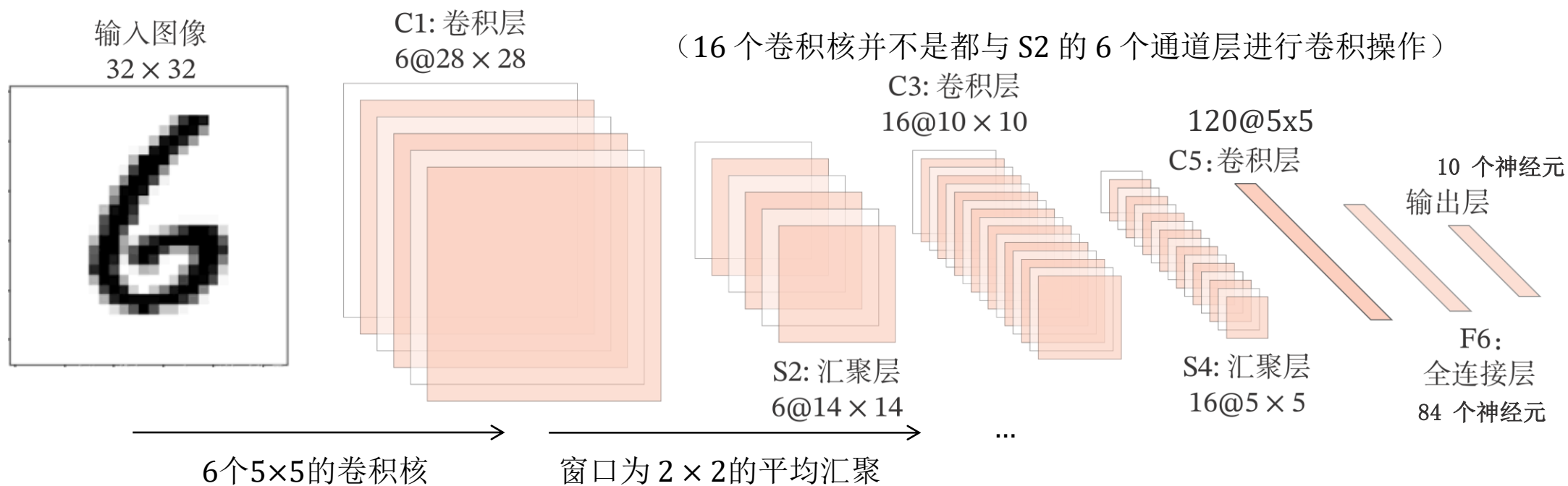
- 汇聚层

汇聚层（Pooling Layer）也叫子采样层（Subsampling Layer），其作用是进行特征选择，降低特征数量，从而减少参数数量。



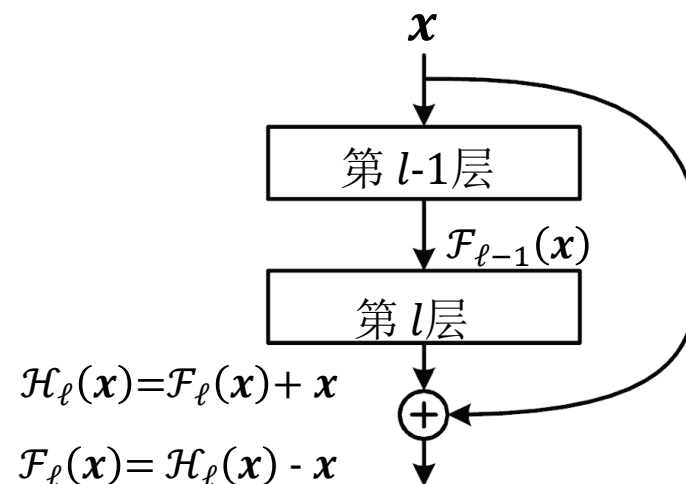
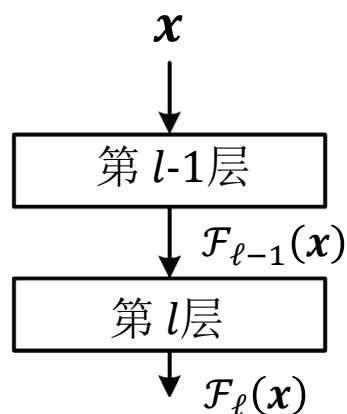
卷积神经网络

- 经典卷积神经网络
LeNet-5模型



卷积神经网络

• 经典卷积神经网络 ResNet 残差网络



- 简化学习任务为学习到高阶表示相较于低阶表示之间的残差
- 即使网络架构上，网络很深。但如果网络在某一层上已经达到了最优，在后续层上可以通过学习到 $\mathcal{F}_l(x) = 0$ ，来使得网络一直维持在最优状态上。如此缓解了由深度带来的过拟合
- 活用1x1卷积核来做维度的约简与还原（bottle-neck building block）

