

## 1. 基本卷积函数的变体

神经网络中的卷积通常指由多个并行卷积组成的运算，因为具有单个核的卷积只能提取一种类型的特征，尽管它作用在多个空间位置上。我们通常希望网络的每一次能够在多个位置上提取多种类型的特征。

输入通常不仅仅是实值的网格，而是由一系列观测数据的向量构成的网络。

因为卷积网络通常使用多通道的卷积，所以即使使用了核翻转，也不一定保证网络的线性运算是可交换的。只有当其中的每个运算的输出和输入具有相同的通道数时，这些多通道的运算才是可交换的。

假设有一个 4 为核张量  $K$ ，它的每个元素是  $K_{i,j,k,l}$ ，表示输出中处于通道  $i$  的一个单元和输入中通道  $j$  中的一个单元的连接强度，并且在输入单元之间有  $k$  行  $l$  列的偏置。 $V$  为输入的观测数据，每个元素为  $V_{i,j,k}$ 。设输入  $Z$  和输出  $V$  有相同的形式且不涉及翻转卷积核张量。则输出  $Z$ ：

$$Z_{i,j,k} = \sum_{l,m,n} V_{l,j+m-1,k+n-1} K_{i,l,m,n}$$

我们可以跳过核中的位置来降低计算开销，即看作对全卷积函数输出的下采样 (downsampling)。在输出的每个方向上每间隔  $s$  个像素进行采样，可定义一个下采样卷积函数：

$$Z_{i,j,k} = c(K, V, s)_{i,j,k} = \sum_{l,m,n} [V_{l,(j-1) \times s + m, (k-1) \times s + n} K_{i,l,m,n}]$$

将  $s$  称为下采样卷积的步幅 (stride)

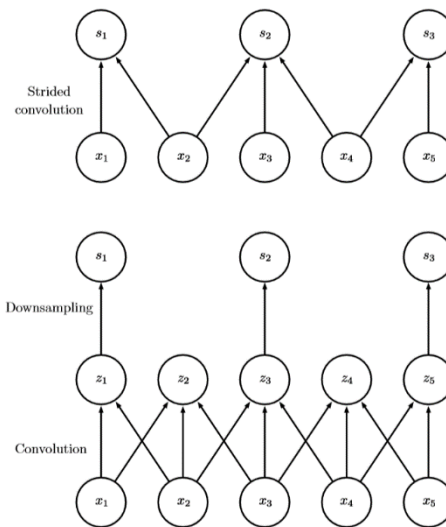


图 9.12: 带有步幅的卷积。在这个例子中，我们的步幅为二。(上)在单个操作中实现的步幅为二的卷积。(下)步幅大于一个像素的卷积在数学上等价于单位步幅的卷积随后降采样。显然，涉及降采样的两步法在计算上是浪费的，因为它计算了许多将被丢弃的值。

在任何卷积网络的实现中都有一个重要性质，即对输入  $V$  用 0 进行填充 (pad) 使之加宽。

有三种零填充设定的情况值得注意：

**有效 (valid) 卷积：**无论怎样都不使用零填充的情况，并且卷积核只允许访问那些图像中能够完全包含整个核的位置。输出的所有像素都是输入中相同数量像素的函数，使得输出像素的表示更加规范，而输出的大小在每一层都会缩减。

**相同 (same) 卷积：**只进行足够的零填充来保持输入和输出具有相同的大小。在这种情况下卷积运算不改变下一层的结构，而输入像素靠近边界的部分相比较于中间部分对于

输出像素的影响更小，可能会导致边界像素存在一定的欠表示

**全（full）卷积：**进行足够多的零填充使得每个像素在每个方向上下好被访问了  $k$  次，最终输出图像的宽度为  $m+k-1$ 。这种情况下，输出像素中靠近边界的部分相比于中间的部分是更少像素的函数。导致学得一个在卷积特征映射的所有位置都表现不错的单核更为困难。

通常零填充的最优数量处于“有效卷积”与“相同卷积”之间。

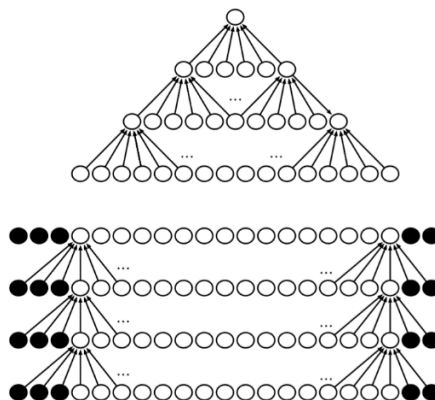


图 9.13: 零填充对网络大小的影响。考虑一个卷积网络，每层有一个宽度为六的核。在这个例子中，我们不使用任何池化，所以只有卷积操作本身缩小网络的大小。(上) 在这个卷积网络中，我们不使用任何隐含的零填充。这使得表示在每层缩小五个像素。从十六个像素的输入开始，我们只能有三个卷积层，并且最后一层不能移动核，所以可以说只有两层是真正的卷积层。可以通过使用较小的核来减缓收缩速率，但是较小的核表示能力不足，并且在这种结构中一些收缩是不可避免的。(下) 通过向每层添加五个隐含的零，我们防止了表示随深度收缩。这允许我们设计一个任意深的卷积网络。

**非共享卷积（unshared convolution）：**具有一个小核的离散卷积运算很像，但并不横跨位置来共享参数，是一些局部连接的网路层，此时多层感知机对应的邻接矩阵是相同的，但每个连接都有它自己的权重。局部连接层的线性部分：

$$Z_{i,j,k} = \sum_{l,m,n} [V_{l,j+m-1,k+n-1} w_{i,j,k,l,m,n}]$$

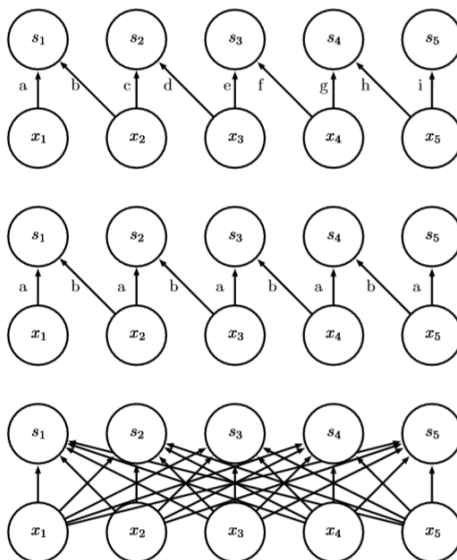


图 9.14: 局部连接，卷积和全连接层的比较。(上) 每一小片（接受域）有两个像素的局部连接层。每条边用唯一的字母标记，来显示每条边都有自身的权重参数。(中) 核宽度为两个像素的卷积层。该模型与局部连接层具有完全相同的连接。区别不在于哪些单元相互交互，而在于如何共享参数。局部连接层没有参数共享。正如用于标记每条边的字母重复出现所指示的，卷积层在整个输入上重复使用相同的两个权重。(下) 全连接层类似于局部连接层，它的每条边都有其自身的参数（在该图中用字母明确标记的话就太多了）。然而，它不具有局部连接层的连接受限的特征。

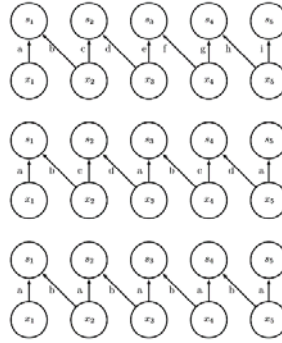
**平铺卷积** (tilled convolution): 学习一组核使得核在空间移动时可以循环利用, 意味着在近邻的位置上拥有不同的过滤器, 但对于这些参数的存储需求仅仅会增长常数倍, 这个常数为核的集合的大小。注意输出的位置在每个方向上在  $t$  个不同的核组成的集合中进行循环。

$$Z_{i,j,k} = \sum_{l,m,n} V_{l,j+m-1,k+n-1} K_{i,l,m,n,j\%t+1,k\%t+1}$$

这里的 $\%$ 代表取模运算, 性质包括  $t\%t=0$ ,  $(t+1)\%t=1$  等。

局部连接层与平铺卷积层与最大池化存在关联: 这些层的探测单元都是由不同的过滤器驱动的。卷积层对于平移具有内置的不变性。

例: 局部连接层、平铺卷积、标准卷积比较:



**关于反向传播:**

需要训练一个卷积网络, 其中包含步幅为  $s$  的步幅卷积, 该卷积的核为  $k$ , 作用于多通道的图像  $V$ , 定义为  $c(K, V, s)$ 。假设需要最小化某个损失函数  $J(V, K)$ 。在前向传播过程中, 需要用  $c$  本身来输出  $Z$ , 然后  $Z$  传递到网络的其余部分并被用来计算损失

函数  $J$ 。在反向传播过程中, 会得到一个张量  $G$  满足  $G_{i,j,k} = \frac{\partial}{\partial Z_{i,j,k}} J(V, K)$ 。

对核中的权重求导:

$$g(\mathbf{G}, \mathbf{V}, s)_{i,j,k,l} = \frac{\partial}{\partial K_{i,j,k,l}} J(\mathbf{V}, \mathbf{K}) = \sum_{m=n} G_{i,m,n} V_{j,(m-1)\times s+k,(n-1)\times s+l}$$

若此层不是网络的底层, 需要对  $V$  求梯度来对误差进一步反向传播。可以使用:

$$\begin{aligned} h(\mathbf{K}, \mathbf{G}, s)_{i,j,k} &= \frac{\partial}{\partial V_{i,j,k}} J(\mathbf{V}, \mathbf{K}) \\ &= \sum_{\substack{l,m \\ \text{s.t.} \\ (l-1)\times s+m=j}} \sum_{\substack{n,p \\ \text{s.t.} \\ (n-1)\times s+p=k}} \sum_q K_{q,i,m,p} G_{q,l,n} \end{aligned}$$

关于如何在偏置项中共享参数:

对于局部连接层, 对每个单元都给定它特有的偏置, 对于平铺卷积, 用与核一样的平铺模式来共享参数。对于卷积层来说, 通常在输出的每个通道上都设置一个偏置, 这个派年至在每个卷积映射的所有位置上共享, 若输入是已知的固定大小, 亦可以在输出映射的每个位置学习一个单独的偏置。

## 2. 结构化输出

卷积神经网络可以用于输出高维的结构化对象, 不仅是预测分类任务的类标签或回归任务的实数值。通常这个对象是一个 tensor, 由标准卷积层产生。

经常出现的一个问题是输出平面可能比输入平面要小。用于对图像中单个对象分类的常用结构中，网络空间维数的最大减少来源于使用大步幅的池化层。

为了产生与输入大小相似的输出，可以避免把池化放在一起，或单纯地产生一张低分辨率的标签网格，原则上可以使用具有单位步幅的池化操作。

对图像逐个像素标记的一种策略是先产生图像标签的原始猜测，然后使用相邻像素之间的交互来修正该原始猜测。（形成一种特殊的循环神经网络）

3. 数据类型

卷积网络使用的数据通常包含多个通道, 每个通道是时间上或空间中某一点的不同观测量。

	单通道	多通道
1 维	音频波形：卷积的输入对应于时间。我们将时间离散化并且在每个时间点上测量一次波形的振幅。	骨架动画 (skeleton animation) 数据：计算机渲染的 3D 角色动画是通过随时间调整“骨架”的姿势而生成的。在每个时间点，角色的姿势通过骨架中的每个关节的角度来描述。我们输入到卷积模型的数据的每个通道，表示一个关节关于一个轴的角度。
2 维	已经使用傅里叶变换预处理过的音频数据：我们可以将音频波形变换成 2 维张量，不同的行对应不同的频率，不同的列对应不同的时间点。在时间轴上使用卷积使模型等效于在时间上移动。在频率轴上使用卷积使模型等效于在频率上移动，这使得在不同入度音频中捕捉的相同旋律产生相同的表示，但处于网络输出中的不同位置。	彩色图像数据：其中一个通道包含红色像素，另一个包含绿色像素，最后一个包含蓝色像素。在图像的水平轴和垂直轴上移动卷积核，赋予了两个方向上半移等变性的。
3 维	体积数据：这种数据一般来源于医学成像技术，例如 CT 扫描等。	彩色视频数据：其中一个轴对应着时间，另一个轴对应着视场轴的高度，最后一个轴对应着视场轴的宽度。

表 9.1 用于卷积网络的不同数据格式的例子。

卷积网络一个有点是可以处理具有可变空间尺度的输入。

注意：使用卷积处理可变尺寸的输入，仅对输入是因为包含对同种事物的不同量的观察而导致的尺寸变化这种情况才有意义。若输入是因为它可以选择性地包括不同种类的观察而具有可变尺寸，使用卷积是不合理的。

4. 高效的卷积算法

卷积等效于使用傅里叶变换将输入与核都转换到频域、执行两个信号的住店相乘，再使用傅里叶逆变换转换回时域。

当一个 d 维的核可以表成 d 个向量（每一维一个向量）的外积时，该核称为**可分离** (separable) 的。当核可分离，它等价于组合 d 个一维卷积，每个卷积使用这些向量中的一个。