

9.1 卷积

卷积网络

- 指那些至少在网络的一层中使用卷积运算来替代一般的矩阵乘法运算的神经网络

卷积运算

- 卷积是对两个实变函数的一种数学运算，通常用星号表示
- $s(t) = (x * w)(t)$ $s(t) = \int x(a)w(t-a)da.$
- 第一个参数是输入，第二个是核函数
- 输出有时被称为特征映射
- 计算机数据会离散化，于是， $s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a)$
- 多维度上卷积运算
$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i-m, j-n)$$
- 卷积是可交换的， $S(i, j) = (K * I)(i, j) = \sum_m \sum_n I(i-m, j-n)K(m, n)$

9.1 卷积

卷积网

•

卷积运

•

•

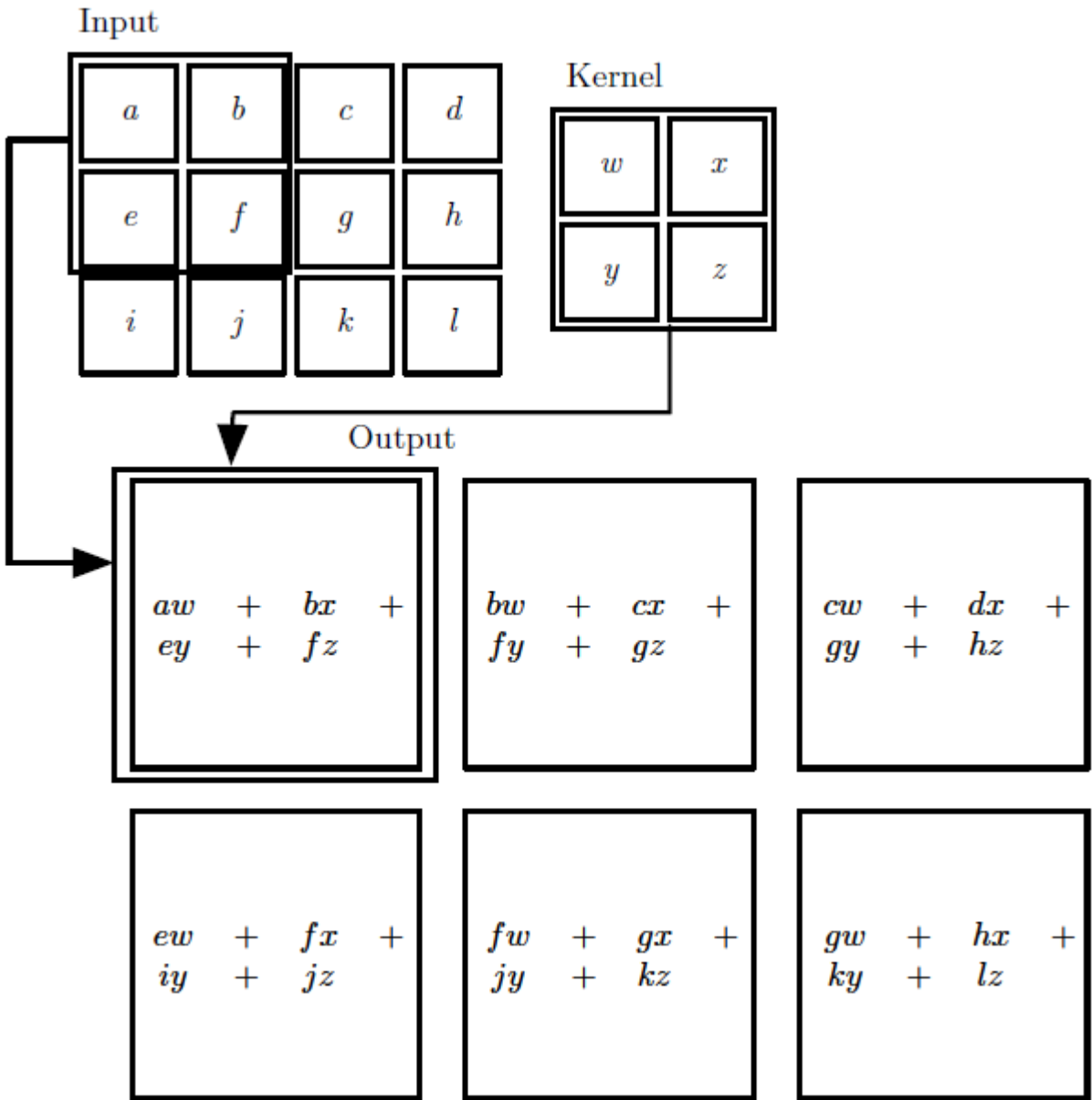
•

•

•

•

•



矩阵乘法

⋮

$- a)da.$

$v(t - a)$

$- n)$

$m, n)$

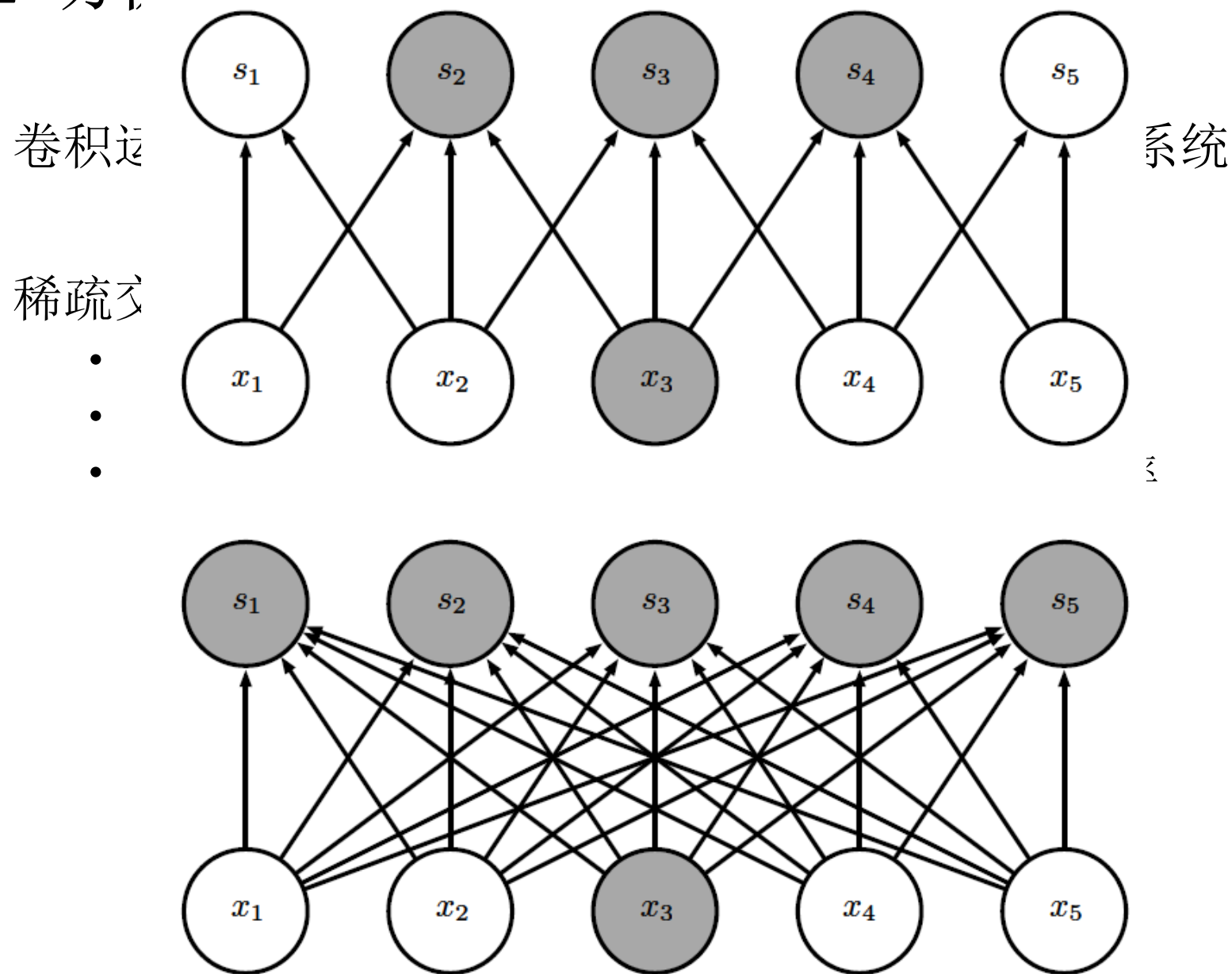
9.2 动机

卷积运算通过三个重要的思想来帮助改进机器学习系统

稀疏交互

- 稀疏连接、稀疏权重
- 使核的大小远小于输入的大小来达到的
- 不仅减少了模型的存储需求，而且提高了它的统计效率

9.2 动机



9.2 动机

卷积运算通过三个重要的思想来帮助改进机器学习系统

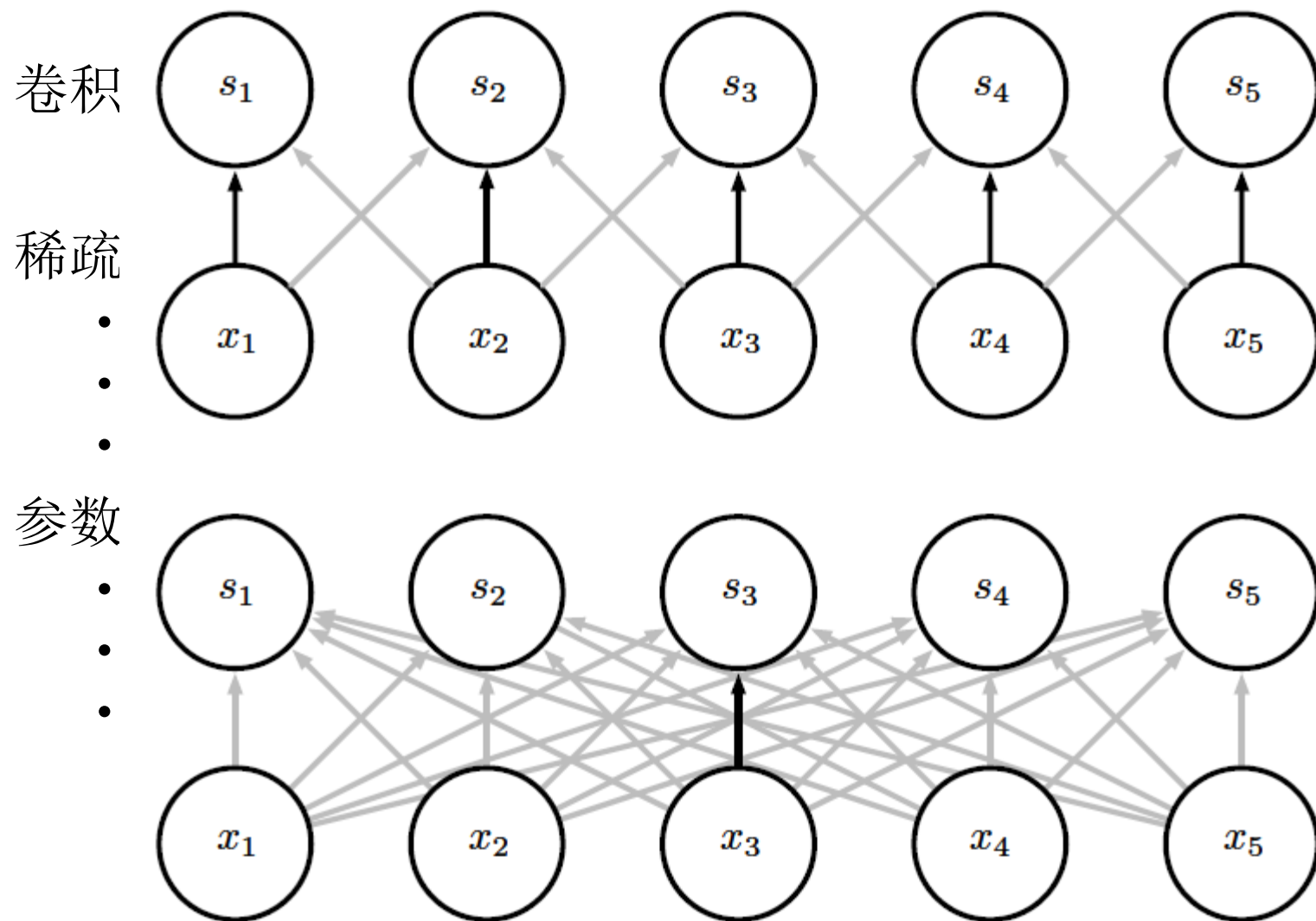
稀疏交互

- 稀疏连接、稀疏权重
- 使核的大小远小于输入的大小来达到的
- 不仅减少了模型的存储需求，而且提高了它的统计效率

参数共享

- 绑定的权重
- 在一个模型的多个函数中使用相同的参数
- 显著地降低了模型的存储需求

9.2 动机



9.2 动机

卷积运算通过三个重要的思想来帮助改进机器学习系统

稀疏交互

- 稀疏连接、稀疏权重
- 使核的大小远小于输入的大小来达到的
- 不仅减少了模型的存储需求，而且提高了它的统计效率

参数共享

- 绑定的权重
- 在一个模型的多个函数中使用相同的参数
- 显著地降低了模型的存储需求

等变表示

- 参数共享的特殊形式使得神经网络层具有对平移等变
- 一个函数满足输入改变，输出也以同样的方式改变

9.2 动机

稀疏连接和参数共享可以显著提高线性函数在一张图像上进行边缘检测的效率的

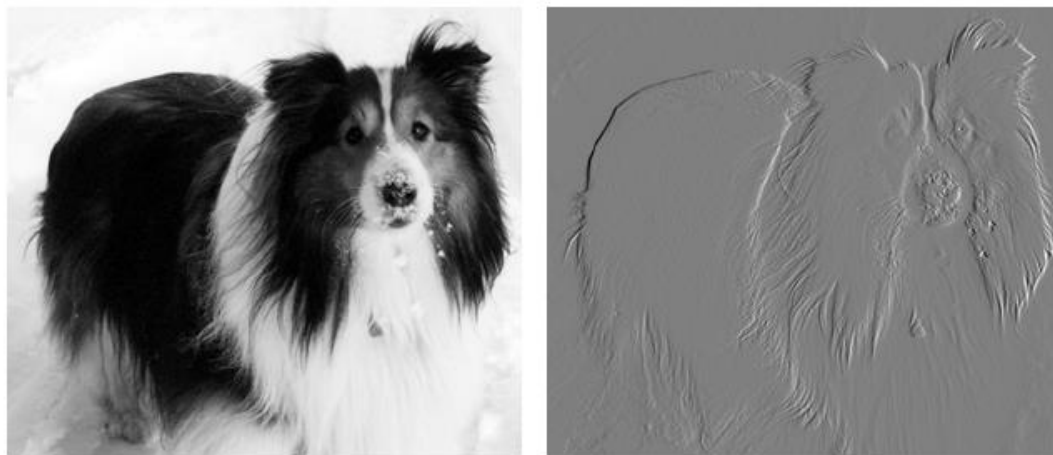


图 9.6: 边缘检测的效率。右边的图像是通过先获得原始图像中的每个像素, 然后减去左边相邻像素的值而形成的。这个操作给出了输入图像中所有垂直方向上的边缘的强度, 对目标检测来说是有用的。两个图像的高度均为 280 个像素。输入图像的宽度为 320 个像素, 而输出图像的宽度为 319 个像素。这个变换可以通过包含两个元素的卷积核来描述, 使用卷积需要 $319 \times 280 \times 3 = 267,960$ 次浮点运算 (每个输出像素需要两次乘法和一次加法)。为了用矩阵乘法描述相同的变换, 需要一个包含 $320 \times 280 \times 319 \times 280$ 个或者说超过 80 亿个元素的矩阵, 这使得卷积对于表示这种变换更有效 40 亿倍。直接运行矩阵乘法的算法将执行超过 160 亿次浮点运算, 这使得卷积在计算上大约有 60,000 倍的效率。当然, 矩阵的大多数元素将为零。如果我们只存储矩阵的非零元, 则矩阵乘法和卷积都需要相同数量的浮点运算来计算。矩阵仍然需要包含 $2 \times 319 \times 280 = 178,640$ 个元素。将小的局部区域上的相同线性变换应用到整个输入上, 卷积是描述这种变换的极其有效的方法。照片来源: Paula Goodfellow。

9.3 池化

卷积网络中一个典型层包含三级

- 第一级：并行地计算多个卷积产生一组线性激活响应
- 探测级：每一个线性激活响应将会通过一个非线性的激活函数
- 第三级：使用池化函数来进一步调整这一层的输出

池化函数

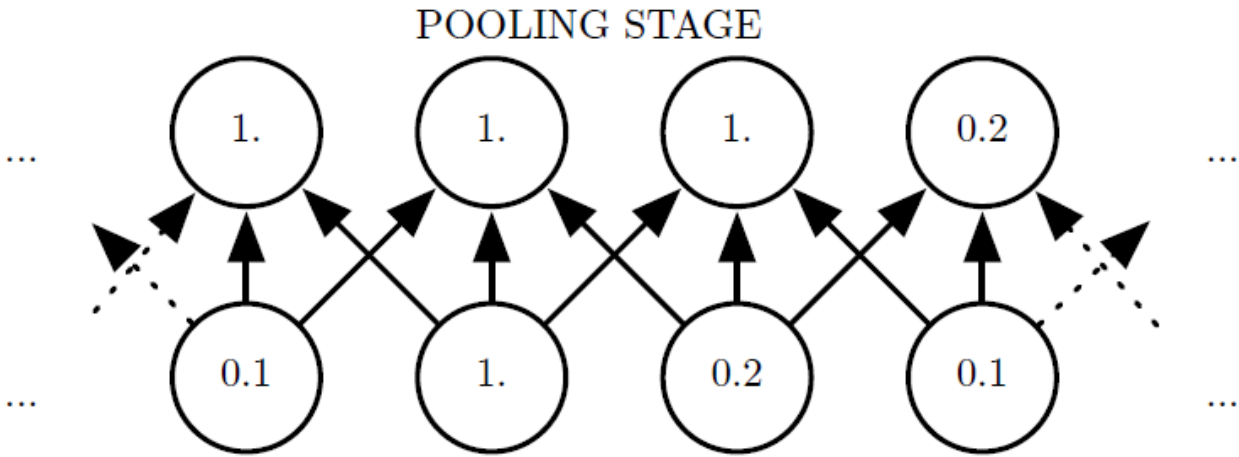
- 使用某一位置的相邻输出的总体统计特征来代替网络在该位置的输出
- 最大池化：给出相邻矩形区域内的最大值
- 相邻矩形区域内的平均值、 L^2 范数以及基于据中心像素距离的加权平均函数

局部平移不变性

- 当输入作出少量平移时，池化能够帮助输入的表达近似不变

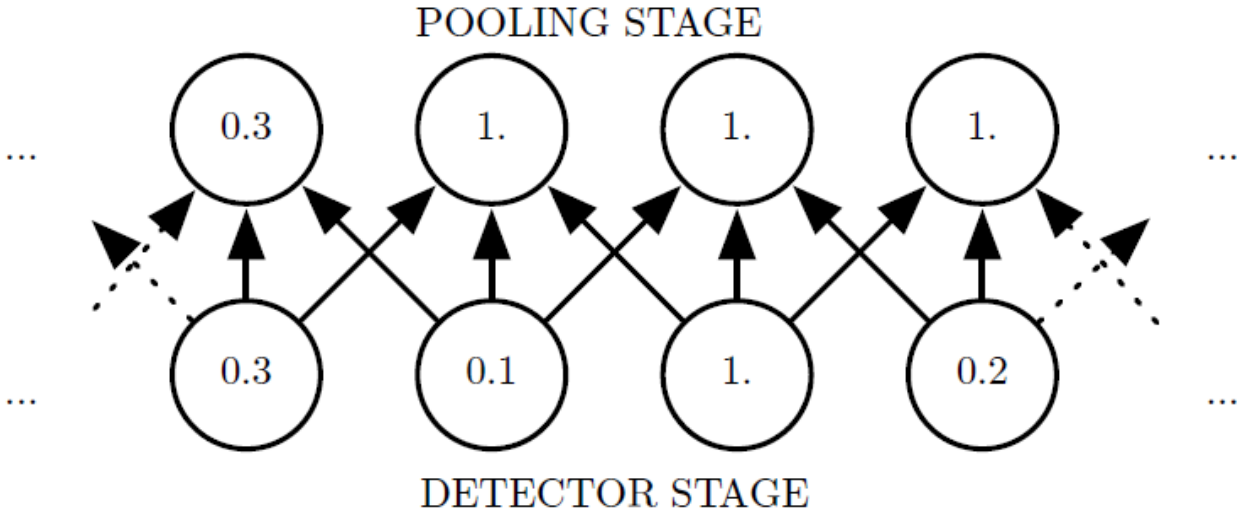
9.3 池化

卷积



活函数

池化



该位置的

距离的加

局音

不变

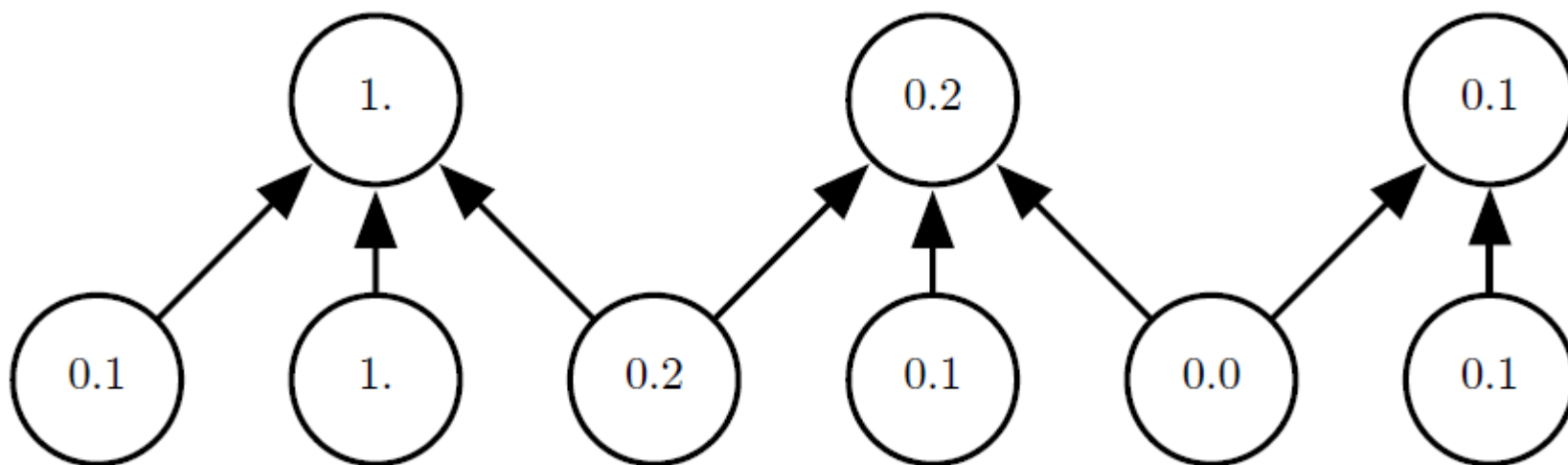
9.3 池化

使用池化可以看作是增加了一个无限强的先验

- 这一层学得的函数必须具有对少量平移的不变性

假设成立时

- 池化可以极大地提高网络的统计效率
- 池化综合了全部邻居的反馈，所以池化单元少于探测单元



9.4 卷积与池化作为一种无限强的先验

先验被认为是强或者弱取决于先验中概率密度的集中程度

- 弱先验具有较高的熵值
- 强先验具有较低的熵值

一个无限强的先验需要对一些参数的概率置零并且完全禁止对这些参数赋值，无论数据对于这些参数的值给出了多大的支持

把卷积网络类比为全连接网络

- 一个隐藏单元的权重必须和它邻居的权重相同，但可以在空间上移动
- 除了那些处在隐藏单元的小的空间连续的接受域内的权重以外，其余的权重都为零

卷积和池化可能导致欠拟合

- 如果一项任务依赖于保存精确的空间信息，在所有的特征上使用池化将会增大训练误差

9.5 基本卷积函数的变体

神经网络中的卷积

- 指由多个并行卷积组成的

输入

- 输入通常也不仅仅是实值的网格
- 由一系列观测数据的向量构成的网格

假设:

- 4维核张量 \mathbf{K}
- 输入 \mathbf{V}
- 输出 \mathbf{Z} 是通过 \mathbf{K} 和 \mathbf{V} 进行卷积
$$Z_{i,j,k} = \sum_{l,m,n} V_{l,j+m-1,k+n-1} K_{i,l,m,n}$$
- 跳过核中的一些位置来降低计算的开销

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

9.5 基本卷积函数的变体

任何卷积网络的实现中都有一个重要性质

- 能够隐含地对输入 \mathbf{v} 用零进行填充使得它加宽

有三种零填充设定的情况值得注意

- 无论怎样都不使用零填充的极端情况，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，这使得输出像素的表示更加规范
- 只进行足够的零填充来保持输出和输入具有相同的大小
 - 只要硬件支持，网络就能包含任意多的卷积层，这是因为卷积运算不改变下一层的结构
- 足够多的零填充使得每个像素在每个方向上恰好被访问了 k 次，最终输出图像的宽度为 $m + k - 1$
 - 输出像素中靠近边界的部分相比于中间部分是更少像素的函数

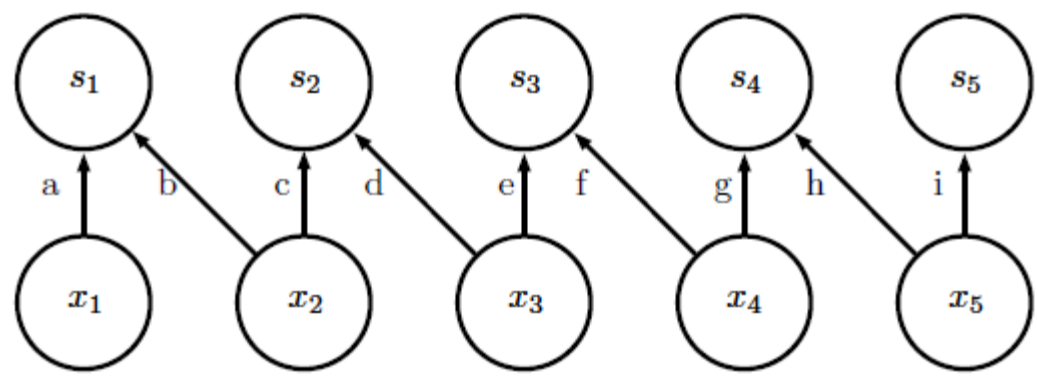
非共享卷积

- 只用一些局部连接的网络层，不使用卷积

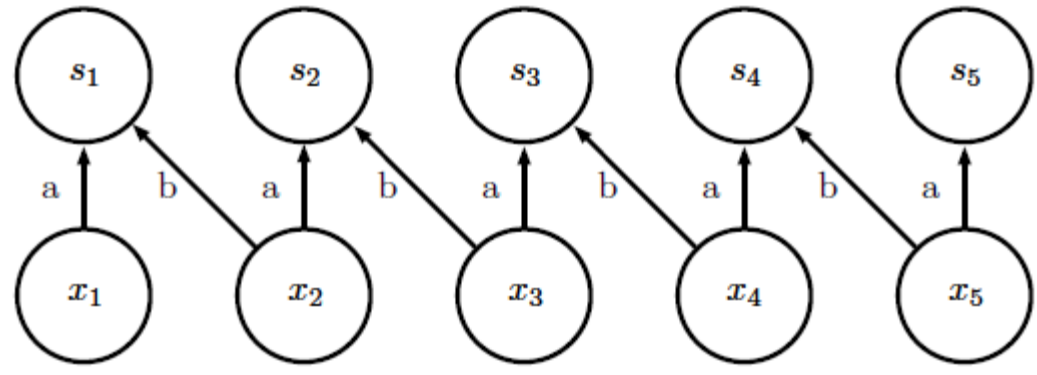
9.5 基本卷积和新的变体

任何卷积网

- 能够降维
- 有三种零填充



- 无论怎样这些图例
- 输出的大小
- 只进行卷积
- 只不改变大小

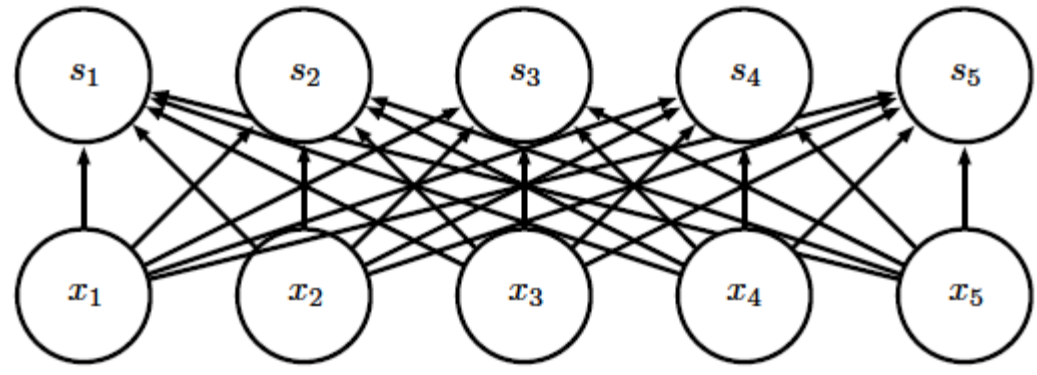


只允许访问那
使得输出像素
大小
因为卷积运算

- 足够多最终输出
- 输出像素的函数

非共享卷积

- 只用一



访问了k次，
像素的函数

9.5 基本卷积函数的变体

平铺卷积

- 对卷积层和局部连接层进行了折衷
- 学习一组核使得当我们在空间移动时它们可以循环利用

转置化卷积

- 类似核梯度运算
- 转置运算返回的输出的大小取决于三个方面
 - 零填充的策略
 - 前向传播运算的步幅
 - 前向传播的输出映射的大小

9.5 基本卷积函数的变体

训练神经网络

- 步幅为 s 的步幅卷积、卷积核为 \mathbf{K}
- 输入是多通道图像 \mathbf{V} 、输出为 $\mathbf{Z} = \mathbf{c}(\mathbf{K}; \mathbf{V}; s)$
- 损失函数 $J(\mathbf{V}; \mathbf{K})$
- 前向传播时， \mathbf{Z} 递到网络的其余部分并且被用来计算损失函数 J
- 反向传播时，得到张量 \mathbf{G} ， $G_{i,j,k} = \frac{\partial}{\partial Z_{i,j,k}} J(\mathbf{V}, \mathbf{K})$
- 对核中的权重求导

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

- 如果这一层不是网络的底层，我们需要对 \mathbf{V} 求梯度来使得误差进一步反向传播

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

9.6 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

- 这个对象只是一个张量，由标准卷积层产生
- 但是，输出平面可能比输入平面要小

为了产生与输入大小相似的输出映射

- 避免把池化放在一起
- 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

- 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
- 重复这个修正步骤数次对应于在每一步使用相同的卷积，该卷积在深层网络的最后几层之间共享权重

9.7 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：卷积的轴对应于时间。我们将时间离散化并且在每个时间点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	已经使用傅立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得模型等效于在频率上移动，这使得在不同八度音阶中播放的相同旋律产生相同的表示，但处于网络输出中的不同高度。	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。在图像的水平轴和竖直轴上移动卷积核，赋予了两个方向上平移等变性。
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，另一个轴对应着视频帧的高度，最后一个对应着视频帧的宽度。

9.7 数据类型

卷积网络的一个优点是它们还可以处理具有可变的空间尺度的输入

考虑一组图像的集合，其中每个图像具有不同的高度和宽度

- 卷积就可以很直接地应用
- 核依据输入的大小简单地被使用不同次，并且卷积运算的输出也相应地放缩
- 卷积可以被视为矩阵乘法
- 相同的卷积核为每种大小的输入引入了一个不同大小的双重分块循环矩阵

9.8 高效的卷积算法

现代卷积网络的应用通常需要包含超过百万个单元的网络

- 并行计算资源
- 通过选择适当的卷积算法来加速卷积

可分离卷积

- 当一个 d 维的核可以表示成 d 个向量的外积时，该核被称为可分离的
- 当核可分离时，朴素的卷积是低效的
- 如果核在每一维都是 w 个元素宽，那么朴素的多维卷积需要 $O(w^d)$ 的运行时间和参数存储空间
- 此时，朴素卷积等价于组合 d 个一维卷积，每个卷积使用这些向量中的一个
- 可分离卷积只需要 $O(w \times d)$ 的运行时间和参数存储空间
- 但是，并不是每个卷积都可以表示成这种形式

9.9 随机或无监督的特征

卷积网络训练中最昂贵的部分是学习特征

- 减少卷积网络训练成本的一种方式是使用那些不是由监督方式训练得到的特征

有三种基本策略可以不通过监督训练而得到卷积核

- 简单地随机初始化
- 手动设计
 - 设置每个核在一个特定的方向或尺度来检测边缘
- 使用无监督的标准来学习核
 - 使用无监督的标准来学习特征，允许这些特征的确定与位于网络结构顶层的分类层相分离

卷积网络为我们提供了相对于多层感知机更进一步采用预训练策略的机会

- 可以训练一小块模型
- 用来自这个小块模型的参数来定义卷积层的核

9.10 卷积网络的神经科学基础

卷积网络层被设计为描述初级视觉皮层 **V1** 的三个性质

- **V1** 可以进行空间映射，卷积网络通过用二维映射定义特征的方式来描述该特性
- **V1** 包含许多简单细胞，卷积网络的检测器单元被设计为模拟简单细胞的这些性质
- **V1** 还包括许多复杂细胞，对于特征的位置微小偏移具有不变性，这激发了卷积网络中的一些跨通道池化策略，例如 **maxout** 单元

未解决的问题

- 人眼分辨率低，场景是大脑产生的错觉——扫视之后，拼接小区域
- 人类视觉系统集成了许多其他感觉，神经网络仅仅是视觉
- 视觉可以理解场景
- 简单的大脑区域也受到来自较高级别的反馈的严重影响，神经网络的反馈进展缓慢
- 前馈IT放电频率刻画了与卷积网络特征很多相同的信息，但是仍不清楚中间计算的相似程度

9.11 卷积网络与深度学习的历史

卷积网络在深度学习的历史中发挥了重要作用

- 是将研究大脑获得的深刻理解成功用于机器学习应用的关键例子
- 是首批表现良好的深度模型之一
- 是第一个解决重要商业应用的神经网络
 - 20 世纪90 年代，AT&T 的神经网络研究小组开发了一个用于读取支票的卷积网络
 - 90 年代末，NEC 部署的这个系统已经被用于读取美国10%以上的支票
 - 后来，微软部署了若干个基于卷积网络的OCR 和手写识别系统
- 卷积网络也被用作在许多比赛中的取胜手段
- 卷积网络是第一批能使用反向传播有效训练的深度网络之一
- 卷积网络提供了一种方法来特化神经网络，使其能够处理具有清楚的网格结构拓扑的数据，以及将这样的模型扩展到非常大的规模