



云栖社区 (/) > 翻译小组 (/teams/132) > 博客 (/teams/132/type_blog) > 正文

://service.weibo.com/share/share.php?

=%E3%80%90%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0%E4%B

(https://www.aliyun.com/acts/product-section-2019/home)

【深度学习之美】局部连接来减参，权值共享肩并肩（入门系列之十一）

【方向】 (/users/5m4kk67c65aqq)

🕒 2017-08-06 17:41:23

👁 浏览20818

💬 评论4

云栖社区 (/tags/type_blog-tagid_1/)

深度学习 (/tags/type_blog-tagid_16/)

算法 (/tags/type_blog-tagid_37/)

神经网络 (/tags/type_blog-tagid_13435/)

卷积神经网络 (/tags/type_blog-tagid_15900/)

品味大数据 (/tags/type_blog-tagid_17970/)

张玉宏 (/tags/type_blog-tagid_18831/)

权值共享 (/tags/type_blog-tagid_20181/)

局部连接 (/tags/type_blog-tagid_20182/)

摘要：《战狼2》中吴京的那句“那他妈是以前！”令人热血沸腾。在以前，全连接网络的扩展性很差，原因就在于它难以承受参数太多之痛。但当卷积神经网络兴起后，我们也可以说一句，“那他妈是以前！”因为现在我们有“局部连接”和“权值共享”。来，看看它们是如何工作的吧？

- 系列文章：
- 一入侯门“深”似海，深度学习深几许（深度学习入门系列之一） (https://yq.aliyun.com/articles/86580)

人工“破”索意犹尽，智能“硅”来未可知（深度学习入门系列之二） (https://yq.aliyun.com/articles/88300)

神经网络不胜数，M-P模型似可寻（深度学习入门系列之三） (https://yq.aliyun.com/articles/90565)

“机器学习”三重门，“中庸之道”趋若人（深度学习入门系列之四） (https://yq.aliyun.com/articles/91436)

Hello World感知机，懂你我心才安息（深度学习入门系列之五） (https://yq.aliyun.com/articles/93540)

损失函数减肥用，神经网络调权重（深度学习入门系列之六） (https://yq.aliyun.com/articles/96427)

山重水复疑无路，最快下降问梯度（深度学习入门系列之七） (https://yq.aliyun.com/articles/105339)

BP算法双向传，链式求导最缠绵（深度学习入门系列之八） (https://yq.aliyun.com/articles/110025)

全面连接困何处，卷积网络见解深（深度学习入门系列之九） (https://yq.aliyun.com/articles/152935)

卷地风来忽吹散，积得飘零美如画（深度学习入门系列之十） (https://yq.aliyun.com/articles/156269)

很多年前，著名物理学家爱因斯坦说过一句名言：Everything should be made as simple as possible, but not simpler（越简单越好，但是还不能过分简单）。"是的，只有名人才能说名言。如果这句话是我的说的，你可能认为这不是废话吗？

我把爱老爷子搬出来，自然不是想唬你，而是因为他的名言和我们今天要讲的主题有点相关。我们知道，相比于全连接的前馈网络，卷积神经网络的结构要简单得多，可是它并不是那么简单，否则也不会有这么多初学者对卷积神经网络“望而生畏”。

卷积神经网络之所以这么成功，套用爱老爷子的观点来说，它简单得“恰如其分”。在前面的章节中，我们重点讨论了卷积的概念以及卷积核在图像处理中的应用。在本章，我们将重点讨论它“恰如其分”的拓扑结构，一旦理解清楚它的设计原理，再动手在诸如Theano或Keras等深度学习框架下，写个卷积神经网络的实战小Demo，聪慧如你，一定不在话下。

11.1 卷积神经网络的拓扑结构

下面我们先感性认识一下卷积神经网络中的几个重要结构，如图11-1所示。在不考虑输入层的情况下，一个典型的卷积神经网络通常由若干个卷积层（Convolutional Layer）、激活层（Activation Layer）、池化层（Pooling Layer）及全连接层（Fully Connected Layer）组成。

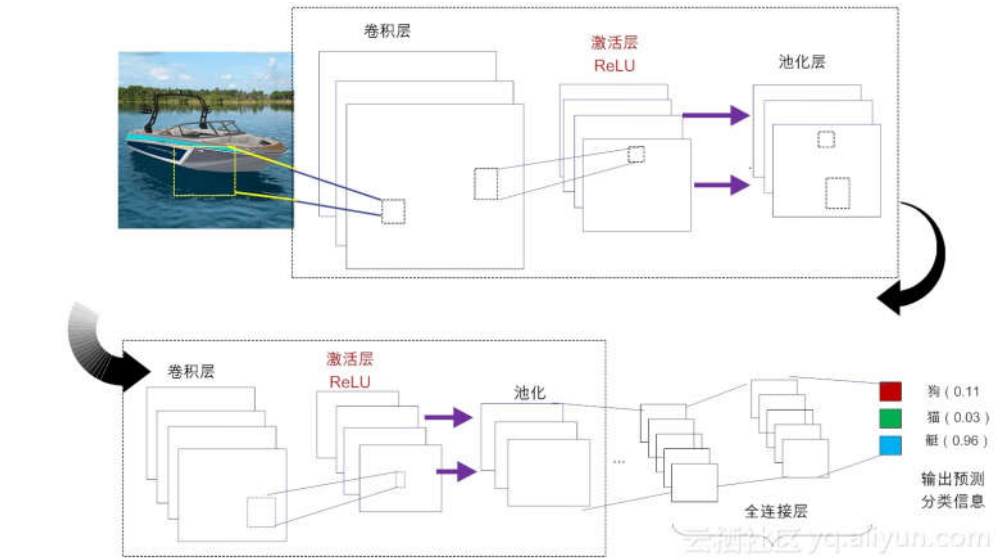


图11-1 典型卷积神经网络的结构

下面先给予简单地介绍，后文会逐个进行详细介绍：

卷积层：这个是卷积神经网络的核心所在（作为数学概念，我们在第10章已做了详细介绍。不熟悉的读者可前往查阅 (<https://yq.aliyun.com/articles/156269>) [1]）。在卷积层，通过实现“局部感知”和“权值共享”等系列的设计理念，可达到两个重要的目的：对高维输入数据实施降维处理和实现自动提取原始数据的核心特征。

激活层：其作用是将前一层的线性输出，通过非线性激活函数处理，从而可模拟任意函数，进而增强网络的表征能力。在深度学习领域，ReLU（Rectified-Linear Unit，修正线性单元）是目前使用较多的激活函数，原因是它收敛更快，且不会产生梯度消失问题。

池化层：亦称亚采样层（Subsampling Layer）。简单来说，利用局部相关性，“采样”在较少数据规模的同时保留了有用信息。巧妙的采样还具备局部线性转换不变性，从而增强卷积神经网络的泛化处理能力。

全连接层：这个网络层相当于传统的多层感知机（Multi-Layer Perceptron，简称MLP，例如我们已经讲解过的BP算法 (<https://yq.aliyun.com/articles/110025>) [2]）。通常来说，“卷积-激活-池化”是一个基本的处理栈，通过多个前栈处理之后，待处理的数据特性已有了显著变化：一方面，输入数据的维度已下降到可用“全连接”网络来处理了；另一方面，此时全连接层的输入数据已不再是“泥沙俱下、鱼龙混杂”，而是经过反复提纯过的结果，因此最后输出的结果要可控得高。

事实上，我们还可以根据不同的业务需求，构建出不同拓扑结构的卷积神经网络，常见架构模式如图11-2所示。

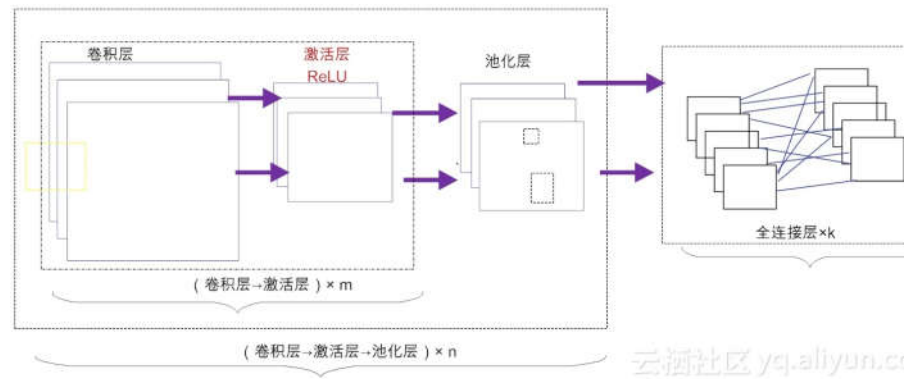


图11-2 卷积神经网络的拓扑结构

也就是说，可以先由 m 个卷积层和激活层叠加，然后(可选)进行一次池化操作，重复这个结构 n 次，最后叠加 k 个全连接层 ($m, n, k \geq 1$)。总体来讲，卷积神经网络通过多层处理，逐渐将初始的“低层”特征表示，转化为“高层”特征表示，之后再“简单模型”即可完成复杂的分类等学习任务。因此在本质上，深度学习就是一个“特征学习 (feature learning)”或“表示学习 (representation learning)”[3]。

下面我们将——讲解卷积神经网络中这几个层的设计理念。在讲解之前，我们有必要补充介绍一下大名鼎鼎的CIFAR-10图像集，因为后面的案例会频频用到这个数据集。

11.2不得不提的 CIFAR-10图像集

CIFAR-10最早是由Hinton教授带领他的两名学生Alex Krizhevsky与Vinod Nair等人一起收集的微型图像集。该图像集由60,000张 32×32 的RGB彩色图片构成，共10个大分类，其中50,000张图片用作训练，另外随机抽取10,000张用作测试（交叉验证）。如图11-3所示[4]。

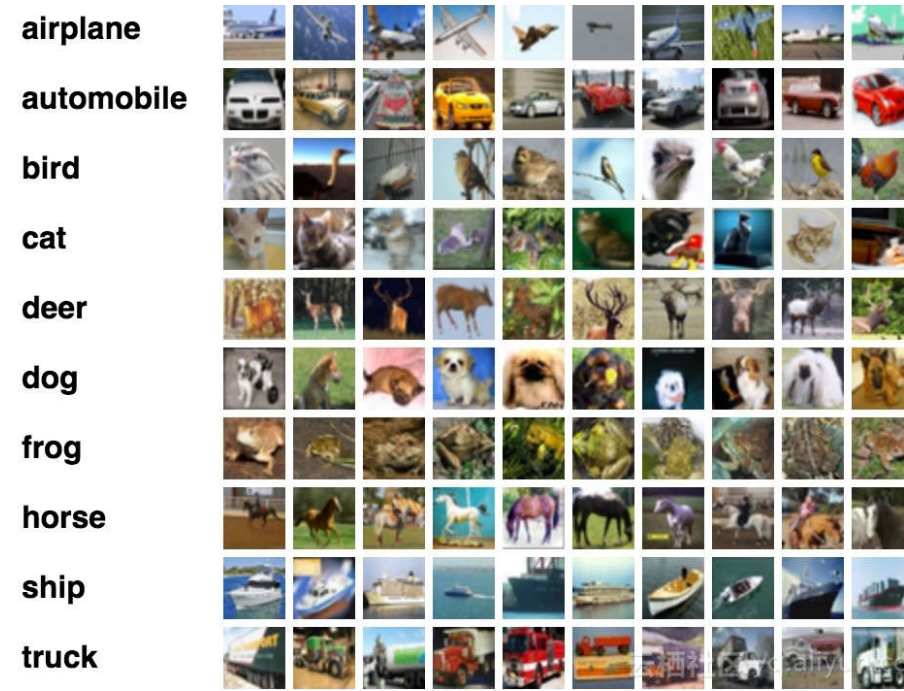


图11-3 从CIFAR-10 (<http://www.cs.toronto.edu/~kriz/cifar.html>)随机抽取的10类图像

CIFAR-10最大的特点莫过于，它将识别的范围扩大到普适物体。相比于已经非常成熟的人脸识别，普适物体的识别更具有挑战性，因为普适图像数据集中含有各式各样的特征，甚至噪声，而且图像中的物体大小比例不一，这些都无疑增加了普适物体判别的难度。

CIFAR-10在深度学习等领域非常有影响力。这是因为它是很多人“深度学习”实战的起点（比如说，Theano、TensorFlow等深度学习框架都常用到这个数据集来做实战练习）。它的存在，在客观上加速推动了“深度学习”的普及进程。可以说，Hinton教授的功劳，不仅仅体现他对深度学习算法的创新上，还体现于他对深度学习的普及上。

11.3 卷积层的3个核心概念

有了上面的工作的铺垫，下面我们来聊聊卷积层的三个核心概念：局部连接、空间位置排列及权值共享。

11.3.1 局部连接

前面我们也提到过，全连接的前馈神经网络有个非常致命的缺点，那就是可扩展性（Scalability）非常差。原因非常简单，网络规模一大，需要调参的个数以神经元数的平方倍增，导致它难以承受参数太多之痛。

局部连接（Local Connectivity）在能某种程度上缓解这个“参数之痛”。下面我们以CIFAR-10图像集为输入数据，来探究一下局部连接的工作原理。

通过前面的介绍可知，每一幅CIFAR-10图像都是 $32 \times 32 \times 3$ 的RGB图像。对于隐藏层的某个神经元，如果是全连接前馈网络中，它不得不和前一层的所有神经元（ 32×32 ）都保持连接。

但现在，对于卷积神经网络而言，隐藏层的这个神经元仅仅需要与前向层的部分区域相连接。这个局部连接区域有个特别的名称叫“感知域（receptive field）”，其大小等同于卷积核的大小（比如说 5×5 ），如图11-4所示。相比于原来的 32×32 连接个数，变成现在的 5×5 个连接，连接的数量自然是稀疏得多，因此，局部连接也被称为“稀疏连接（Sparse Connectivity）”。

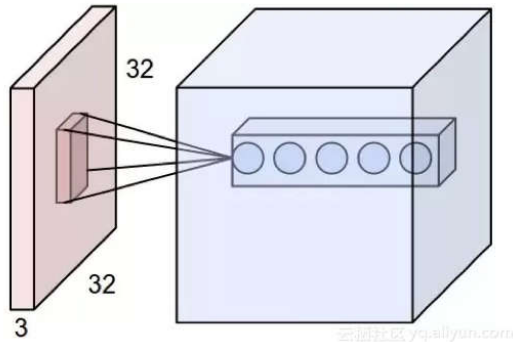


图11-4 局部连接示意图

但需要注意的是，这里的稀疏连接，仅仅是指卷积核的感知域（ 5×5 ）相对于原始图像的高度和宽度（ 32×32 ）而言的。卷积核的深度（depth）则需要与原始数据保持一致，不能缩减。在这里，卷积核的深度实际上就是卷积核的个数。对于RGB图像而言，如果我们需要在红色、蓝色和绿色等三个通道提取特征，那么卷积核个数就是3）。所以对于隐藏层的某个神经元，它的前向连接个数是由全连接的 $32 \times 32 \times 3$ 个，通过卷积操作，减少到局部连接的到 $5 \times 5 \times 3$ 个。

请读者思考，为了提取更多特征，如果卷积核的深度不是3个，而是100个，又会发生什么？很显然，这样一来，局部连接带来的参数个数减少量，就要大打折扣。

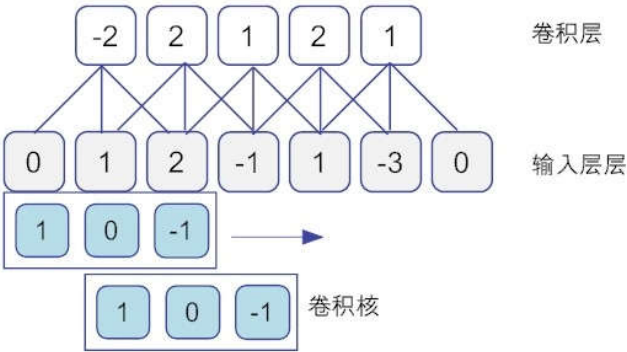
11.3.2 空间排列

在讲解完毕局部连接的原理之后。下面我们来谈谈决定卷积层的空间排列（Spatial arrangement）的4个参数，它们分别是：卷积核的大小、深度、步幅及补零。其中，卷积核的大小（通常是 3×3 或 5×5 的方矩阵）我们已经在第10章 (<https://yq.aliyun.com/articles/156269>) 讨论了，这里仅仅对另外三个结构进行说明[5]。

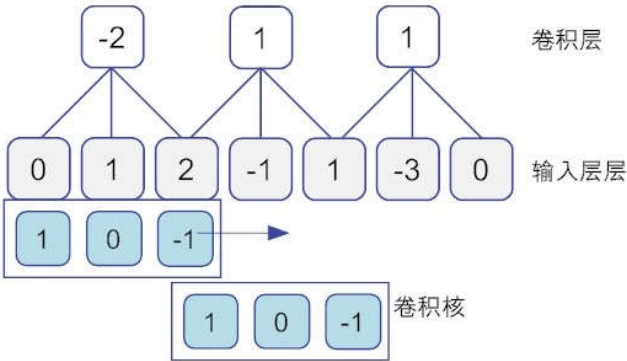
（1）卷积核的深度（depth）：卷积核的深度对应的是卷积核的个数。每个卷积核只能提取输入数据的部分特征。每一个卷积核与原始输入数据执行卷积操作，会得到一个卷积特征，这样的多个特征汇集在一起，我们称为特征图谱。在图11-1所示的示例中（左上部），我们使用三个不同的滤波器（即卷积核）对原始图像进行卷积操作，这样就可以生成三个不同的特征图。你可以把这三个特征图看作是堆叠在一起的2D（二维）矩阵。

事实上，每个卷积核提取的特征都有各自的侧重点。因此，通常说来，多个卷积核的叠加效果要比单个卷积核的分类效果要好得多。例如在2012年的ImageNet竞赛中，Hinton教授和他的学生Krizhevsky等人打造了第一个“大型的深度卷积神经网络”，也即现在众所周知的AlexNet。在这个夺得冠军的深度学习算法中，他们使用的卷积核高达96个！可以说，自那时起，深度卷积神经网络一战成名，才逐渐被世人瞩目。

（2）步幅（stride）：即滤波矩阵在输入矩阵上滑动跨越的单元个数。设步幅大小为S，当S为1时，滤波器每次移动一个像素的位置。当S为2时，每次移动滤波器会跳过2个像素。S越大，卷积得到特征图就越小。以一维数据为例，当卷积核为 $[1, 0, -1]$ ，输入矩阵为 $[0, 1, 2, -1, 1, -3, 0]$ 时，图11-5显示了步幅分别为1和2卷积层的神经元分布情况。



(a) 步幅S=1，卷积层神经元分布

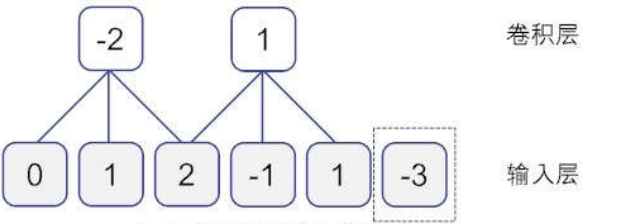


(b) 步幅S=2，卷积层神经元分布 [云栖社区 yq.aliyun.com](https://yq.aliyun.com)

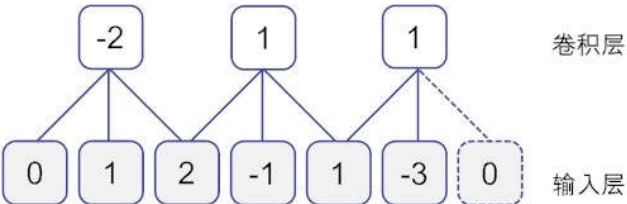
图11-5 当步幅为1和2时，输入层和卷积层的神经元空间分布

(3) 补零 (zero-padding)：补零操作通常用于边界处理。在有些场景下，卷积核的大小并不一定刚好就被输入数据矩阵的维度大小整除。因此，就会出现卷积核不能完全覆盖边界元素的情况。这时，我们就需要在输入矩阵的边缘使用零值进行填充，使得在输入矩阵的边界处的大小刚好和卷积核大小匹配。这样做的结果，相当于对输入图像矩阵的边缘进行了一次滤波。零填充的好处在于，它可以让我们控制特征图的大小。使用零填充的卷积叫做泛卷积 (wide convolution)，不适用零填充的叫做严格卷积 (narrow convolution)。

下面我们举例说明这个概念。假设步幅S的大小为2，为了简单起见，我们假设输入数据为一维矩阵 [0, 1, 2, -1, 1, -3]，卷积核为[1, 0, -1]，在卷积核滑动两次之后，此时输入矩阵边界多余一个“-3”，不够滑动第3次，如图11-6(a)所示。此时，便可以在输入矩阵填入额外的0元素，使得输入矩阵变成[0, 1, 2, -1, 1, -3, 0]，这样一来，所有数据都能得到处理，如图11-6(b)所示。。



(a) 边界有剩余元素



(b) 边界补零 [云栖社区 yq.aliyun.com](https://yq.aliyun.com)

图11-6 在输入矩阵边界处补零

综上所述，在构造卷积层时，对于给定的输入数据，如果确定了卷积核的大小，卷积核的深度（个数）、步幅以及补零个数，那么卷积层的空间安排就能确定下来。以一维数据为例，假设数据的大小（数据元素的长度）为 W ，卷积核的深度为 F ，步幅大小为 S ，补零的数目为 P ，那么对于每个卷积核，在它输入数据实施卷积操作后得到特征图谱，它包含的神经元个数 N 可以用（11-1）公式计算得到。

$$N = \lfloor (W - F + 2P) / S \rfloor + 1 \quad (11-1)$$

对于高维数据而言，对每一个维度的数据均按照（11-1）计算即可。

11.3.3 权值共享

卷积层设计的第三个核心概念就是权值共享（Shared Weights），由于这些权值实际上就是不同神经元之间的连接参数，所以有时候，也将权值共享称为参数共享（Parameter Sharing）。

为什么要设置权值共享呢？其实这也是无奈之举。前文我们提到，通过局部连接处理后，神经元之间的连接个数已经有所减少。可到底减少多少呢？还以CIFAR-10数据集为例，一个原始的图像大小为 $32 \times 32 \times 3$ ，假设我们有100个卷积核，每个卷积核的大小为 $5 \times 5 \times 3$ ，步幅为1，没有补零。先单独考虑一个卷积核，将公式11-1扩展到二维空间，可以很容易计算得到每一个卷积核对应的特征图谱大小是 28×28 。也就是说，这个特征图谱对应有 28×28 神经元。而每个神经元以卷积核大小（ $5 \times 5 \times 3$ ）连接前一层的“感知域（receptive field）”，也就是说，它的连接参数个数为 $(28 \times 28) \times (5 \times 5 \times 3)$ 。如果考虑所有的100个卷积核，（在不考虑偏置参数的情况下）连接的参数个数为 $(5 \times 5 \times 3) \times (28 \times 28) \times 100 = 5,880,000$ 。

那么全连接参数个数又是多少呢？仅仅考虑两层网络的情况下，其连接个数为 $(32 \times 32 \times 3) \times (32 \times 32 \times 3) = 9,437,184$ 。对比这二者的数字可以发现，局部连接虽然降低了连接的个数，但整体幅度并不大，需要调节的参数个数依然非常庞大，因此还是无法满足高效训练参数的需求。

而权值共享就是来解决这个问题的，它能显著降低参数的数量。该如何理解权值共享呢？首先从生物学意义上来看，相邻神经元的活性相似，从而可以它们共享相同的连接权值。

其次单从数据特征上来看，我们可以把每个卷积核（即过滤核）当作一种特征提取方式，而这种方式与图像等数据的位置无关。这就意味着，对于同一个卷积核，它在一个区域提取到的特征，也能适用于其他区域。基于权值共享策略，将卷积层神经元与输入数据相连，同属于一个特征图谱的神经元，将共用一个权值参数矩阵，如图11-7所示。经过权值共享处理后，CIFAR-10的连接参数一下子锐减为 $5 \times 5 \times 3 \times 1 \times 100 = 7500$ 。

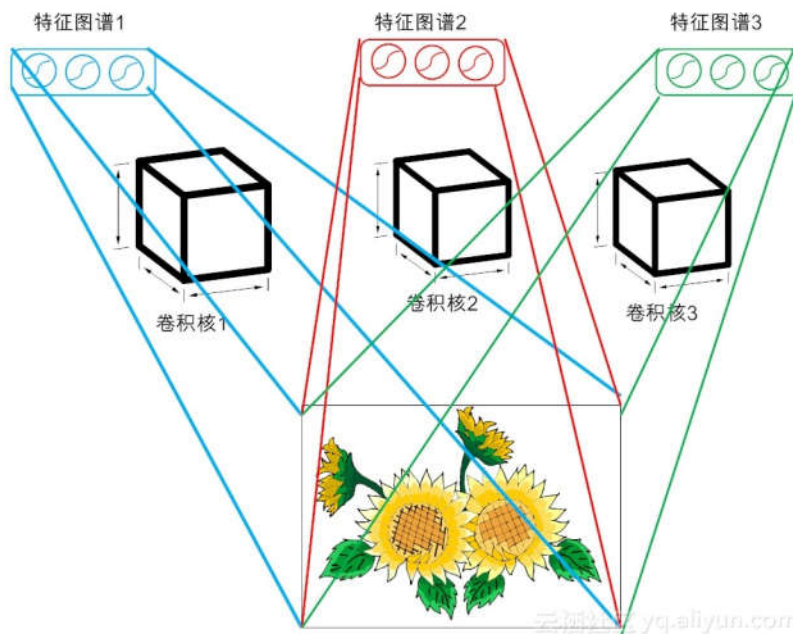


图11-7 权值共享策略

权值共享保证了在卷积时只需要学习一个参数集合即可，而不是对每个位置都再学习一个单独的参数集合。因此参数共享也被称为绑定的权值（tied weights）。

11.4 小节与思考

在本章，我们讨论了卷积神经网络的拓扑结构，并重点讲解了卷积层的设计动机和卷积层的3个核心概念：空间位置排列、局部连接和权值共享。前者确定了神经网络的结构参数，而局部连接和权值共享等策略显著降低了神经元之间的连接数。示意图11-8演示了三种不同的连接类型带来的参

数变化，从图中可以看出，全连接（不包括偏置的权值连接）的参数为15个，局部连接为7个，而权值共享的参数为3个（即红绿蓝线分别共用一个参数）。

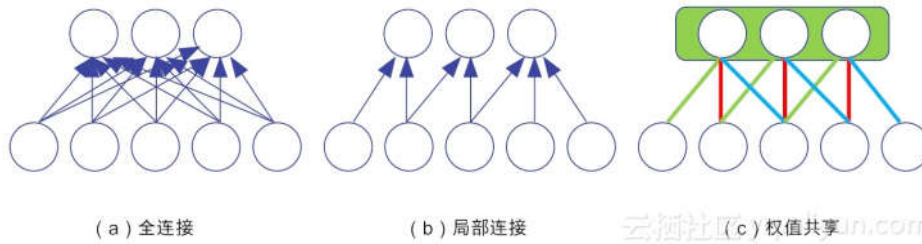


图11-8 神经元连接的三种类型

通过上面的学习，请你思考如下问题：

(1) 虽然局部连接、权值共享等策略大大降低了卷积层与输入层之间的权值调整个数，但并没有提升网络的前向传播速度，你知道为什么吗？

(2) 前文我们提到“肤浅而全面”的全连接，不如“深邃而局部”部分连接。在2016年商汤科技团队在ImageNet图片分类比赛中勇夺冠军，其网络深度已达到1207层。那么，深度学习是不是越深越好？为什么？深度学习就没有未来吗？

在下一讲中，我们将讲解卷积神经网络的剩余部分：池化层、激活层和全连接层。请你关注。

参考文献

- [1] 张玉宏. 云栖社区. 全面连接困何处，卷积网络见解深 (<https://yq.aliyun.com/articles/152935>) (深度学习入门系列之九)
- [2] 张玉宏. 云栖社区. BP算法双向传，链式求导最缠绵 (<https://yq.aliyun.com/articles/110025>) (深度学习入门系列之八)
- [3] 周志华. 机器学习. 清华大学出版社. 2016.1
- [4] The CIFAR-10 dataset. <http://www.cs.toronto.edu/~kriz/cifar.html>
- [5] 黄安埠. 深入浅出深度学习. 中国工信出版社. 2017.6

文章作者：张玉宏，著有《品味大数据 (<http://product.dangdang.com/24048575.html>)》一书。本文节选自《深度学习之美》 (<https://item.jd.com/12382640.html>)一书，电子工业出版社2018年7月出版

(未完待续)

系列文章：

Hello World感知机，懂你我心才安息(深度学习入门系列之五) (<https://yq.aliyun.com/articles/93540>)

损失函数减肥用，神经网络调权重（深度学习入门系列之六） (<https://yq.aliyun.com/articles/96427>)

山重水复疑无路，最快下降问梯度（深度学习入门系列之七） (<https://yq.aliyun.com/articles/105339>)

BP算法双向传，链式求导最缠绵（深度学习入门系列之八） (<https://yq.aliyun.com/articles/110025>)

全面连接困何处，卷积网络见解深（深度学习入门系列之九） (<https://yq.aliyun.com/articles/152935>)

卷地风来忽吹散，积得飘零美如画（深度学习入门系列之十） (<https://yq.aliyun.com/articles/156269>)

局部连接来减参，权值共享肩并肩（深度学习入门系列之十一） (<https://yq.aliyun.com/articles/159710>)

激活引入非线性，池化预防过拟合（深度学习入门系列之十二） (<https://yq.aliyun.com/articles/167391>)