

人工智能学习笔记三——卷积神经网络

卷积神经网络（Convolutional Neural Networks, CNN）是一类包含卷积计算且具有深度结构的前馈神经网络，是深度学习的代表算法之一。卷积神经网络具有表征学习能力，能够按其阶层结构对输入信息进行平移不变分类，因此也被称为“平移不变人工神经网络”。

在正式介绍之前，默认已经了解了神经网络的相关知识。

下面我们演示一下怎么对一个图像做卷积：

首先，我们要搞清楚一张照片是如何输入到神经网络中的。众所周知，计算机适合处理的是矩阵运算，所以必须要把图片转换成矩阵后计算机才能认识。所有的彩色图像都由红绿蓝（RGB）叠加而成，成为图像的三个通道，一张图片在计算机中存储也是通过这三个矩阵完成的。

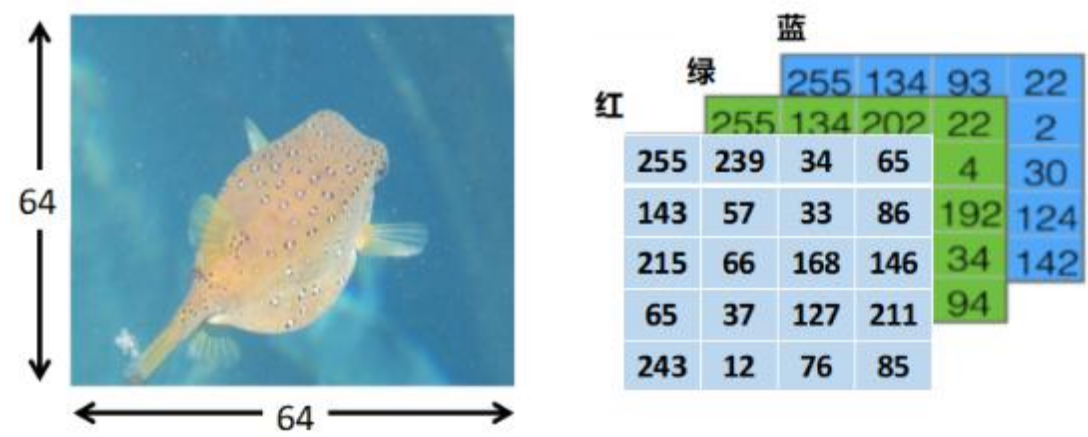


图 1 图像的 RGB 图

如图 1 所示，一张 64×64 个像素大小的图片（例如白色可以表示成 RGB (255, 255, 255)，可以用 3 个 64×64 大小的矩阵来代表这个图。这边为了演示，只画了三个 5×4 的矩阵来代表 64×64 的全尺寸矩阵。RGB 这三个矩阵称为为图像的 3 个通道，也作为神经网络的输入数据。

接着来明确几个基本的概念：卷积核(convolution kernel)、深度(depth)、步长(stride)和补零(zero-padding)。

卷积核(convolution kernel)：卷积核是一个矩阵，可以根据需求设置大小，常见大小为 3×3 ， 5×5 和 7×7 。卷积核的参数一开始是随机赋值，之后通过反向传播进行更新。就相当于标准神经网络模型的权重参数。

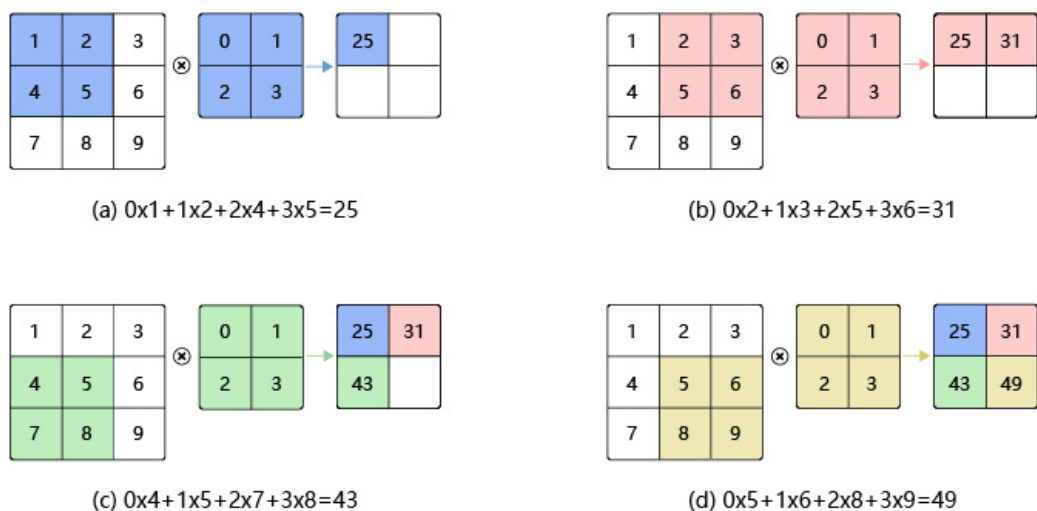


图 2 卷积核运算示意图

图 2 是网上的通用图像卷积解释图了，将两个图像对应的位置的两个数字相乘并累加在一起。然后随着卷积核的滑动，我们就计算出了一个新的图像，这就是图像卷积操作。为了保证模型的深度，通常在卷积操作结束后还会加上偏置项 b ， b 的值一开始是随机的，后续通过反向传播更新。最后将得到的每一个值经过激活函数处理，和神经网络相同。

深度 (depth): 深度指的是图的深度。灰度图只需要一个矩阵就能表示，深度为 1。RGB 图需要三个矩阵表示，深度为 3。在卷积运算中，为了保证网络的学习深度，通常会对同一个矩阵进行多次卷积操作，进行的次数就是图像的深度。

如图 3 所示，图中就将一个矩阵进行了五次卷积操作，得到了五个特征图。通常

用 $F \times F \times D$ 表示卷积核的尺寸， F 表示卷积核大小， D 表示有多少个不同的卷积核，图 3 中 D 的值就为 5。

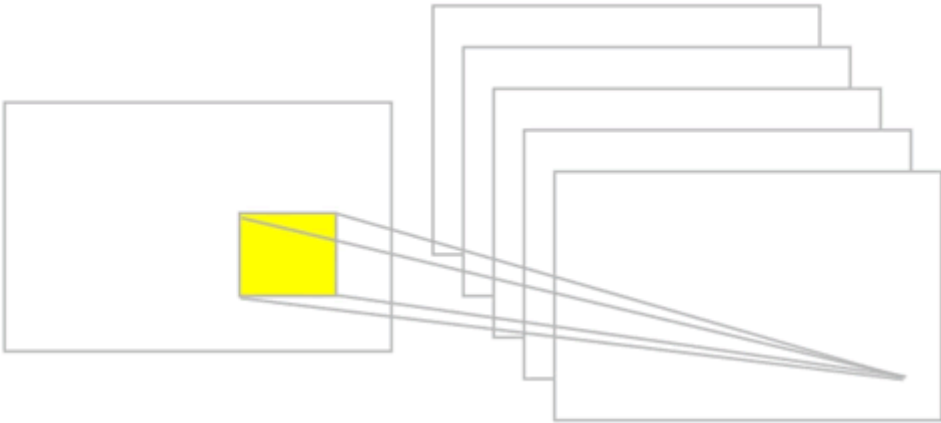


图 3 卷积深度示意图

步长(stride)：用来描述卷积核移动的步长。通常来说，卷积核的步长为 1，也就是卷积核每次都滑动一个像素单位，如图 4 所示。

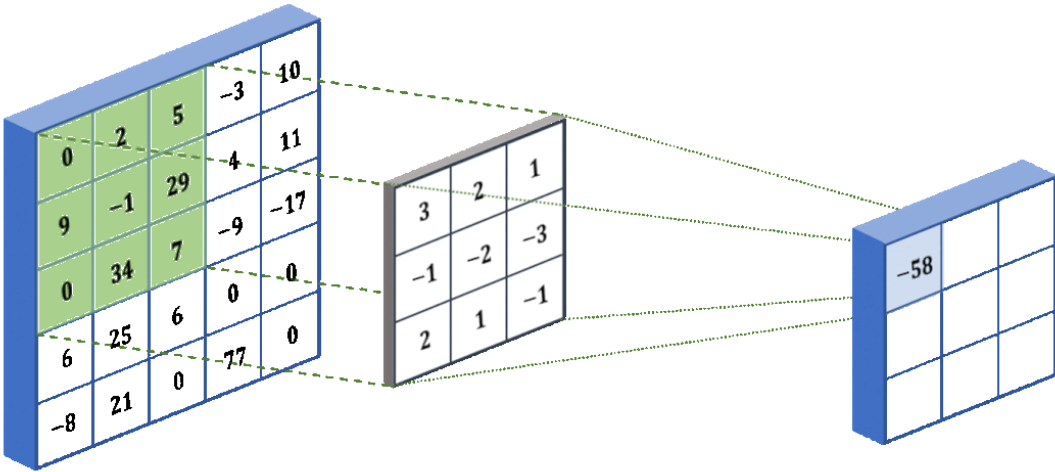


图 4 卷积核滑动示意图

当然，步长也可以设置为其它的数值，如图 5 所示。

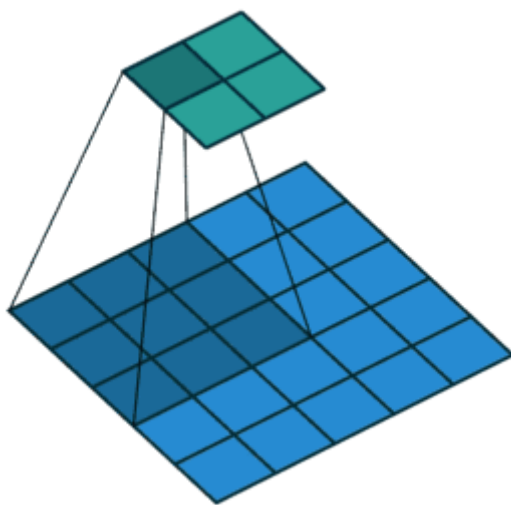


图 5 步长为 2 的滑动示意图

补零 (zero-padding): 通过对图片边缘补零来填充图片边缘，从而控制输出单元的空间大小。之所以在图像边缘补零，是因为图像边缘被卷积的次数小于图像中心的卷积次数，但有时，图像边缘往往也含有重要的特征信息，因而通过补零，让图像边缘的信息能够被学习，通常周围补一圈零，当然也可以根据情况补多圈零。

0	0	0	0	0	0	0	0
0	18	54	51	239	244	188	0
0	55	121	75	78	95	88	0
0	35	24	204	113	109	221	0
0	3	154	104	235	25	130	0
0	15	253	225	159	78	233	0
0	68	85	180	214	245	0	0
0	0	0	0	0	0	0	0

图 6 补零示意图

如果规定输入图片的大小为 $W_1 \times H_1 \times D_1$ ，卷积核的尺寸为 $F \times F \times D$ ，步长为 S ，填充圈数为 P 。

那么经过卷积操作后的图像大小 $W_2 \times H_2 \times D_2$ 满足：

$$W_2 = (W_1 - F + 2P)/S + 1$$

$$H_2 = (H_1 - F + 2P)/S + 1$$

$$D_2 = K$$

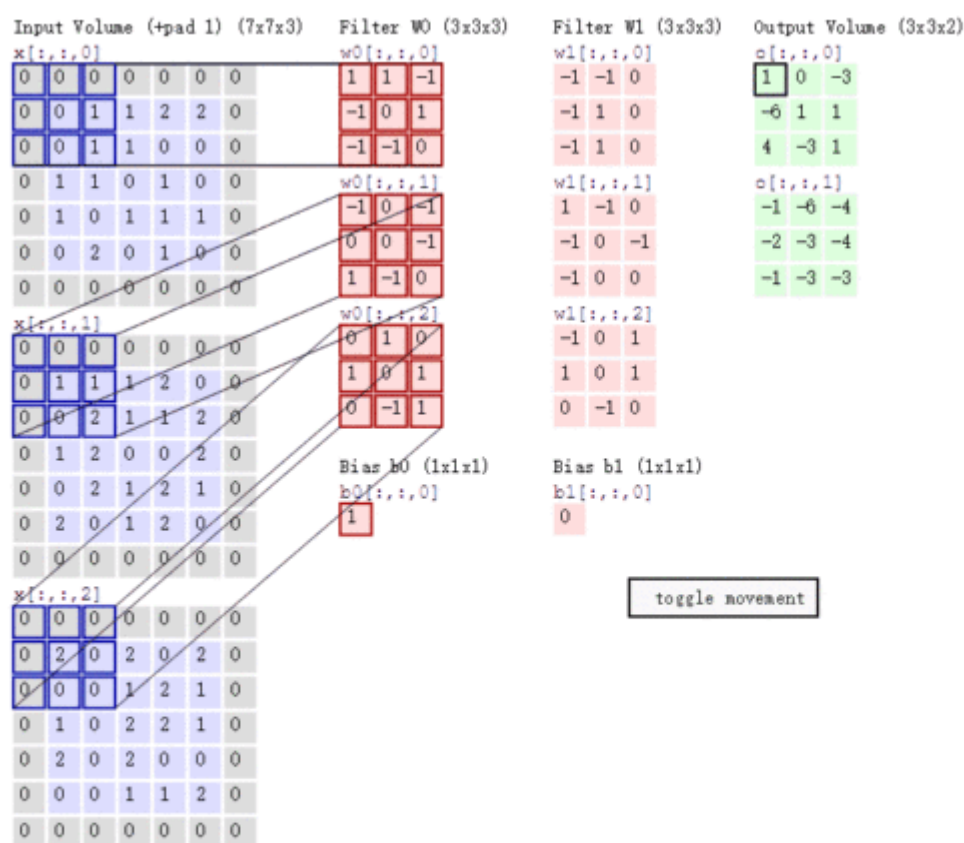


图 7 卷积示意图

举个例子，如图 7 所示，输入图片的大小为 $5 \times 5 \times 3$ ，卷积核的尺寸为 $3 \times 3 \times 2$ ，步长为 2，填充圈数为 1，没有设置激活函数，那么根据公式，输出图片的尺寸为 $3 \times 3 \times 2$ 。

然后来介绍一下池化操作，池化层是当前卷积神经网络中常用组件之一，它最早见于 LeNet 一文，称之为 Subsample。自 AlexNet 之后采用 Pooling 命

名。池化层是模仿人的视觉系统对数据进行降维，用更高层次的特征表示图像。

实施池化的目的：（1）降低信息冗余；（2）提升模型的尺度不变性、旋转不变性；（3）防止过拟合。

池化层的常见操作包含最大值池化，均值池化等。

最大值池化

最大值池化是最常见、也是用的最多的池化操作。它的运算规则是将图像化成一个个的小区域，每一个区域内取最大值，通常区域的划分是 2×2 ，如图8所示。

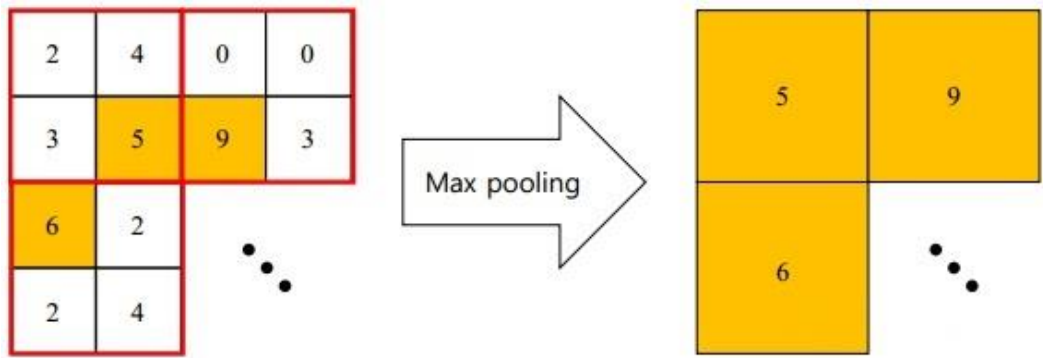


图 8 最大池化示意图

最大值池化的优点在于它能学习到图像的边缘和纹理结构。

均值池化

均值池化的运算规则也是将图像化成一个个的小区域，然后计算图像区域中

的均值作为该区域池化后的值，如图 9 所示。

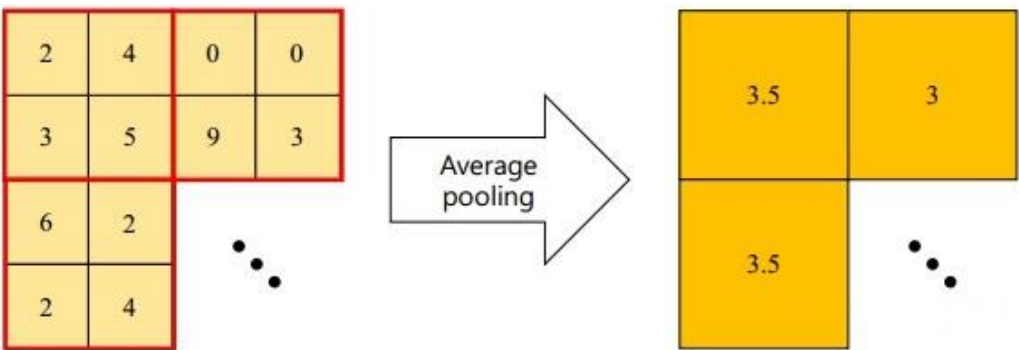


图 9 均值池化示意图

均值池化的优点在于可以减小估计均值的偏移，提升模型的鲁棒性。

最后来介绍一下 flatten 操作。flatten 是用来对数组进行展平操作的，首

先我们假设有一张灰度图片，这个图片只有 3x3 个像素点，分别是 1 到 9，我

们对其进行 flatten 操作。首先它会把每一行进行分开，然后用下一行接在前一

行的后面，形成一个新的数组 1, 2, 3, 4, 5, 6, 7, 8, 9。如图 10 所示：

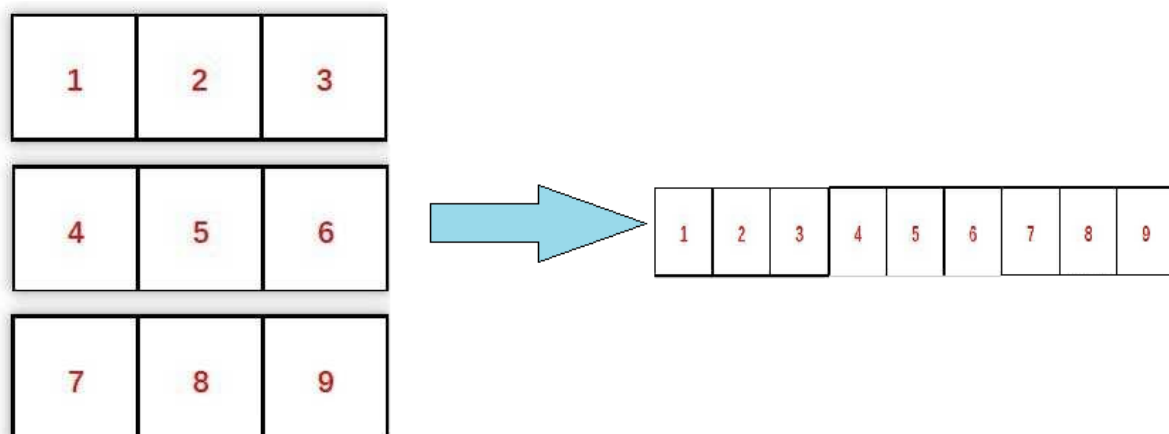


图 10 flatten 示意图

展平的好处是可以将数据变成一维的，从而便于后续进行神经网络的运算。

一个卷积神经网络模型一般都需要经过卷积层，池化层和展平层的运算，如

图 11 所示：

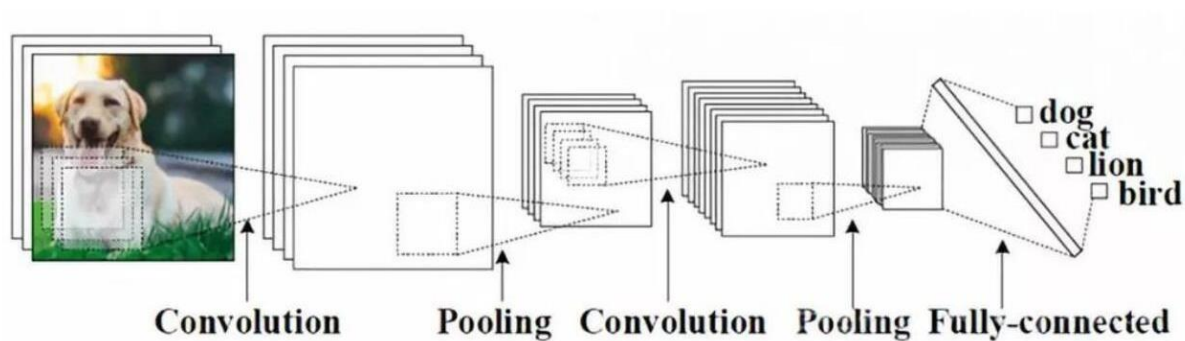


图 11 卷积神经网络模型

参考文章: [图像的处理原理：CNN（卷积神经网络）的实现过程 | 人人都是产品经理](#)

[\(\[woshipm.com\]\(http://woshipm.com\)\)](http://woshipm.com)

[CNN 基础知识——卷积 \(Convolution\)、填充 \(Padding\)、步长\(Stride\)](#)

- [知乎 \(zhihu.com\)](#)

[深度学习入门之池化层 - 知乎 \(zhihu.com\)](#)

本文地址: [TLearning \(caodong0225.github.io\)](https://caodong0225.github.io/TLearning)