

卷积神经网络 (CNN)

学生: E02014154 王浩南

1.1 卷积

在学习卷积神经网络之前,必须先搞懂何为卷积。在不同学科的应用领域,卷积的含义不尽相同,可以说上至天文物理下至工程科学再到生活中无处不在的信号处理和人工智能,卷积在各个领域大放异彩并已经成为强有力的科学工具。

1.1.1 卷积的数学理解

理解卷积在 CNN 领域的应用之前,我认为需要了解其数学定义。在数学中给出的定义为:“在泛函分析中,卷积是透过两个函数 f 和 g 生成第三个函数的一种数学算子,表征函数 f 与经过翻转和平移的 g 的乘积函数所围成的曲边梯形的面积。如果将参加卷积的一个函数看作区间的指示函数,卷积还可以被看作是“移动平均”的推广。” [1]

数学公式分别有离散和连续两种形式,连续形式为:

$$\int_{-\infty}^{\infty} f(\tau)g(x-\tau) d\tau$$

离散形式为:

$$\begin{aligned}(f * g)[n] &\stackrel{\text{def}}{=} \sum_{m=-\infty}^{\infty} f[m]g[n-m] \\ &= \sum_{m=-\infty}^{\infty} f[n-m]g[m].\end{aligned}$$

这里把函数定义域以外的值当成零,所以可以扩展函数到所有整数上,当 $g[n]$ 的支撑集合 (support) 为有限长度 M , 上式会变成有限和:

$$(f * g)[n] = \sum_{m=-M}^M f[n-m]g[m].$$

我的理解是卷积表示了一种特殊的数学运算,它可以用来计算一个输入不稳定但输出是稳定的系统的系统存量。

例如,设 $f(t)$ 为系统在 t_1 时刻的输入, $g(t)$ 为系统在 t_2 时刻剩余量比率,则有图 1.1, 其中每一条连线都代表着一对 $f(x)$ 和 $g(t-x)$ 相乘,所有乘积进行累

加（积分），便是最后系统的存量。

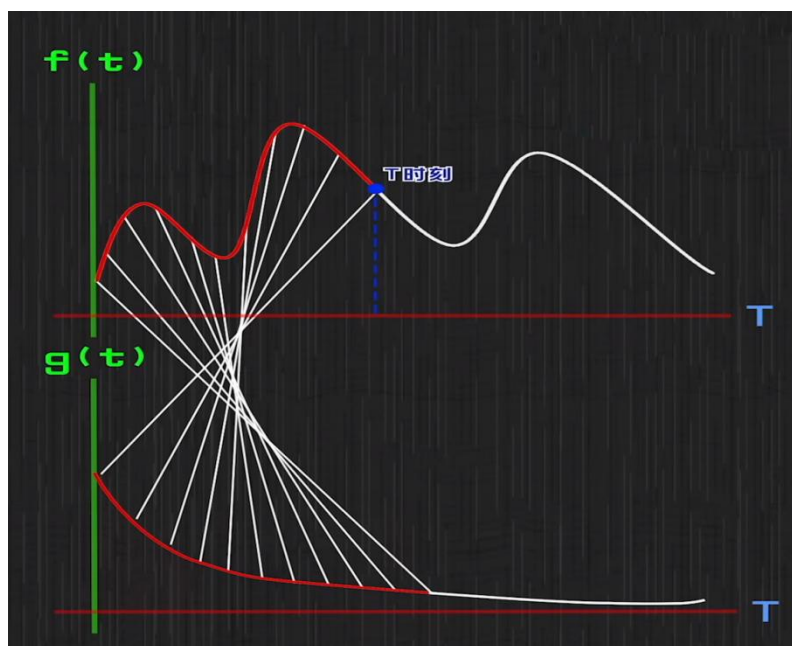


图 1.1

至于为什么叫卷积，我认为把上图翻转则可以解释，图 1.1 中的卷积操作就像把图 1.2 “卷”起来了。

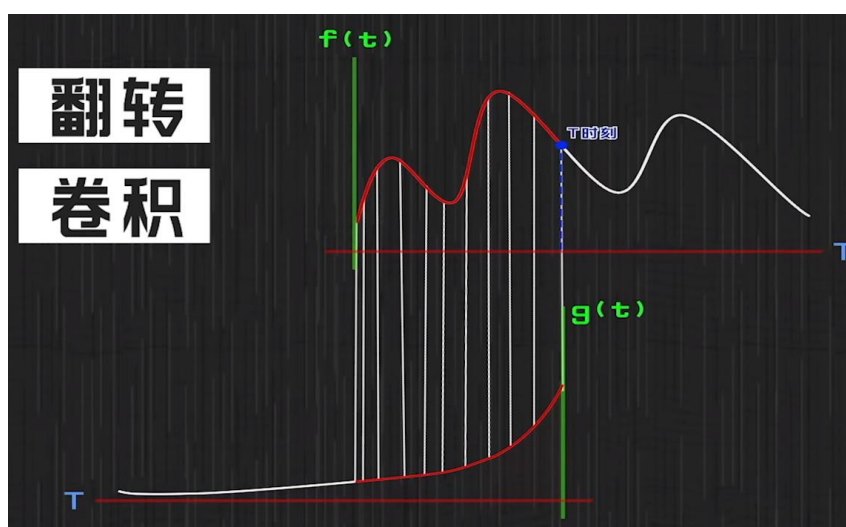


图 1.2

1.1.2 CNN 中的卷积

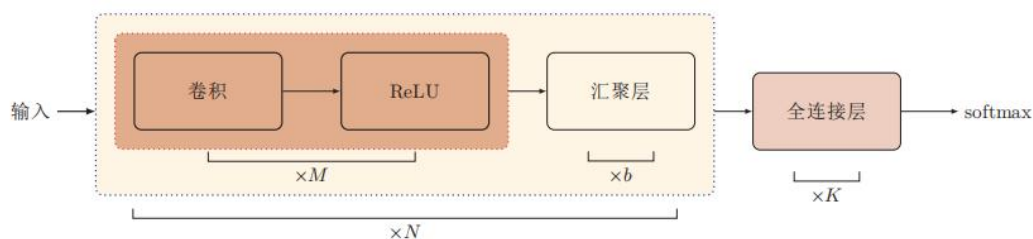
CNN 中的卷积操作并不和上述严格的数学定义完全一致。拿 CNN 在图像处理中的应用来分析，CNN 中的卷积是一种计算周围像素点与当前像素点关系（相互作用）的操作。简单来说，卷积层是用来对输入层进行卷积，提取更高层次的特征。这里卷积的作用和信号处理中的滤波操作类似。而各种各样的卷积核（kernels）就相当于滤波器，通过使用合适的卷积核，我们就可以选择性的过滤掉图像中不

同的信息从而得到不同的特征图。简而言之，我理解的 CNN 中的卷积操作是一种过滤器操作，它可以进行图像的特征提取，而卷积核则规定了如何提取特征，因此使用不同的卷积核就可以提取出不同的特征图。

1.2 卷积神经网络

使用全连接神经网络处理大尺寸图像具有三个明显的缺点：（1）首先将图像展开为向量会丢失空间信息；（2）其次参数过多效率低下，训练困难；（3）同时大量的参数也很快会导致网络过拟合。计算机科学家受动物视觉的皮层结构（动物视觉的神经元在感知外界物体的过程中起作用的只有一部分神经元）启发，将卷积操作和神经网络结合在一起。先使用卷积操作进行局部的、多尺度的特征提取和深度学习，再连接全连接层进行整体的汇总。这种操作基于每个神经元没有必要对全局图像进行感知，只需要对局部进行感知，然后在更高层将局部的信息综合起来就可以得到全局信息的性质。相比于神经元之间的全连接方式，CNN 极大程度上的减少了参数的数量，同时也一定程度上避免了模型的过拟合，很好的解决了全连接神经网络的缺陷。

1.2.1 典型的卷积网络结构



典型的卷积网络结构示意图 [2]

- **卷积层：**卷积层可以产生一组平行的特征图（feature map），它通过在输入图像上滑动不同的卷积核并执行一定的运算而组成。在前馈期间，每个滤波器对输入进行卷积，计算滤波器和输入之间的点积，并产生该滤波器的二维激活图（输入一般是二维向量，但可能有高度（即 RGB））简单来说，卷积层是用来对输入层进行卷积，提取更高层次的特征。类似于人类的视觉感受系统，是 CNN 的“眼睛”。
- **激活函数：**目前一般使用 Relu 函数作为激励函数。Relu 函数全称为线性整流（Rectified Linear Units），数学表达式为 $f(x) = \max(0, x)$ ，它可以增强判定函数和整个神经网络的非线性特性，而本身并不会改变卷积层。其他的一些函数如：双曲正切函数 $f(x) = \tanh(x)$ ，或者 sigmoid 函数 $f(x) = (1 + e^{-x})^{-1}$ ，也可以用于增强网络的非线性特性。

- **领域聚合层/池化层：**在 CNN 中，池化层通常在卷积或者激励函数的后面，池化的方式有两种：最大池化或者平均池化。“最大池化（Max pooling）”是最为常见的。它将输入的图像划分为若干个矩形区域，对每个子区域输出最大值。池化层会不断地减小数据的空间大小，因此参数的数量和计算量也会下降，这在一定程度上控制了过拟合。一般来说，CNN 网络结构中的卷积层之间都会周期性地插入池化层。总结来说，加入池化层可以更好的抓住问题的主要矛盾，降低计算量，减小过拟合。
- **完全连接层：**经过卷积层和池化层降维过的数据，全连接层才能”跑得动”，不然数据量太大，计算成本高，效率低下。和常规的非卷积人工神经网络相同，将学习到的特征表示映射到样本的标记空间，对结果进行识别分类。

1.2.2 CNN 的应用

如今 CNN 应用范围十分广泛，在此仅举出一些例子。

- **图像分类、检索**
图像分类是比较基础的应用，CNN 可以节省大量的人工成本，将图像进行有效的分类。对于一些特定领域的图片，分类的准确率可以达到 95%以上。
- **目标定位检测**
可以在图像中定位目标，并确定目标的位置及大小。可以应用于自动驾驶、安防、医疗等领域。
- **目标分割**
简单理解就是一个像素级的分类。CNN 可以对前景和背景进行像素级的区分、再高级一点还可以识别出目标并且对目标进行分类。应用于美图秀秀、视频后期加工、图像生成等领域。
- **人脸识别**
人脸识别已经走进千家万户，在社会的各个领域发挥着巨大作用，该技术的成熟与广泛使用也要归功于 CNN。

[1] 参考 维基百科: <https://zh.wikipedia.org/wiki/%E5%8D%B7%E7%A7%AF>

[2] 参考 邱锡鹏《神经网络与深度学习》