

COMP90086 Computer Vision

Week 4

CNN

Lecture Notes summarized by Neo

Semester 2 2021

1 Artificial Neural Network

1.1 Definition

人工神经网络¹（英語：Artificial Neural Network, ANN），简称神经网络（Neural Network, NN）或類神經網絡，在机器学习和认知科学领域，是一种模仿生物神经网络（动物的中樞神經系統，特别是大脑）的结构和功能的数学模型或计算模型，用于对函数进行估计或近似。神经网络由大量的人工神经元联结进行计算。大多数情况下人工神经网络能在外界信息的基础上改变内部结构，是一种自适应系统，通俗地讲就是具备学习功能。现代神经网络是一种非线性统计性数据建模工具，神经网络通常是通过一个基于数学统计学类型的学习方法（Learning Method）得以优化，所以也是数学统计学方法的一种实际应用，通过统计学的标准数学方法我们能够得到大量的可以用函数来表达的局部结构空间，另一方面在人工智能学的人工感知领域，我们通过数学统计学的应用可以来做人工感知方面的决定问题，这种方法比起正式的逻辑学推理演算更具有优势。

ANN 是神经网络的一个统称，有不同种类，例如 MLP 和 CNN 等等。

1.2 Type

- Multilayer Perceptron (MLP)
- Convolutional Neural Network (CNN)
- ...

¹https://en.wikipedia.org/wiki/Artificial_neural_network

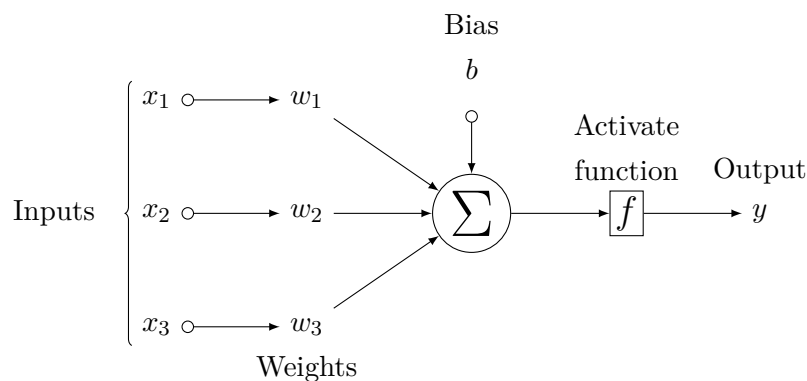
2 Multilayer Perceptron (MLP) 多层感知器

2.1 Perceptron

Perceptron is a **single layer neural network** and a multi-layer perceptron is called Neural Networks.

感知机²是生物神经细胞的简单抽象。神经细胞结构大致可分为：树突、突触、细胞体及轴突。单个神经细胞可被视为一种只有两种状态的机器——激动时为『是』，而未激动时为『否』。神经细胞的状态取决于从其它的神经细胞收到的输入信号量，及突触的强度（抑制或加强）。当信号量总和超过了某个阈值时，细胞体就会激动，产生电脉冲。电脉冲沿着轴突并通过突触传递到其它神经元。为了模拟神经细胞行为，与之对应的感知机基础概念被提出，如权量（突触）、偏置（阈值）及激活函数（细胞体）。

在人工神经网络领域中，感知机也被指为单层的人工神经网络，以区别于较复杂的多层感知机（Multilayer Perceptron）。作为一种线性分类器，（单层）感知机可说是最简单的前向人工神经网络形式。尽管结构简单，感知机能够学习并解决相当复杂的问题。感知机主要的本质缺陷是它不能处理线性不可分问题。



- Weights shows the strength of the particular node.
- The bias value allows you to shift the activation function curve up or down.
- The activation functions are used to map the input between the required values like (0, 1) or (-1, 1).

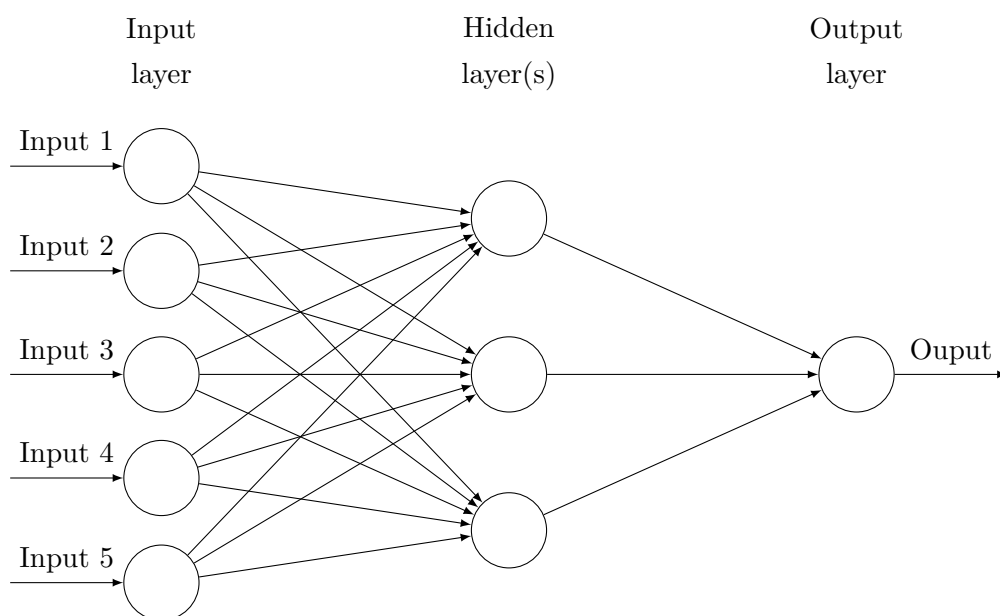
Perceptron is usually used to classify the data into two parts. Therefore, it is also known as a **Linear Binary Classifier**.

²<https://en.wikipedia.org/wiki/Perceptron>

2.2 Multilayer Perceptron

A multilayer perceptron (MLP) is a class of feedforward artificial neural network (ANN). The term MLP is used ambiguously, sometimes loosely to mean any feed-forward ANN, sometimes strictly to refer to networks composed of multiple layers of perceptrons (with threshold activation).

多层感知器³ (Multilayer Perceptron, 缩写 MLP) 是一种前向结构的人工神经网络, 映射一组输入向量到一组输出向量。MLP 可以被看作是一个有向图, 由多个的节点层所组成, 每一层都全连接到下一层。除了输入节点, 每个节点都是一个带有非线性激活函数的神经元 (或称处理单元)。一种被称为反向传播算法的监督学习方法常被用来训练 MLP。MLP 是感知器的推广, 克服了感知器不能对线性不可分数据进行识别的弱点。



MLP 通常被视为“最经典的”人工神经网络。

2.3 Problem

MLP is now deemed insufficient for modern advanced computer vision tasks. Has the characteristic of **fully connected layers**, where each perceptron is connected with every other perceptron. Disadvantage is that the number of total parameters

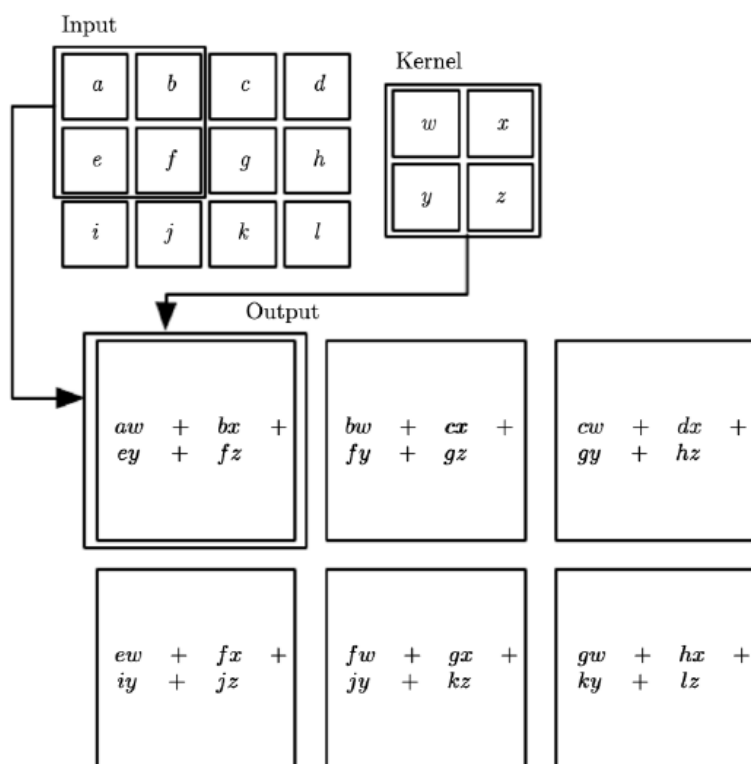
³https://en.wikipedia.org/wiki/Multilayer_perceptron

can grow to very high (number of perceptron in layer 1 multiplied by # of p in layer 2 multiplied by # of p in layer 3...). This is inefficient because there is redundancy in such high dimensions. Another disadvantage is that it disregards spatial information. It takes flattened vectors as inputs. (input 图像会被展平成 1d-vector 导致图像物体位置无法体现。)

3 Convolutional Neural Network (CNN) 卷积神经网络

3.1 Convolution 卷积 recap

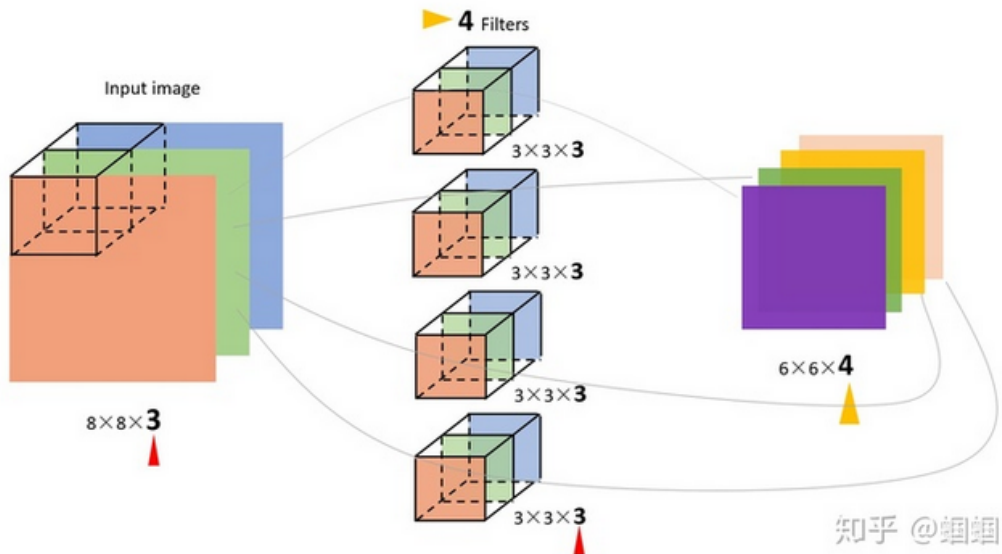
The “convolution” operation in CNN is actually “correlation” operation, i.e., the filter is not flipped before applying to the input. The difference between convolution and correlation (flip the filter or not) does not affect the performance of the algorithm. During training of CNN, the filter will adjust weights that are best suited for the correlation operation, so the flip is omitted.



3.2 Motivation

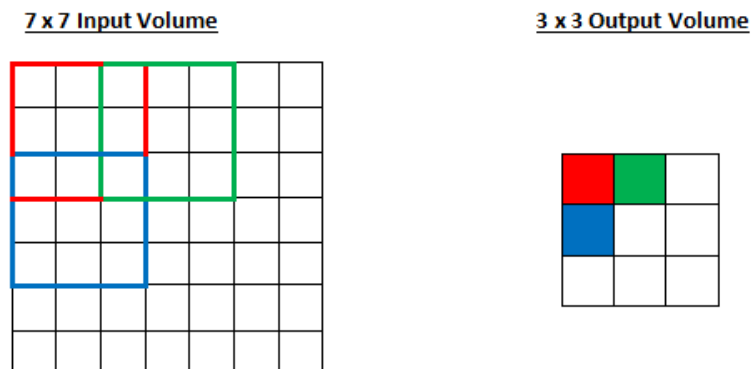
卷积可以用来探测图片与 Kernel 的相似程度，(通过把 Kernel 定义成不同的组合)，所以我们可以使用很多个 Kernel 分别代表不同的“图像特征”与原图像做卷积。

图中的输入图像是 $(8,8,3)$ ，filter 有 4 个，大小均为 $(3,3,3)$ ，得到的输出为 $(6,6,4)$ 。(每一个 Kernel 与原图像三个 RGB 频道做卷积得到其中一层输出。)



3.3 Stride 步长

每两步卷积操作的距离，默认为 1 (Kernel 一格一格地挪)。Stride 为 2 时如下图所示：



- No padding: $\text{output_size} = \text{ceiling}((\text{input_size} - \text{kernel_size} + 1) / \text{stride})$
- With padding: $\text{output_size} = \text{ceiling}(\text{input_size} / \text{stride})$

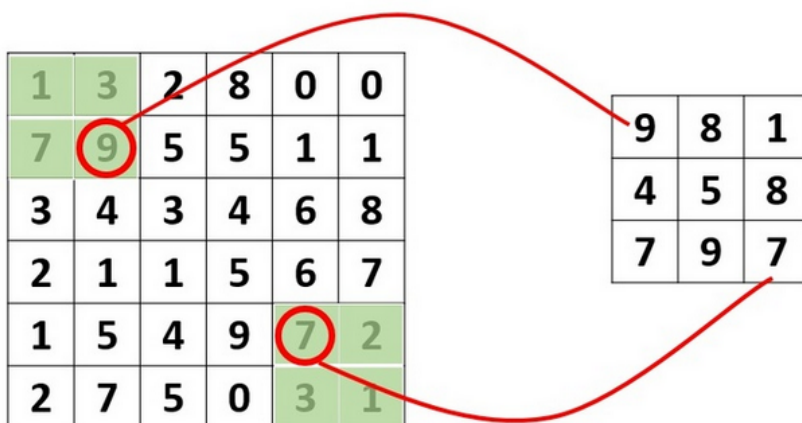
3.4 Padding 填白

由于经过卷积操作图像会变小，需要通过填白先使原图变大再做卷积，具体方式可详见 Week2 笔记。在使用代码框架时我们把“让卷积之后的大小不变”的 padding 方式，称为“Same”方式，把不经过任何填白的，称为“Valid”方式。

3.5 Pooling 池化

为了提取一定区域的主要特征，并减少参数数量，防止模型过拟合。不需要考虑 Kernel。

- Max-Pooling
取 Kernel 大小区域中的最大值，示例见下图。可以最大化提取区域特征。
- Average-Pooling
取 Kernel 大小区域中的平均值
- ...



4 CNN Layers

部分笔记来自知乎 @ 蝈蝈⁴

4.1 Convolutional Layer 卷积层

由 Filters 和 Activate Function 构成。一般要设置的参数包括 filters 的数量、大小、步长，以及 padding。

⁴<https://zhuanlan.zhihu.com/p/42559190>

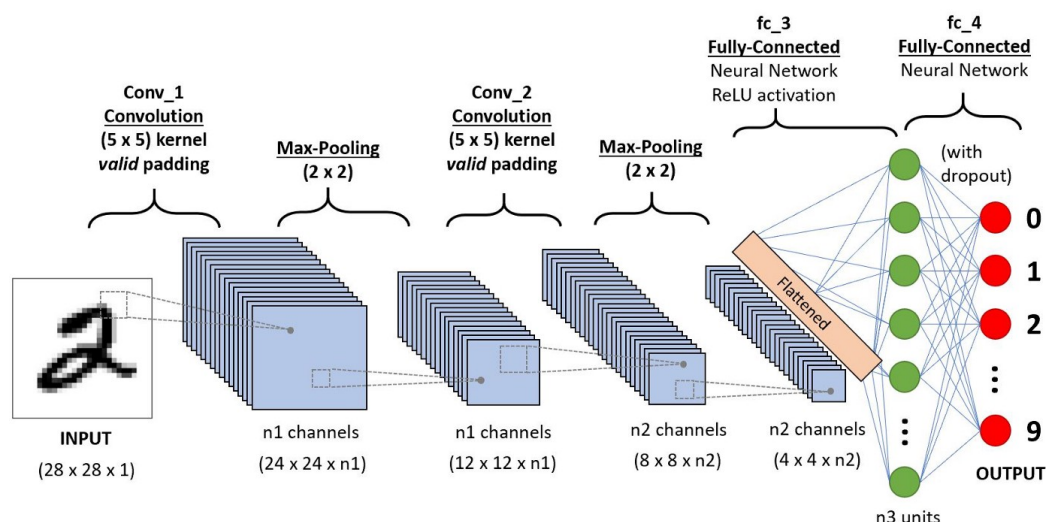
4.2 Pooling Layer 池化层

这里里面没有参数需要我们学习，因为这里里面的参数都是我们设置好了，要么是 Maxpooling，要么是 Averagepooling。需要指定的超参数，包括是 Max 还是 average，窗口大小以及步长。通常，我们使用的比较多的是 Maxpooling，而且一般取大小为 (2,2) 步长为 2 的 filter，这样，经过 pooling 之后，输入的长宽都会减半，channels 不变。

4.3 Fully-Connected Layer 全连接层

发生在一次或数次卷积层和池化层之后：我们最后会先将多维的数据进行“扁平化”，也就是把 (height,width,channel) 的数据压缩成长度为 $\text{height} \times \text{width} \times \text{channel}$ 的一维数组，然后再与 FC 层连接，这之后就跟普通的神经网络无异了。

4.4 Visualization



5 Separable Convolutions

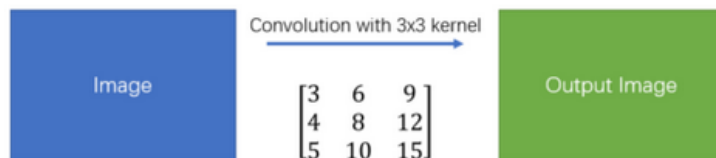
介绍两种如何把一个 kernel 分成两个 kernel 连续卷积的方法，具体的数学推导请详见 Lecture Slides...

5.1 Spatial Separable Convolutions

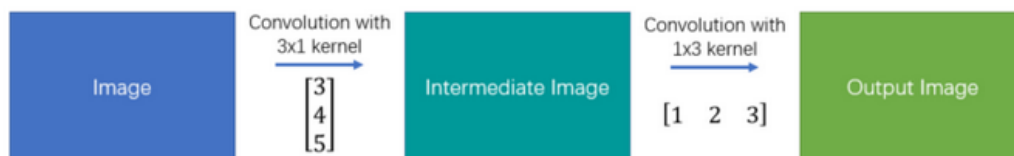
一些 Kernel 可以分成两个 1D 矩阵相乘，如：这样就可以对一个图像进行两次 1D Filter 的卷积，可以有效减少计算复杂度，提高训练速度。

$$\begin{bmatrix} 3 & 6 & 9 \\ 4 & 8 & 12 \\ 5 & 10 & 15 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} \times [1 \ 2 \ 3]$$

Simple Convolution



Spatial Separable Convolution

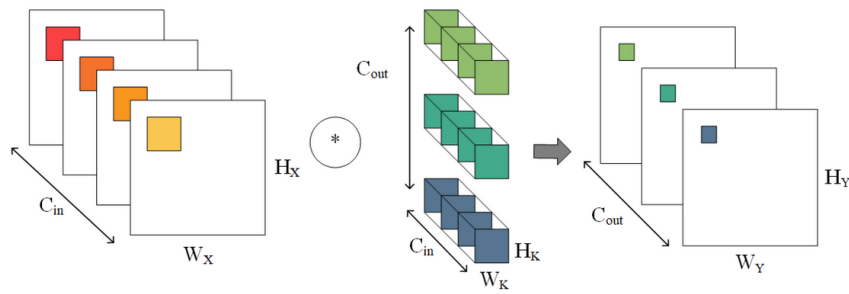


之前提到的 Sobel Filter 用来探测边界就是利用了 Spatial Seperable Convolutions 的思想。

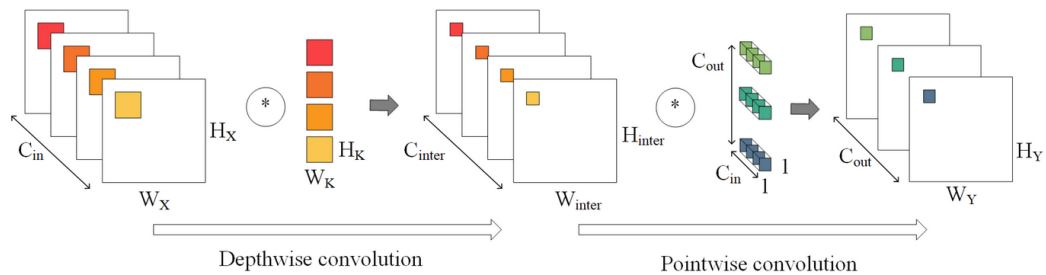
问题: 不是所有的 Kernel Matrix 都可以分成两个 Matrix 相乘的形式。

5.2 Depthwise Seperable Convolutions

讲 Kernel 分为两个部分分别对原图像做 Depth-wise Convolution 和 Point-wise Convolution。要注意不同步骤之间的维度转化, 可见图:



(a) Standard convolution



(b) Depthwise separable convolution

6 Train CNN 训练模型

6.1 Gradient Descent 梯度下降

... 一些数学公式和算法, 详见 Lecture Slides, 一般会使用第三方库直接操作...

7 CNN Visualization (推荐)

<https://poloclub.github.io/cnn-explainer/>