

一、卷积神经网络.

1. 卷积由来.

卷积最初在数学上定义为 $\int_{-\infty}^{+\infty} f(t)g(t-x)dx$. 在深度学习中则有所不同。在受到科学家研究人类视觉的基础之上，人们设计卷积层来模仿人类视觉系统中的感受野。

2. 卷积层.

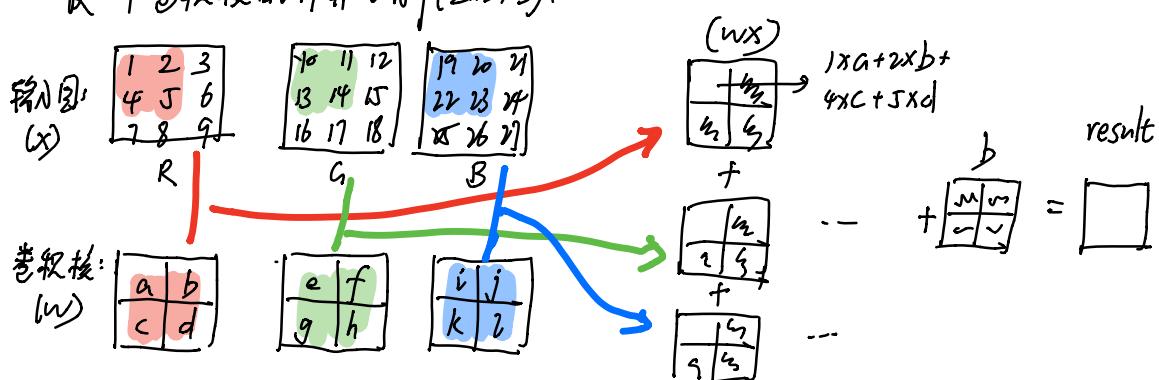
在人类的视觉当中，一个点并不能发挥其作用，而是局部的一块面积起到相应的作用。

计算机中图像以三维张量的形式表示。

卷积层也以类似的形式表示，如 $4 \times 4 \times 3 \times m$

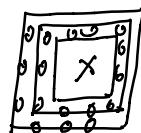
m 为卷积核的数量（类似于隐藏层神经元个数）。

以一个卷积核的例子为例 $(2 \times 2 \times 3)$:



在上面计算，我们并没有引入填充与步幅。

填充：在输入中给矩阵增厚若干圈 0. 如丘陵矩阵为 3×3 , 增加 6 是 5×5

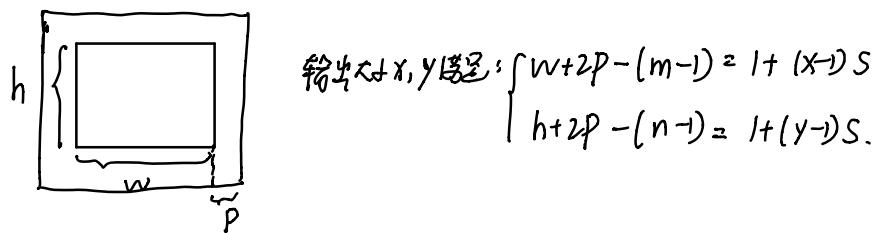


步幅：步幅指卷积核每次向右、向下移动几个单位格。

向右向右
步幅可以调。填充的主要用途在于控制输出的大小以及配合步幅进行卷积。

步幅大的好处在于减少卷积核的参数，但可能会忽略局部特征。

假如我们使用 $h \times w$ 的输入，填充为 P ，步幅为 S ，卷积核大小 $m \times n$.



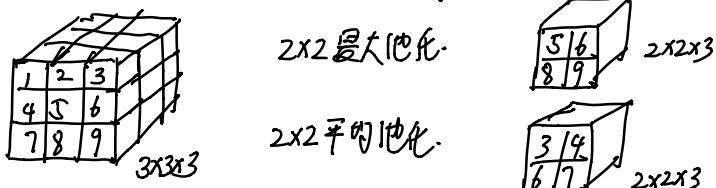
3. 池化层 (pooling)

池化层是为了进一步减少模型的参数而提出的。分为最大池化与平均池化。

池化的操作类似于卷积层，但有两个不同之处。

第一，它没有参数，除了设置步长 s 、步幅 p 。

第二，它单独对每一个输入通道进行。



最大池化一般能进行锐化的作用。

平均池化则是平缓的作用。

池化层也可有 stride, padding 等。

4. 典型的卷积操作。

常见的 CNN 中，一般均为 输入层 \rightarrow 卷积 \rightarrow 卷积 $\rightarrow \dots \rightarrow$ 全连接 $\rightarrow \dots \rightarrow$ 输出。

卷积一般由一个卷积层加上池化层组成。

卷积一般为 $3 \times 3, 5 \times 5, 7 \times 7$ 大小， $stride=2$ 或 3 ， $padding = "same"$ 。池化层一般为 $2 \times 2, 3 \times 3$ 的最大池化网格。

二、典型 CNN

1. 图象增强。

许多图片只对其进行翻转、裁剪、锐化等操作并不会改变其类别，由此可以产生

更多的数据进行学习，但对某些图片进行此操作会改变其类别，如：



2. 表征学习

在对于 CNN 的解释中发现，训练成功的 CNN 中间卷积层的输出为一些目标物体的局部(低级)特征，如一只猫的嘴巴、耳朵等。连接位于全连接层，它们将用于更高阶的特征。

3. dropout

一种抵抗过拟合的解决方案，避免了神经元之间的串通。

4. NIN 层

1x1 卷积层可以捕捉全局维度上的信息。放弃了它相邻的元素关联的模式。
但如将其全局维度作为特征值，高维作为数据，它又相当于全连接层。

AlexNet 与 VGG： $\boxed{\text{卷积层}} \rightarrow \boxed{\text{卷积层}} \rightarrow \boxed{\text{全连接}} \rightarrow \boxed{\text{全连接}}$

NIN : $\boxed{\text{卷积层}} \rightarrow \boxed{1 \times 1 \text{卷积层}} \rightarrow \boxed{\text{卷积层}} \rightarrow \boxed{1 \times 1 \text{卷积层}}$

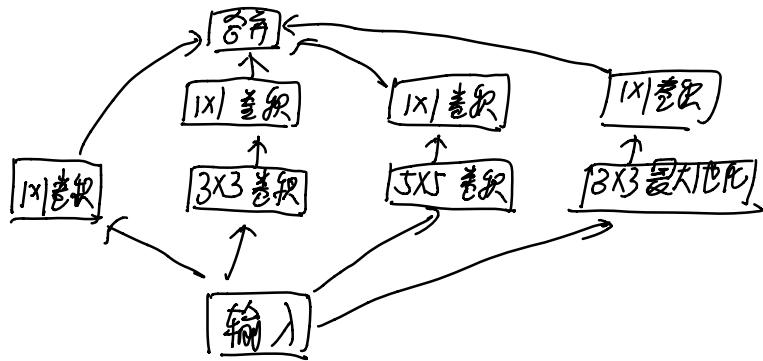
1×1 卷积层充当了全连接层，并且可以捕捉全局上的信息。

5. GooleNet

在 GooleNet 中，使用了一种称为 Inception 块的结构。

在 4 条岔路中，分别捕捉潜在的可能信息。

1×1 卷积主要用于捕捉全局上的信息与减少模型的参数。



三、优化技巧.

1. 批量归一化.

在训练当中，十分容易出现数值的不稳定状态。批量归一化层可以缓解该问题。

(1) 针对全连接层.

归一化操作在仿射变换与激活函数之间，使用 $BN(\cdot)$ 进行归一化操作：

$$\Phi(BN(xw + b))$$

初始化时： $H \leftarrow \frac{m}{m} \sum_{i=1}^m x_i / m$.

$$\sigma^2 \leftarrow \frac{m}{m} \sum_{i=1}^m (x_i - H)^2 / m. \quad (\text{忽略 } m-1)$$

训练时：每一个经过仿射变换的样本 $y_i \leftarrow \frac{y_i - H}{\sqrt{\sigma^2 + \epsilon}}$.

除此之外，批量归一化层引入了两个参数 γ, β .

对上述 y_i 做 y'_i ： $y'_i \leftarrow \gamma \odot y_i + \beta$. \odot 为逐元素乘法。

上面参数引入的目的是，保留了不需要进行归一化的可能性。

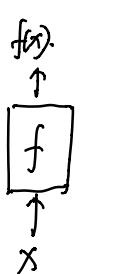
(2) 针对卷积层.

对于卷积层以固定值 γ .

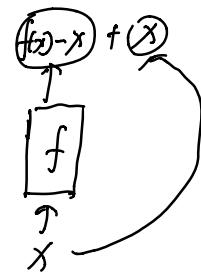
2. 残差神经网络.

理论上，神经网络的层数越多，其效果越好，但实际上并不是这样，由于过深

的神经网络会出现梯度弥散现象.



输出期望的
输出 $f(x)$.



理论可以.
并传给 x 更多的梯度.

3. 稠密连接层.

将 Residual Net 中的 + 变为拼接操作.