# CNN的公式推导

CNN工作的思想可以参照DNN，但也有一些不同：

1）池化层没有激活函数，这个问题倒比较好解决，我们可以令池化层的激活函数为σ(z)=z，即激活后就是自己本身。这样池化层激活函数的导数为1.

2）池化层在前向传播的时候，对输入进行了压缩，那么我们现在需要向前反向推导，这个推导方法和DNN完全不同。

3) 卷积层是通过张量卷积，或者说若干个矩阵卷积求和而得的当前层的输出，这和DNN很不相同，DNN的全连接层是直接进行矩阵乘法得到当前层的输出。这样在卷积层反向传播的时候，上一层的递推计算方法肯定有所不同。

4）对于卷积层，由于使用的运算是卷积，那么从推导出该层的所有卷积核的和的方式也不同。

在研究过程中，需要注意的是，由于卷积层可以有多个卷积核，各个卷积核的处理方法是完全相同且独立的，为了简化算法公式的复杂度，下面提到卷积核都是卷积层中若干卷积核中的一个。

## 符号说明

|  |  |
| --- | --- |
|  | 第层图片的元素个数 |
|  | 神经元的激活函数 |
|  | 第层到第层的某个权重矩阵 |
|  | 为中的元素，表示第层到第层的连接权重 |
|  | 表示第层到第层某个权重矩阵的偏置 |
|  | 等于，表示第层神经元的状态值（输入值） |
|  | 等于，表示第层神经元的激活值（输出值） |
| **\*** | 卷积 |
|  | 哈达玛乘积 |

## 一、已知卷积层的，推导上一隐藏层的

### 1、信息向前传播

即当前层的状态值和激活值可由前一层的状态值和激活值求出：

### 2、误差后向传播

我们已知和两者的关系如下：

其中j表示第层图片像素编号（从左到右，从上到下，依次增大），k表示第层的图片像素编号（从左到右，从上到下，依次增大）。

已知和的关系为：

=

=

因此我们有：

以一个简单的例子推导上面的式子：

假设我们第层的输出为一个3x3的矩阵，第层的卷积核为一个2x2的矩阵，采用1像素的步幅，则输出为一个2x2的矩阵。假设为0，则：

即：

\* =

利用卷积的定义，很容易得出：

假设第层的反向传播误差为

则：

**+ + +**

**=**

**+ + +**

**=**

**+ + +**

**=**

**+ + +**

**=**

**+ + +**

**=**

**+ + +**

**=**

**+ + +**

**=**

**+ + +**

**=**

**+ + +**

**=**

以上的式子可以用一个矩阵卷积的形式来表示：

为了符合梯度计算，我们在误差矩阵的周围填充了一圈0，此时我们将卷积核翻转后和反向传播的梯度误差进行卷积，就得到了前一次的梯度误差。

## 二、已知池化层的，推导上一隐藏层的

在前向传播算法时，池化层一般我们会用MAX或者Average对输入进行池化，池化的区域大小已知。现在我们反过来，要从缩小后的误差，还原前一次较大区域对应的误差。

在反向传播时，我们首先会把的所有子矩阵矩阵大小还原成池化之前的大小。

如果是Max pool，则把的所有子矩阵的各个池化局域的值放在之前做前向传播算法得到最大值的位置。

如果是Average pool，则把的所有子矩阵的各个池化局域的值取平均后放在还原后的子矩阵位置。这个过程一般叫做upsample。

假设池化区域大小为2x2，该层的某个矩阵的为：

由于池化区域为2x2，我们先将还原：

若的池化区域最大值的位置为：

若为MaxPool，则：

若为AveragePool，则：

即：

其中== 1

## 三、已知卷积层的，推导该层的、梯度

由卷积层的前向传播公式：

=

可得：

其中j表示第层图片像素编号（从左到右，从上到下，依次增大），k表示第层的某个卷积核编号（从左到右，从上到下，依次增大）。

假设为4x4的矩阵**，**为3x3的矩阵**，**则输出为2x2的矩阵，那么也是2x2的矩阵。那么根据第一节假设的例子有：

* ·
* ·
* ·

最终可得到9个式子，整理后可用如下矩阵形式表达：

\*

即：

对于，由：

## 四、总结

CNN的前向传播公式可以归纳为：

CNN的后向传播公式可以归纳为：

其中N代表样本数量，j代表该层图片的像素编号（从左到右，从上到下，依次增大）。

对于卷积层的公式有以下一些注意事项：

1、对上式而言，其中卷积的步长和padding为下一个卷积层的步长和padding。

1、对上式而言，其中卷积的步长和padding为前向传播时所用的卷积步长和padding

## 五、编程注意事项

在实际应用中，网络中的各层分布存在以下几种情况，需要分别对原有的公式进行重写。

以下公式中的符号若无再次说明，则其意义为：

操作为：将高维向量以一种方式映射为低维向量。操作为：将低维向量以一种方式映射为高维向量。每层的操作和操作一一对应。

表示层卷积核的个数，代表层卷积核的编号，从“0”开始，其中表示层特征maps的个数，代表层特征maps的编号，从“0”开始。a % b代表a对b取余。下标中的\*代表数字相乘。

### 卷积层

1、前一层为池化层，对应公式重写为：

2、前一层为卷积层，对应公式重写为：

A、一对一连接

B、多对一连接

与前一层为池化层的情况相同

3、前一层为全连接层，对应公式重写为：

4、后一层为池化层，对应公式重写为：

5、后一层为卷积层，对应公式重写为：

A、一对一连接

B、一对多连接

6、后一层为全连接层，对应公式重写为：

以上情况中，偏置的更新都改写为以下公式：

### 池化层

1、前一层为池化层，对应公式重写为：

2、前一层为卷积层，对应公式重写为：

与前一层为池化层的情况相同

3、前一层为全连接层，对应公式重写为：

4、后一层为池化层，对应公式重写为：

5、后一层为卷积层，对应公式重写为：

6、后一层为全连接层

### 全连接层

1、前一层为池化层，对应公式重写为：

2、前一层为卷积层，对应公式重写为：

与前一层为池化层的情况相同

3、前一层为全连接层，对应公式重写为：

4、后一层为池化层，对应公式重写为：

5、后一层为卷积层，对应公式重写为：

6、后一层为全连接层，对应公式重写为：

以上情况中，偏置的更新都表达为：