卷积的反向传播

这篇文章我们讲解一下卷积的反向传播理论，我们会给相应的数学证明。卷积反向传播比普通的全连接层要复杂一些。刚开始接触卷积的对其反向传播会觉得很神秘，很难理解。本篇文章尽量用通俗的语言，并且尽量细致的讲解一下卷积的反向传播计算方法和原理。本文参考了知乎的一篇文章, CNN 卷积层反向传播(https://zhuanlan.zhihu.com/p/40951745)。在这篇文章上做了更为细致的说明，并且给出多通道情况下卷积的反向传播数学证明，而知乎这篇仅仅给出了单通道的反向传播证明。好了，让我们开始吧。

卷积前向传播算法过程：

我们再讲解反向传播理论之前，首先得了解卷积的前向过程是如何计算的。为了便于描述，我们给出一些数学表达符号。用$X_{B\times H\times W\times C_{in}}$表示单通道的输入数据。其形状定义为(批量大小,高度,宽度,通道数)，用(B,H,W,C_{in})来表示。卷积核用$K_{k\times k\times C_{in}\times C_{out}}$表示。该符号表达的意思是卷积核高和宽大小为$k$，通道数和输入$X$一样都是$C_{in}$，输出通道数用$C_{out}$表示，即是卷积核的深度，步长 stride 设置为 1。输出为$Y_{B\times(H-k+1)\times(H-k+1)\times C_{out}}$。我们首先考虑一种最简单的情况，即是$C_{in}$和$C_{out}$都是 1 的时候，然后我们再推广到多个输入和输出通道的情况。这样比较好理解一些。考虑C_{in}和C_{out}都是 1，说明我们处理的是个单通道的输入数据，并且卷积核的深度也是 1。我们将这种情况绘制成矩阵：

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} Conv \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix}$$

我们知道运算过程就是用卷积核根据步长 stride 在输入数据X上先左右滑动再上下滑动，用卷积核覆盖住的X上的元素和卷积核的元素按位相乘，即是以下表达式：

$$y_{11} = x_{11}k_{11} + x_{12}k_{12} + x_{21}k_{21} + x_{22}k_{22}$$

$$y_{12} = x_{12}k_{11} + x_{13}k_{12} + x_{22}k_{21} + x_{23}k_{22}$$

$$y_{21} = x_{21}k_{11} + x_{22}k_{12} + x_{31}k_{21} + x_{32}k_{22}$$

$$y_{22} = x_{22}k_{11} + x_{23}k_{12} + x_{32}k_{21} + x_{33}k_{22}$$

可以写成矩阵相乘的形式：

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & x_{21} & x_{22} \\ x_{12} & x_{13} & x_{22} & x_{23} \\ x_{21} & x_{22} & x_{31} & x_{32} \\ x_{22} & x_{23} & x_{32} & x_{33} \end{pmatrix} \cdot \begin{pmatrix} k_{11} \\ k_{12} \\ k_{21} \\ k_{22} \end{pmatrix}$$

由上面给出的矩阵相乘，我们得到启发，其实卷积的前向运算本质上可以转换为矩阵乘

法运算的。如果是矩阵乘法的运算， 那么和全连接层的计算方式就一致了。我们完全熟悉矩阵乘法形式的前向传播，是如何进行反向传播计算的。之前写的深度学习的反向传播的数学推导已经详细说明了。$\begin{pmatrix} k_{11} \\ k_{12} \\ k_{21} \\ k_{22} \end{pmatrix}$和$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{pmatrix}$就是单通道卷积核$K$做一个 reshape(4,1)，输出结果$Y$做个 reshpe(4,1)。就这个$\begin{pmatrix} x_{11} & x_{12} & x_{21} & x_{22} \\ x_{12} & x_{13} & x_{22} & x_{23} \\ x_{21} & x_{22} & x_{31} & x_{32} \\ x_{22} & x_{23} & x_{32} & x_{33} \end{pmatrix}$麻烦一些。如果读者看过一些框架的源码，或者一些书籍介绍卷积的内容，会把将输入X转换为$\begin{pmatrix} x_{11} & x_{12} & x_{21} & x_{22} \\ x_{12} & x_{13} & x_{22} & x_{23} \\ x_{21} & x_{22} & x_{31} & x_{32} \\ x_{22} & x_{23} & x_{32} & x_{33} \end{pmatrix}$的过程称为 im2col（image to column）。其实就是把每次做卷积运算时，卷积窗口对应的X的元素给排成一行，比方说第一个值y_{11}，它是卷积核$\begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix}$和$\begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{13} \end{pmatrix}$按位乘得到的。我们把$\begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{13} \end{pmatrix}$转换为 1 行，即$(x_{11} \quad x_{12} \quad x_{21} \quad x_{22})$，每行的大小是$k \times k$。这样的一行和卷积核转换成的列，进行点乘就是输入值了。依次类推 ，每个输出都是将卷积窗口中对应X的元素排成一行，和卷积核的列相乘。那么总共要计算的次数就是输出值的高×宽，$(H - k + 1) \times (W - k + 1)$。

这是单通道的情况，下面我们考虑多个通道。其实道理也是一样的。我们将输入数据的形状设置为$X_{3\times3\times3}$，高和宽以及通道数都设置为 3，卷积核设置为$K_{2\times2\times3\times2}$，其深度设置为 2.其余保持不变。为了方便描述，我们给出具体的矩阵形式，请看如下表达式：

$$X_{3\times3}^{[1]} = \begin{pmatrix} x_{11}^{[1]} & x_{12}^{[1]} & x_{13}^{[1]} \\ x_{21}^{[1]} & x_{22}^{[1]} & x_{23}^{[1]} \\ x_{31}^{[1]} & x_{32}^{[1]} & x_{33}^{[1]} \end{pmatrix}$$

$$X_{3\times3}^{[2]} = \begin{pmatrix} x_{11}^{[2]} & x_{12}^{[2]} & x_{13}^{[2]} \\ x_{21}^{[2]} & x_{22}^{[2]} & x_{23}^{[2]} \\ x_{31}^{[2]} & x_{32}^{[2]} & x_{33}^{[2]} \end{pmatrix}$$

$$X_{3\times3}^{[3]} = \begin{pmatrix} x_{11}^{[3]} & x_{12}^{[3]} & x_{13}^{[3]} \\ x_{21}^{[3]} & x_{22}^{[3]} & x_{23}^{[3]} \\ x_{31}^{[3]} & x_{32}^{[3]} & x_{33}^{[3]} \end{pmatrix}$$

这里$X_{3\times3}^{[1]}$表示输入数据中第一个通道代表的 9 个元素。以此类推，$X_{3\times3}^{[2]}$和$X_{3\times3}^{[3]}$是第二个

和第三个通道对应的 9 个元素。我们再按照这个方式，对卷积核$K_{2\times2\times3\times2}$做如下分解：

$$K_{2\times2}^{\{1\}[1]}=\begin{pmatrix}k_{11}^{\{1\}[1]} & k_{12}^{\{1\}[1]}\\ k_{21}^{\{1\}[1]} & k_{22}^{\{1\}[1]}\end{pmatrix}\ K_{2\times2}^{\{1\}[2]}=\begin{pmatrix}k_{11}^{\{1\}[2]} & k_{12}^{\{1\}[2]}\\ k_{21}^{\{1\}[2]} & k_{22}^{\{1\}[2]}\end{pmatrix}\ K_{2\times2}^{\{1\}[3]}=\begin{pmatrix}k_{11}^{\{1\}[3]} & k_{12}^{\{1\}[3]}\\ k_{21}^{\{1\}[3]} & k_{22}^{\{1\}[4]}\end{pmatrix}$$

$$K_{2\times2}^{\{2\}[1]}=\begin{pmatrix}k_{11}^{\{2\}[1]} & k_{12}^{\{2\}[1]}\\ k_{21}^{\{2\}[1]} & k_{22}^{\{2\}[1]}\end{pmatrix}\ K_{2\times2}^{\{1\}[2]}=\begin{pmatrix}k_{11}^{\{2\}[2]} & k_{12}^{\{2\}[2]}\\ k_{21}^{\{2\}[2]} & k_{22}^{\{2\}[2]}\end{pmatrix}\ K_{2\times2}^{\{1\}[3]}=\begin{pmatrix}k_{11}^{\{2\}[3]} & k_{12}^{\{2\}[3]}\\ k_{21}^{\{2\}[3]} & k_{22}^{\{2\}[4]}\end{pmatrix}$$

这里大括号里面的数字表示的是卷积核的深度，即C_{out}，中括号里的数字表示的是通道数，即C_{in}。这样$K_{2\times2}^{\{1\}[1]}$就表示第一个深度为 1，第一通道的四个元素，然后以此类推。那么这种多通道的卷积计算是如何转换为矩阵相乘的形式呢。首先输出数据的形状是$Y_{2\times2\times2}$，即高和宽还有通道数都是 2。我们依然是按照上面的方式把$Y_{2\times2\times2}$按照通道分解一下：

$$Y_{2\times2}^{[1]}=\begin{pmatrix}y_{11}^{[1]} & y_{12}^{[1]}\\ y_{21}^{[1]} & y_{22}^{[1]}\end{pmatrix}\ \ Y_{2\times2}^{[2]}=\begin{pmatrix}y_{11}^{[2]} & y_{12}^{[2]}\\ y_{21}^{[2]} & y_{22}^{[2]}\end{pmatrix}$$

我们先来看下卷积的前向传播计算过程。我们把输出Y的每个元素计算表达式写出来:

$$Y_{2\times2}^{[1]}=X_{3\times3}\ Conv\ K_{2\times2}^{\{1\}}=X_{3\times3}^{[1]}ConvK_{2\times2}^{\{1\}[1]}+X_{3\times3}^{[1]}ConvK_{2\times2}^{\{1\}[1]}+X_{3\times3}^{[1]}ConvK_{2\times2}^{\{1\}[1]}$$

$$\begin{pmatrix}y_{11}^{[1]} & y_{12}^{[1]}\\ y_{21}^{[1]} & y_{22}^{[1]}\end{pmatrix}=\begin{pmatrix}x_{11}^{[1]} & x_{12}^{[1]} & x_{13}^{[1]}\\ x_{21}^{[1]} & x_{22}^{[1]} & x_{23}^{[1]}\\ x_{31}^{[1]} & x_{32}^{[1]} & x_{33}^{[1]}\end{pmatrix}Conv\begin{pmatrix}k_{11}^{\{1\}[1]} & k_{12}^{\{1\}[1]}\\ k_{21}^{\{1\}[1]} & k_{22}^{\{1\}[1]}\end{pmatrix}+$$

$$\begin{pmatrix}x_{11}^{[2]} & x_{12}^{[2]} & x_{13}^{[2]}\\ x_{21}^{[2]} & x_{22}^{[2]} & x_{23}^{[2]}\\ x_{31}^{[2]} & x_{32}^{[2]} & x_{33}^{[2]}\end{pmatrix}Conv\begin{pmatrix}k_{11}^{\{1\}[2]} & k_{12}^{\{1\}[2]}\\ k_{21}^{\{1\}[2]} & k_{22}^{\{1\}[2]}\end{pmatrix}+$$

$$\begin{pmatrix}x_{11}^{[3]} & x_{12}^{[3]} & x_{13}^{[3]}\\ x_{21}^{[3]} & x_{22}^{[3]} & x_{23}^{[3]}\\ x_{31}^{[3]} & x_{32}^{[3]} & x_{33}^{[3]}\end{pmatrix}Conv\begin{pmatrix}k_{11}^{\{1\}[3]} & k_{12}^{\{1\}[3]}\\ k_{21}^{\{1\}[3]} & k_{22}^{\{1\}[4]}\end{pmatrix}$$

$$y_{11}^{[1]} = \begin{pmatrix} x_{11}^{[1]} & x_{12}^{[1]} \\ x_{21}^{[1]} & x_{22}^{[1]} \end{pmatrix} * \begin{pmatrix} k_{11}^{\{1\}[1]} & k_{12}^{\{1\}[1]} \\ k_{21}^{\{1\}[1]} & k_{22}^{\{1\}[1]} \end{pmatrix} + \begin{pmatrix} x_{11}^{[2]} & x_{12}^{[2]} \\ x_{21}^{[2]} & x_{22}^{[2]} \end{pmatrix} * \begin{pmatrix} k_{11}^{\{1\}[2]} & k_{12}^{\{1\}[2]} \\ k_{21}^{\{1\}[2]} & k_{22}^{\{1\}[2]} \end{pmatrix} + \begin{pmatrix} x_{11}^{[3]} & x_{12}^{[3]} \\ x_{21}^{[3]} & x_{22}^{[3]} \end{pmatrix} * \begin{pmatrix} k_{11}^{\{1\}[3]} & k_{12}^{\{1\}[3]} \\ k_{21}^{\{1\}[3]} & k_{22}^{\{1\}[4]} \end{pmatrix}$$

从$y_{11}^{[1]}$的计算表达算式，结合之前说的单通道卷积如何转换成矩阵相乘的形式。我们可以得到多通道转换矩阵相乘的方法。取卷积核$K_{k\times k\times C_{in}}$这一部分，在我们的例子里即是$[K_{2\times2}^{\{1\}[1]}, K_{2\times2}^{\{1\}[2]}, K_{2\times2}^{\{1\}[3]}]$这些元素，将他们排成一列：

$$\begin{pmatrix} k_{11}^{\{1\}[1]} & k_{12}^{\{1\}[1]} & k_{21}^{\{1\}[1]} & k_{22}^{\{1\}[1]} & k_{11}^{\{1\}[2]} & k_{12}^{\{1\}[2]} & k_{21}^{\{1\}[2]} & k_{22}^{\{1\}[2]} & k_{11}^{\{1\}[3]} & k_{12}^{\{1\}[3]} & k_{21}^{\{1\}[3]} & k_{22}^{\{1\}[3]} \end{pmatrix}^T$$

然后取$[K_{2\times2}^{\{1\}[1]}, K_{2\times2}^{\{1\}[2]}, K_{2\times2}^{\{1\}[3]}]$这部分对应的输入数据$X$，即$\begin{pmatrix} x_{11}^{[1]} & x_{12}^{[1]} \\ x_{21}^{[1]} & x_{22}^{[1]} \end{pmatrix}$，$\begin{pmatrix} x_{11}^{[2]} & x_{12}^{[2]} \\ x_{21}^{[2]} & x_{22}^{[2]} \end{pmatrix}$，$\begin{pmatrix} x_{11}^{[3]} & x_{12}^{[3]} \\ x_{21}^{[3]} & x_{22}^{[3]} \end{pmatrix}$，将他们排成一行。即：

$$\begin{pmatrix} x_{11}^{[1]} & x_{12}^{[1]} & x_{21}^{[1]} & x_{22}^{[1]} & x_{11}^{[2]} & x_{12}^{[2]} & x_{21}^{[2]} & x_{22}^{[2]} & x_{11}^{[3]} & x_{12}^{[3]} & x_{21}^{[3]} & x_{22}^{[3]} \end{pmatrix}$$

用这一行和前面卷积组成的列按位乘，就得到输出元素了。每个输出值都是通过这种方式计算的出来的，就不一个一个举例了。所以我们总结到规律就是将卷积核做一次 reshape，从$K_{k\times k\times C_{in}\times C_{out}}$变为$K_{[k\times k\times C_{in}, C_{out}]}$。即变为一个二维矩阵。行数为$k\times k\times C_{in}$，列数为$C_{out}$。输入数据$X$，通过 im2col 这个过程，将对应的和卷积核元素按位乘的数据排成行。整个X转换完后的形状是$[B, (H-k+1)\times(W-k+1), k\times k\times C_{in}]$。这里$B$是批量数大小。然后将这两个矩阵做点乘：

$$Y_{[B,(H-k+1)\times(W-k+1),C_{out}]} = \mathrm{im2col(X)}_{[B,(H-k+1)\times(W-k+1),k\times k\times C_{in}]} \cdot K_{[k\times k\times C_{in},C_{out}]}$$

这样多通道的卷积前向传播的矩阵相乘形式就做出来了。最后我们给出im2col的实现过程，这里用 python 代码描述。

```
def im2col(X, ksize, stride):
    # X shape:[B, H, W, C_in]
    X_col = []

    #总共有(H − k + 1) × (W − k + 1)这个多次
    for h in range(0, X.shape[1] - ksize + 1, stride):
        for w in range(0, X.shape[2] - ksize + 1, stride):
            col = X[:, h:h + ksize, w:w + ksize, :].reshape([-1])
            X_col.append(col)
    #最终得到的 X_col 的 shape:[ (H − k + 1) × (W − k + 1), B × k × k × C_in].
    #这个 shape 不是我们需要的，所以要进行调整
```

```
X_col = np.array(X_col)

#调整为[B,(H-k+1)×(W-k+1),k×k×C_in],
#交换轴向是要考虑数据在内存的存储顺序
X_col = X_col.reshape(X_col,shape[0],X.shape[0],-1).swapaxes(0,1)
return X_col
```

卷积反向传播：

我们开始通过方向传播来计算梯度。我们用 L 来表示损失函数。我们需要求解的是$\frac{\partial L}{\partial X}$，$\frac{\partial L}{\partial K}$以及$\frac{\partial L}{\partial b}$。前面我们得到了将卷积前向传播转为矩阵相乘的计算公式。根据这个公式我们可以很容易求得$\frac{\partial L}{\partial K}$以及$\frac{\partial L}{\partial b}$。我们先将卷积的矩阵相乘关系式列出:

$$Y_{[B,(H-k+1)\times(W-k+1),C_{out}]} = \mathrm{im2col(X)}_{[B,(H-k+1)\times(W-k+1),k\times k\times C_{in}]} \cdot K_{[k\times k\times C_{in},C_{out}]}$$

我们设置当前层传递过来的误差为$\frac{\partial L}{\partial Y}$。为了描述方便，我设置$(B,H,W,K,C_{in},C_{out})$的值为$(1,3,3,2,3,2)$。Stride 步长设置为 1。我们用将所有相关参数用矩阵和向量来表示。

$$\frac{\partial L}{\partial Y_{2\times 2\times 2}} = \begin{pmatrix} \frac{\partial L}{\partial Y_{2*2}^{\{1\}}} & \frac{\partial L}{\partial Y_{2*2}^{\{2\}}} \end{pmatrix}$$

$$\frac{\partial L}{\partial Y_{2*2}^{\{1\}}} = \begin{pmatrix} \frac{\partial L}{\partial y_{11}^{\{1\}}} & \frac{\partial L}{\partial y_{12}^{\{1\}}} \\ \frac{\partial L}{\partial y_{21}^{\{1\}}} & \frac{\partial L}{\partial y_{22}^{\{1\}}} \end{pmatrix}$$

$$\frac{\partial L}{\partial Y_{2*2}^{\{2\}}} = \begin{pmatrix} \frac{\partial L}{\partial y_{11}^{\{2\}}} & \frac{\partial L}{\partial y_{12}^{\{2\}}} \\ \frac{\partial L}{\partial y_{21}^{\{2\}}} & \frac{\partial L}{\partial y_{22}^{\{2\}}} \end{pmatrix}$$

$$K_{2\times 2\times 3\times 2} = \begin{pmatrix} K_{2\times 2\times 3}^{\{1\}} & K_{2\times 2\times 3}^{\{2\}} \end{pmatrix}$$

$$K_{2\times 2\times 3}^{\{1\}} = \begin{pmatrix} K_{2\times 2}^{\{1\}[1]} & K_{2\times 2}^{\{1\}[2]} & K_{2\times 2}^{\{1\}[3]} \end{pmatrix}$$

$$K_{2\times2\times3}^{\{2\}} = \begin{pmatrix} K_{2\times2}^{\{2\}[1]} & K_{2\times2}^{\{2\}[2]} & K_{2\times2}^{\{2\}[3]} \end{pmatrix}$$

$$K_{2\times2}^{\{1\}[1]} = \begin{pmatrix} k_{11}^{\{1\}[1]} & k_{12}^{\{1\}[1]} \\ k_{21}^{\{1\}[1]} & k_{22}^{\{1\}[1]} \end{pmatrix} K_{2\times2}^{\{1\}[2]} = \begin{pmatrix} k_{11}^{\{1\}[2]} & k_{12}^{\{1\}[2]} \\ k_{21}^{\{1\}[2]} & k_{22}^{\{1\}[2]} \end{pmatrix} K_{2\times2}^{\{1\}[3]} = \begin{pmatrix} k_{11}^{\{1\}[3]} & k_{12}^{\{1\}[3]} \\ k_{21}^{\{1\}[3]} & k_{22}^{\{1\}[4]} \end{pmatrix}$$

$$K_{2\times2}^{\{2\}[1]} = \begin{pmatrix} k_{11}^{\{2\}[1]} & k_{12}^{\{2\}[1]} \\ k_{21}^{\{2\}[1]} & k_{22}^{\{2\}[1]} \end{pmatrix} K_{2\times2}^{\{1\}[2]} = \begin{pmatrix} k_{11}^{\{2\}[2]} & k_{12}^{\{2\}[2]} \\ k_{21}^{\{2\}[2]} & k_{22}^{\{2\}[2]} \end{pmatrix} K_{2\times2}^{\{1\}[3]} = \begin{pmatrix} k_{11}^{\{2\}[3]} & k_{12}^{\{2\}[3]} \\ k_{21}^{\{2\}[3]} & k_{22}^{\{2\}[4]} \end{pmatrix}$$

$$X_{3\times3\times3} = \begin{pmatrix} X_{3\times3}^{[1]} & X_{3\times3}^{[2]} & X_{3\times3}^{[3]} \end{pmatrix}$$

$$X_{3\times3}^{[1]} = \begin{pmatrix} x_{11}^{[1]} & x_{12}^{[1]} & x_{13}^{[1]} \\ x_{21}^{[1]} & x_{22}^{[1]} & x_{23}^{[1]} \\ x_{31}^{[1]} & x_{32}^{[1]} & x_{33}^{[1]} \end{pmatrix}$$

$$X_{3\times3}^{[2]} = \begin{pmatrix} x_{11}^{[2]} & x_{12}^{[2]} & x_{13}^{[2]} \\ x_{21}^{[2]} & x_{22}^{[2]} & x_{23}^{[2]} \\ x_{31}^{[2]} & x_{32}^{[2]} & x_{33}^{[2]} \end{pmatrix}$$

$$X_{3\times3}^{[3]} = \begin{pmatrix} x_{11}^{[3]} & x_{12}^{[3]} & x_{13}^{[3]} \\ x_{21}^{[3]} & x_{22}^{[3]} & x_{23}^{[3]} \\ x_{31}^{[3]} & x_{32}^{[3]} & x_{33}^{[3]} \end{pmatrix}$$

我们将所有相关参数均用矩阵和向量方式表达出来了。接着我们就使用这些表达符号来进行反向传播的说明。我们先给出$\frac{\partial L}{\partial K}$以及$\frac{\partial L}{\partial b}$的求解公式。

$$\frac{\partial L}{\partial K} = \frac{\partial L}{\partial Y} \cdot (\mathrm{im2col}(\mathrm{X}))^T$$

$$\frac{\partial L}{\partial b} = sum(\frac{\partial L}{\partial Y}, axis = 0)/B$$

这是根据$Y_{[B,(H-k+1)\times(W-k+1),C_{out}]} = \mathrm{im2col}(\mathrm{X})_{[B,(H-k+1)\times(W-k+1),k\times k\times C_{in}]} \cdot K_{[k\times k\times C_{in},C_{out}]}$得到的，其实和全连接反向传播求导的方式没有区别。但是得到的$\frac{\partial L}{\partial K}$的形状是$(k\times k\times, C_{out})$，我们做一次 reshape 操作，将形状转为$(k, k, C_{in,}, C_{out})$即可。$\frac{\partial L}{\partial b}$的结果是因为，首先$b.shape = (1, C_{out})$。$\frac{\partial L}{\partial Y}.shape = ((H-k+1)\times(W-k+1), C_{out})$，$Y$和$\frac{\partial L}{\partial Y}$的形状是一致的。所以$b$是

通过广播机制将行扩展为$(H-k+1)\times(W-k+1)$这么多，然后和Y相加得到的。最后反向传播的求导的时候，$\frac{\partial Y}{\partial b}=1$，所以是按照行顺序将$\frac{\partial L}{\partial Y}$相加，除以批量数B，求得平均值。

最后我们再看下$\frac{\partial L}{\partial X}$。我们根据这个等式：

$$Y_{[B,(H-k+1)\times(W-k+1),C_{out}]}=\mathrm{im2col(X)}_{[B,(H-k+1)\times(W-k+1),k\times k\times C_{in}]}\cdot K_{[k\times k\times C_{in},C_{out}]}$$

可以得到$\frac{\partial L}{\partial \mathrm{im2col(X)}}$，但是将$\frac{\partial L}{\partial \mathrm{im2col(X)}}$还原到$\frac{\partial L}{\partial X}$，是非常复杂艰难的，计算量非常大。所以我们需要给出另外一种计算方式。根据前向传播的计算过程和链式求导法则$\frac{\partial L}{\partial X}=\frac{\partial L}{\partial Y}\cdot\frac{\partial Y}{\partial X}$，我们可以写出$\frac{\partial L}{\partial X}$的每个元素的表达式，我们列举$X$的第一个通道的相关导数。

$$\frac{\partial L}{\partial x_{11}^{[1]}}=0\times k_{22}^{\{1\}[1]}+0\times k_{21}^{\{1\}[1]}+0\times k_{12}^{\{1\}[1]}+\frac{\partial L}{\partial y_{11}^{\{1\}}}k_{11}^{\{1\}[1]}$$

$$+0\times k_{22}^{\{2\}[1]}+0\times k_{21}^{\{2\}[1]}+0\times k_{12}^{\{2\}[1]}+\frac{\partial L}{\partial y_{11}^{\{2\}}}k_{11}^{\{2\}[1]}$$

$$\frac{\partial L}{\partial x_{12}^{[1]}}=0\times k_{22}^{\{1\}[1]}+0\times k_{21}^{\{1\}[1]}+\frac{\partial L}{\partial y_{11}^{\{1\}}}k_{12}^{\{1\}[1]}+\frac{\partial L}{\partial y_{12}^{\{1\}}}k_{11}^{\{1\}[1]}$$

$$+0\times k_{22}^{\{2\}[1]}+0\times k_{21}^{\{2\}[1]}+\frac{\partial L}{\partial y_{11}^{\{2\}}}k_{12}^{\{2\}[1]}+\frac{\partial L}{\partial y_{12}^{\{2\}}}k_{11}^{\{2\}[1]}$$

$$\frac{\partial L}{\partial x_{13}^{[1]}}=0\times k_{22}^{\{1\}[1]}+0\times k_{21}^{\{1\}[1]}+\frac{\partial L}{\partial y_{12}^{\{1\}}}k_{12}^{\{1\}[1]}+0\times k_{11}^{\{1\}[1]}$$

$$+0\times k_{22}^{\{2\}[1]}+0\times k_{21}^{\{2\}[1]}+\frac{\partial L}{\partial y_{12}^{\{2\}}}k_{12}^{\{2\}[1]}+0\times k_{11}^{\{2\}[1]}$$

$$\frac{\partial L}{\partial x_{21}^{[1]}}=0\times k_{22}^{\{1\}[1]}+\frac{\partial L}{\partial y_{11}^{\{1\}}}k_{21}^{\{1\}[1]}+0\times k_{12}^{\{1\}[1]}+\frac{\partial L}{\partial y_{21}^{\{1\}}}\times k_{11}^{\{1\}[1]}$$

$$+0\times k_{22}^{\{2\}[1]}+\frac{\partial L}{\partial y_{11}^{\{2\}}}\times k_{21}^{\{2\}[1]}+\frac{\partial L}{\partial y_{12}^{\{2\}}}\times k_{12}^{\{2\}[1]}+\frac{\partial L}{\partial y_{21}^{\{2\}}}k_{11}^{\{2\}[1]}$$

$$\frac{\partial L}{\partial x_{22}^{[1]}}=\frac{\partial L}{\partial y_{11}^{\{1\}}}k_{22}^{\{1\}[1]}+\frac{\partial L}{\partial y_{12}^{\{1\}}}k_{21}^{\{1\}[1]}+\frac{\partial L}{\partial y_{21}^{\{1\}}}k_{12}^{\{1\}[1]}+\frac{\partial L}{\partial y_{22}^{\{1\}}}\times k_{11}^{\{1\}[1]}$$

$$+\frac{\partial L}{\partial y_{11}^{\{2\}}}k_{22}^{\{2\}[1]}+\frac{\partial L}{\partial y_{12}^{\{2\}}}k_{21}^{\{2\}[1]}+\frac{\partial L}{\partial y_{21}^{\{2\}}}k_{12}^{\{2\}[1]}+\frac{\partial L}{\partial y_{22}^{\{2\}}}k_{11}^{\{2\}[1]}$$

$$\frac{\partial L}{\partial x_{23}^{[1]}}=\frac{\partial L}{\partial y_{12}^{\{1\}}}k_{22}^{\{1\}[1]}+0\times k_{21}^{\{1\}[1]}+\frac{\partial L}{\partial y_{22}^{\{1\}}}k_{12}^{\{1\}[1]}+0\times k_{11}^{\{1\}[1]}$$

$$+\frac{\partial L}{\partial y_{12}^{\{2\}}}k_{22}^{\{2\}[1]}+0\times k_{21}^{\{2\}[1]}+\frac{\partial L}{\partial y_{22}^{\{2\}}}\times k_{12}^{\{2\}[1]}+0\times k_{11}^{\{2\}[1]}$$

$$\frac{\partial L}{\partial x_{31}^{[1]}}=0\times k_{22}^{\{1\}[1]}+\frac{\partial L}{\partial y_{21}^{\{1\}}}k_{21}^{\{1\}[1]}+0\times k_{12}^{\{1\}[1]}+0\times k_{11}^{\{1\}[1]}$$

$$+0\times k_{22}^{\{2\}[1]}+\frac{\partial L}{\partial y_{21}^{\{2\}}}k_{21}^{\{2\}[1]}+0\times k_{12}^{\{2\}[1]}+0\times k_{11}^{\{2\}[1]}$$

$$\frac{\partial L}{\partial x_{32}^{[1]}}=\frac{\partial L}{\partial y_{21}^{\{1\}}}k_{22}^{\{1\}[1]}+\frac{\partial L}{\partial y_{22}^{\{1\}}}k_{21}^{\{1\}[1]}+0\times k_{12}^{\{1\}[1]}+0\times k_{11}^{\{1\}[1]}$$

$$+\frac{\partial L}{\partial y_{21}^{\{2\}}}k_{22}^{\{2\}[1]}+\frac{\partial L}{\partial y_{22}^{\{2\}}}k_{21}^{\{2\}[1]}+\frac{\partial L}{\partial y_{22}^{\{2\}}}\times k_{12}^{\{2\}[1]}+0\times k_{11}^{\{2\}[1]}$$

$$\frac{\partial L}{\partial x_{33}^{[1]}}=\frac{\partial L}{\partial y_{22}^{\{1\}}}k_{22}^{\{1\}[1]}+0\times k_{21}^{\{1\}[1]}+0\times k_{12}^{\{1\}[1]}+0\times k_{11}^{\{1\}[1]}$$

$$+\frac{\partial L}{\partial y_{22}^{\{2\}}}k_{22}^{\{2\}[1]}+0\times k_{21}^{\{2\}[1]}+0\times k_{12}^{\{2\}[1]}+0\times k_{11}^{\{2\}[1]}$$

那么我们可以写成以下形式:

$$\begin{pmatrix}\frac{\partial L}{\partial x_{11}^{[1]}}\\ \frac{\partial L}{\partial x_{12}^{[1]}}\\ \frac{\partial L}{\partial x_{13}^{[1]}}\\ \frac{\partial L}{\partial x_{21}^{[1]}}\\ \frac{\partial L}{\partial x_{22}^{[1]}}\\ \frac{\partial L}{\partial x_{23}^{[1]}}\\ \frac{\partial L}{\partial x_{31}^{[1]}}\\ \frac{\partial L}{\partial x_{32}^{[1]}}\\ \frac{\partial L}{\partial x_{33}^{[1]}}\end{pmatrix}=\begin{pmatrix}0 & 0 & 0 & \frac{\partial L}{\partial y_{11}^{\{1\}}} & 0 & 0 & 0 & \frac{\partial L}{\partial y_{11}^{\{2\}}}\\ 0 & 0 & \frac{\partial L}{\partial y_{11}^{\{1\}}} & \frac{\partial L}{\partial y_{12}^{\{1\}}} & 0 & 0 & \frac{\partial L}{\partial y_{11}^{\{2\}}} & \frac{\partial L}{\partial y_{12}^{\{2\}}}\\ 0 & 0 & \frac{\partial L}{\partial y_{12}^{\{1\}}} & 0 & 0 & 0 & \frac{\partial L}{\partial y_{12}^{\{2\}}} & 0\\ 0 & \frac{\partial L}{\partial y_{11}^{\{1\}}} & 0 & \frac{\partial L}{\partial y_{21}^{\{1\}}} & 0 & \frac{\partial L}{\partial y_{11}^{\{2\}}} & 0 & \frac{\partial L}{\partial y_{21}^{\{2\}}}\\ \frac{\partial L}{\partial y_{11}^{\{1\}}} & \frac{\partial L}{\partial y_{12}^{\{1\}}} & \frac{\partial L}{\partial y_{21}^{\{1\}}} & \frac{\partial L}{\partial y_{22}^{\{1\}}} & \frac{\partial L}{\partial y_{11}^{\{2\}}} & \frac{\partial L}{\partial y_{12}^{\{2\}}} & \frac{\partial L}{\partial y_{21}^{\{2\}}} & \frac{\partial L}{\partial y_{22}^{\{2\}}}\\ \frac{\partial L}{\partial y_{12}^{\{1\}}} & 0 & \frac{\partial L}{\partial y_{22}^{\{1\}}} & 0 & \frac{\partial L}{\partial y_{12}^{\{2\}}} & 0 & \frac{\partial L}{\partial y_{22}^{\{2\}}} & 0\\ 0 & \frac{\partial L}{\partial y_{21}^{\{1\}}} & 0 & 0 & 0 & \frac{\partial L}{\partial y_{21}^{\{2\}}} & 0 & 0\\ \frac{\partial L}{\partial y_{21}^{\{1\}}} & \frac{\partial L}{\partial y_{22}^{\{1\}}} & 0 & 0 & \frac{\partial L}{\partial y_{21}^{\{2\}}} & \frac{\partial L}{\partial y_{22}^{\{2\}}} & 0 & 0\\ \frac{\partial L}{\partial y_{22}^{\{1\}}} & 0 & 0 & 0 & \frac{\partial L}{\partial y_{22}^{\{2\}}} & 0 & 0 & 0\end{pmatrix}\cdot\begin{pmatrix}k_{22}^{\{1\}[1]}\\ k_{21}^{\{1\}[1]}\\ k_{12}^{\{1\}[1]}\\ k_{11}^{\{1\}[1]}\\ k_{22}^{\{2\}[1]}\\ k_{21}^{\{2\}[1]}\\ k_{12}^{\{2\}[1]}\\ k_{11}^{\{2\}[1]}\end{pmatrix}$$

这是一个通道的，同理，按照我们之前的推导过程。如果是多个通道，则有如下等式:

$$\begin{pmatrix} \frac{\partial L}{\partial x_{11}^{[1]}} & \frac{\partial L}{\partial x_{11}^{[2]}} & \frac{\partial L}{\partial x_{11}^{[3]}} \\ \frac{\partial L}{\partial x_{12}^{[1]}} & \frac{\partial L}{\partial x_{12}^{[2]}} & \frac{\partial L}{\partial x_{12}^{[3]}} \\ \frac{\partial L}{\partial x_{13}^{[1]}} & \frac{\partial L}{\partial x_{13}^{[2]}} & \frac{\partial L}{\partial x_{13}^{[3]}} \\ \frac{\partial L}{\partial x_{21}^{[1]}} & \frac{\partial L}{\partial x_{21}^{[2]}} & \frac{\partial L}{\partial x_{21}^{[3]}} \\ \frac{\partial L}{\partial x_{22}^{[1]}} & \frac{\partial L}{\partial x_{22}^{[2]}} & \frac{\partial L}{\partial x_{22}^{[3]}} \\ \frac{\partial L}{\partial x_{23}^{[1]}} & \frac{\partial L}{\partial x_{23}^{[2]}} & \frac{\partial L}{\partial x_{23}^{[3]}} \\ \frac{\partial L}{\partial x_{31}^{[1]}} & \frac{\partial L}{\partial x_{31}^{[2]}} & \frac{\partial L}{\partial x_{31}^{[3]}} \\ \frac{\partial L}{\partial x_{32}^{[1]}} & \frac{\partial L}{\partial x_{32}^{[2]}} & \frac{\partial L}{\partial x_{32}^{[3]}} \\ \frac{\partial L}{\partial x_{33}^{[1]}} & \frac{\partial L}{\partial x_{33}^{[2]}} & \frac{\partial L}{\partial x_{33}^{[3]}} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & \frac{\partial L}{\partial y_{11}^{\{1\}}} & 0 & 0 & 0 & \frac{\partial L}{\partial y_{11}^{\{2\}}} \\ 0 & 0 & \frac{\partial L}{\partial y_{11}^{\{1\}}} & \frac{\partial L}{\partial y_{12}^{\{1\}}} & 0 & 0 & \frac{\partial L}{\partial y_{11}^{\{2\}}} & \frac{\partial L}{\partial y_{12}^{\{2\}}} \\ 0 & 0 & \frac{\partial L}{\partial y_{12}^{\{1\}}} & 0 & 0 & 0 & \frac{\partial L}{\partial y_{12}^{\{2\}}} & 0 \\ 0 & \frac{\partial L}{\partial y_{11}^{\{1\}}} & 0 & \frac{\partial L}{\partial y_{21}^{\{1\}}} & 0 & \frac{\partial L}{\partial y_{11}^{\{2\}}} & 0 & \frac{\partial L}{\partial y_{21}^{\{2\}}} \\ \frac{\partial L}{\partial y_{11}^{\{1\}}} & \frac{\partial L}{\partial y_{12}^{\{1\}}} & \frac{\partial L}{\partial y_{21}^{\{1\}}} & \frac{\partial L}{\partial y_{22}^{\{1\}}} & \frac{\partial L}{\partial y_{11}^{\{2\}}} & \frac{\partial L}{\partial y_{12}^{\{2\}}} & \frac{\partial L}{\partial y_{21}^{\{2\}}} & \frac{\partial L}{\partial y_{22}^{\{2\}}} \\ \frac{\partial L}{\partial y_{12}^{\{1\}}} & 0 & \frac{\partial L}{\partial y_{22}^{\{1\}}} & 0 & \frac{\partial L}{\partial y_{12}^{\{2\}}} & 0 & \frac{\partial L}{\partial y_{22}^{\{2\}}} & 0 \\ 0 & \frac{\partial L}{\partial y_{21}^{\{1\}}} & 0 & 0 & 0 & \frac{\partial L}{\partial y_{21}^{\{2\}}} & 0 & 0 \\ \frac{\partial L}{\partial y_{21}^{\{1\}}} & \frac{\partial L}{\partial y_{22}^{\{1\}}} & 0 & 0 & \frac{\partial L}{\partial y_{21}^{\{2\}}} & \frac{\partial L}{\partial y_{22}^{\{2\}}} & 0 & 0 \\ \frac{\partial L}{\partial y_{22}^{\{1\}}} & 0 & 0 & 0 & \frac{\partial L}{\partial y_{22}^{\{2\}}} & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} k_{22}^{\{1\}[1]} & k_{22}^{\{1\}[2]} & k_{22}^{\{1\}[3]} \\ k_{21}^{\{1\}[1]} & k_{21}^{\{1\}[2]} & k_{21}^{\{1\}[3]} \\ k_{12}^{\{1\}[1]} & k_{12}^{\{1\}[2]} & k_{12}^{\{1\}[3]} \\ k_{11}^{\{1\}[1]} & k_{11}^{\{1\}[2]} & k_{11}^{\{1\}[3]} \\ k_{22}^{\{2\}[1]} & k_{22}^{\{2\}[2]} & k_{22}^{\{2\}[3]} \\ k_{21}^{\{2\}[1]} & k_{21}^{\{2\}[2]} & k_{21}^{\{2\}[3]} \\ k_{12}^{\{2\}[1]} & k_{12}^{\{2\}[2]} & k_{12}^{\{2\}[3]} \\ k_{11}^{\{2\}[1]} & k_{11}^{\{2\}[2]} & k_{11}^{\{2\}[3]} \end{pmatrix}$$

我们将这几个矩阵从左到右命名为$\nabla X_{9\times 3}$、$\nabla Y_{9\times 8}$、$\nabla K_{8\times 3}$。它们满足$\nabla X = \nabla Y \cdot \nabla K$。

$$\Delta X^{[1]} = \begin{pmatrix} \frac{\partial L}{\partial x_{11}^{[1]}} & \frac{\partial L}{\partial x_{12}^{[1]}} & \frac{\partial L}{\partial x_{13}^{[1]}} \\ \frac{\partial L}{\partial x_{21}^{[1]}} & \frac{\partial L}{\partial x_{22}^{[1]}} & \frac{\partial L}{\partial x_{23}^{[1]}} \\ \frac{\partial L}{\partial x_{31}^{[1]}} & \frac{\partial L}{\partial x_{32}^{[1]}} & \frac{\partial L}{\partial x_{33}^{[1]}} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial L}{\partial y_{11}^{\{1\}}} & \frac{\partial L}{\partial y_{12}^{\{1\}}} & 0 \\ 0 & \frac{\partial L}{\partial y_{21}^{\{1\}}} & \frac{\partial L}{\partial y_{22}^{\{1\}}} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} Conv \begin{pmatrix} k_{22}^{\{1\}[1]} & k_{21}^{\{1\}[1]} \\ k_{12}^{\{1\}[1]} & k_{11}^{\{1\}[1]} \end{pmatrix} +$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial L}{\partial y_{11}^{\{2\}}} & \frac{\partial L}{\partial y_{12}^{\{2\}}} & 0 \\ 0 & \frac{\partial L}{\partial y_{21}^{\{2\}}} & \frac{\partial L}{\partial y_{22}^{\{2\}}} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} Conv \begin{pmatrix} k_{22}^{\{2\}[1]} & k_{21}^{\{2\}[1]} \\ k_{12}^{\{2\}[1]} & k_{11}^{\{2\}[1]} \end{pmatrix}$$

$\Delta X^{[2]}$和$\Delta X^{[3]}$也是如此。就不再罗列出。从这个式子可以看到这是个卷积操作。$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial L}{\partial y_{11}^{\{1\}}} & \frac{\partial L}{\partial y_{12}^{\{1\}}} & 0 \\ 0 & \frac{\partial L}{\partial y_{21}^{\{1\}}} & \frac{\partial L}{\partial y_{22}^{\{1\}}} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$和$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial L}{\partial y_{11}^{\{2\}}} & \frac{\partial L}{\partial y_{12}^{\{2\}}} & 0 \\ 0 & \frac{\partial L}{\partial y_{21}^{\{2\}}} & \frac{\partial L}{\partial y_{22}^{\{2\}}} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$就是$\frac{\partial L}{\partial Y_{2\times2\times2}}$矩阵进行了填充 0 的操作。填充之后得到的矩阵和旋转 180 度后的卷积核$K_{2\times2\times3\times2}$做卷积运算。得到的结果的高$(H)$和宽$W$就是原输入数据$X_{3\times3\times3}$的高$(H)$和宽$W$，这里是 3。至于填充多少呢，设置填充大小$P$，卷积核大小为$k$。给出计算填充大小的公式:

$$OH = \frac{H + 2P - k}{S} + 1, OW = \frac{W + 2P - k}{S} + 1$$

这里的OH和OW是卷积后的输出大小，即是$X_{3\times3\times3}$的高(H)和宽W，也就是 3。H和W就是$\frac{\partial L}{\partial Y_{2\times2\times2}}$的高$(H)$和宽$W$，均是 2。$S$是步长，这里是 1。代入公式计算得到$P = 1$。表示在四周都填充一个单元。原来的$H$和$W$是 2，经过此填充变为 4。即是$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial L}{\partial y_{11}^{\{1\}}} & \frac{\partial L}{\partial y_{12}^{\{1\}}} & 0 \\ 0 & \frac{\partial L}{\partial y_{21}^{\{1\}}} & \frac{\partial L}{\partial y_{22}^{\{1\}}} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$和$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial L}{\partial y_{11}^{\{2\}}} & \frac{\partial L}{\partial y_{12}^{\{2\}}} & 0 \\ 0 & \frac{\partial L}{\partial y_{21}^{\{2\}}} & \frac{\partial L}{\partial y_{22}^{\{2\}}} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$。这里要注意的是此时的卷积核$K_{2\times2\times3\times2}$不仅仅是旋转 180 度，而且它的$C_{in}$和$C_{out}$发生了互换。从前面我们推导的式子不难发现这一点。那么这种操作和原先前向传播做卷积的操作，似乎是个逆操作。之前对输入数据X通过卷积核$K_{[k,k,C_{in},C_{out}]}$，输出Y。而现在我们实行的这个卷积是将$\frac{\partial L}{\partial Y}$通过对卷积核旋转 180，再对$K_{[k,k,C_{in},C_{out}]}$做轴变换，转换成$K_{[k,k,C_{out},C_{in}]}$，然后执行卷积操作输出$\frac{\partial L}{\partial X}$。这里$C_{out}, C_{in}$发生了对换，所以要对要对$C_{in}$组进行这个操作，每组有$k \times k \times C_{out}$元素。

而$\nabla Y_{9\times 8}=\begin{pmatrix} 0 & 0 & 0 & \frac{\partial L}{\partial y_{11}^{\{1\}}} & 0 & 0 & 0 & \frac{\partial L}{\partial y_{11}^{\{2\}}} \\ 0 & 0 & \frac{\partial L}{\partial y_{11}^{\{1\}}} & \frac{\partial L}{\partial y_{12}^{\{1\}}} & 0 & 0 & \frac{\partial L}{\partial y_{11}^{\{2\}}} & \frac{\partial L}{\partial y_{12}^{\{2\}}} \\ 0 & 0 & \frac{\partial L}{\partial y_{12}^{\{1\}}} & 0 & 0 & 0 & \frac{\partial L}{\partial y_{12}^{\{2\}}} & 0 \\ 0 & \frac{\partial L}{\partial y_{11}^{\{1\}}} & 0 & \frac{\partial L}{\partial y_{21}^{\{1\}}} & 0 & \frac{\partial L}{\partial y_{11}^{\{2\}}} & 0 & \frac{\partial L}{\partial y_{21}^{\{2\}}} \\ \frac{\partial L}{\partial y_{11}^{\{1\}}} & \frac{\partial L}{\partial y_{12}^{\{1\}}} & \frac{\partial L}{\partial y_{21}^{\{1\}}} & \frac{\partial L}{\partial y_{22}^{\{1\}}} & \frac{\partial L}{\partial y_{11}^{\{2\}}} & \frac{\partial L}{\partial y_{12}^{\{2\}}} & \frac{\partial L}{\partial y_{21}^{\{2\}}} & \frac{\partial L}{\partial y_{22}^{\{2\}}} \\ \frac{\partial L}{\partial y_{12}^{\{1\}}} & 0 & \frac{\partial L}{\partial y_{22}^{\{1\}}} & 0 & \frac{\partial L}{\partial y_{12}^{\{2\}}} & 0 & \frac{\partial L}{\partial y_{22}^{\{2\}}} & 0 \\ 0 & \frac{\partial L}{\partial y_{21}^{\{1\}}} & 0 & 0 & 0 & \frac{\partial L}{\partial y_{21}^{\{2\}}} & 0 & 0 \\ \frac{\partial L}{\partial y_{21}^{\{1\}}} & \frac{\partial L}{\partial y_{22}^{\{1\}}} & 0 & 0 & \frac{\partial L}{\partial y_{21}^{\{2\}}} & \frac{\partial L}{\partial y_{22}^{\{2\}}} & 0 & 0 \\ \frac{\partial L}{\partial y_{22}^{\{1\}}} & 0 & 0 & 0 & \frac{\partial L}{\partial y_{22}^{\{2\}}} & 0 & 0 & 0 \end{pmatrix}$就是$\frac{\partial L}{\partial Y_{2\times 2\times 2}}$矩阵进行了填充 0，再执行im2col操作就可以得到。我们给出一个具体的公式:

$$\nabla Y=\mathrm{im2col}(\mathrm{Padding}(\frac{\partial \mathrm{L}}{\partial \mathrm{Y}}))$$

得到∇Y后，再做之前将的卷积操作就得到了∇X。此时∇X的 shape 是$[(\frac{H+2P-k}{S}+1)\times\left(\frac{W+2P-k}{S}+1\right),C_{in}]$。进行一下 reshape 调整为$[(\frac{H+2P-k}{S}+1,(\frac{W+2P-k}{S}+1,C_{in}]$。问题就得到解决了。好了这就是卷积的反向传播过程。最后给出如何对卷积核旋转 180 度再调整C_{in}和C_{out}两个轴向的 python 代码:

```
#conv.weights 维度为:(k, k, C_in, C_out)
swap_weights = conv.weights.swapaxes(2,3) #调整C_in和C_out两个轴
re_weights = swap_weights.reshape(-1, swap_weights.shape[2], swap_weights.shape[3])
flip_weights = re_weights [::-1,...] #实现旋转 180 度的操作
flip_weights = flip_weights.reshape(-1, swap_weights.shape[3])
return flip_weights
```