卷积神经网络的反向传播

传统的神经网络是全连接形式的，如果进行反向传播，只需要由下一层对前一层不断的求偏导，即求链式偏导就可以求出每一层的误差敏感项，然后求出权重和偏置项的梯度，即可更新权重。而卷积神经网络有两个特殊的层：卷积层和池化层。池化层输出时不需要经过激活函数，是一个滑动窗口的最大值，一个常数，那么它的偏导是1。池化层相当于对上层图片做了一个压缩，这个反向求误差敏感项时与传统的反向传播方式不同。从卷积后的feature_map反向传播到前一层时，由于前向传播时是通过卷积核做卷积运算得到的feature_map，所以反向传播与传统的也不一样，需要更新卷积核的参数。下面我们介绍一下池化层和卷积层是如何做反向传播的。

在介绍之前，首先回顾一下传统的反向传播方法：

1.通过前向传播计算每一层的输入值$net_{i,j}$ (如卷积后的feature_map的第一个神经元的输入：$net_{i_{11}}$)

2.反向传播计算每个神经元的误差项$\delta_{i,j}$ ，$\delta_{i,j} = \frac{\partial E}{\partial net_{i,j}}$ ，其中E为损失函数计算得到的总体误差，可以用平方差，交叉熵等表示。

3.计算每个神经元权重$w_{i,j}$ 的梯度，$\eta_{i,j} = \frac{\partial E}{\partial net_{i,j}} \cdot \frac{\partial net_{i,j}}{\partial w_{i,j}} = \delta_{i,j} \cdot out_{i,j}$

4.更新权重 $w_{i,j} = w_{i,j} - \lambda \cdot \eta_{i,j}$(其中$\lambda$为学习率)

卷积层的反向传播

由前向传播可得：

每一个神经元的值都是上一个神经元的输入作为这个神经元的输入，经过激活函数激活之后输出，作为下一个神经元的输入，在这里我用i_{11}表示前一层,o_{11}表示i_{11}的下一层。那么$net_{i_{11}}$就是i11这个神经元的输入，$out_{i_{11}}$就是i11这个神经元的输出，同理，$net_{o_{11}}$就是o11这个神经元的输入，$out_{o_{11}}$就是o_{11}这个神经元的输出,因为上一层神经元的输出 = 下一层神经元的输入，所以$out_{i_{11}}$= $net_{o_{11}}$ ，这里我为了简化，直接把$out_{i_{11}}$记为i_{11}

$$
\begin{aligned}
i_{11} &= out_{i_{11}} \\
&= activators(net_{i_{11}}) \\
net_{o_{11}} &= conv(input, filter) \\
&= i_{11} \times h_{11} + i_{12} \times h_{12} + i_{21} \times h_{21} + i_{22} \times h_{22} \\
out_{o_{11}} &= activators(net_{o_{11}}) \\
&= max(0, net_{o_{11}})
\end{aligned} \tag{4}
$$

$net_{i_{11}}$表示上一层的输入，$out_{i_{11}}$表示上一层的输出

首先计算卷积的上一层的第一个元素i_{11}的误差项δ_{11}：

$$
\delta_{11} = \frac{\partial E}{\partial net_{i_{11}}} = \frac{\partial E}{\partial out_{i_{11}}} \cdot \frac{\partial out_{i_{11}}}{\partial net_{i_{11}}} = \frac{\partial E}{\partial i_{11}} \cdot \frac{\partial i_{11}}{\partial net_{i_{11}}}
$$

先计算$\frac{\partial E}{\partial i_{11}}$

此处我们并不清楚$\frac{\partial E}{\partial i_{11}}$怎么算，那可以先把input层通过卷积核做完卷积运算后的输出feature_map写出来:

$$
\begin{aligned}
net_{o_{11}} &= i_{11} \times h_{11} + i_{12} \times h_{12} + i_{21} \times h_{21} + i_{22} \times h_{22} \\
net_{o_{12}} &= i_{12} \times h_{11} + i_{13} \times h_{12} + i_{22} \times h_{21} + i_{23} \times h_{22} \\
net_{o_{12}} &= i_{13} \times h_{11} + i_{14} \times h_{12} + i_{23} \times h_{21} + i_{24} \times h_{22} \\
net_{o_{21}} &= i_{21} \times h_{11} + i_{22} \times h_{12} + i_{31} \times h_{21} + i_{32} \times h_{22} \\
net_{o_{22}} &= i_{22} \times h_{11} + i_{23} \times h_{12} + i_{32} \times h_{21} + i_{33} \times h_{22} \\
net_{o_{23}} &= i_{23} \times h_{11} + i_{24} \times h_{12} + i_{33} \times h_{21} + i_{34} \times h_{22} \\
net_{o_{31}} &= i_{31} \times h_{11} + i_{32} \times h_{12} + i_{41} \times h_{21} + i_{42} \times h_{22} \\
net_{o_{32}} &= i_{32} \times h_{11} + i_{33} \times h_{12} + i_{42} \times h_{21} + i_{43} \times h_{22} \\
net_{o_{33}} &= i_{33} \times h_{11} + i_{34} \times h_{12} + i_{43} \times h_{21} + i_{44} \times h_{22}
\end{aligned} \tag{5}
$$

然后依次对输入元素$i_{i,j}$求偏导

i_{11}的偏导：

$$\begin{aligned}\frac{\partial E}{\partial i_{11}} &= \frac{\partial E}{\partial net_{o_{11}}} \cdot \frac{\partial net_{o_{11}}}{\partial i_{11}} \\ &= \delta_{11} \cdot h_{11}\end{aligned} \tag{6}$$

i_{12}的偏导：

$$\begin{aligned}\frac{\partial E}{\partial i_{12}} &= \frac{\partial E}{\partial net_{o_{11}}} \cdot \frac{\partial net_{o_{11}}}{\partial i_{12}} + \frac{\partial E}{\partial net_{o_{12}}} \cdot \frac{\partial net_{o_{12}}}{\partial i_{12}} \\ &= \delta_{11} \cdot h_{12} + \delta_{12} \cdot h_{11}\end{aligned} \tag{7}$$

i_{13}的偏导：

$$\begin{aligned}\frac{\partial E}{\partial i_{13}} &= \frac{\partial E}{\partial net_{o_{12}}} \cdot \frac{\partial net_{o_{12}}}{\partial i_{13}} + \frac{\partial E}{\partial net_{o_{13}}} \cdot \frac{\partial net_{o_{13}}}{\partial i_{13}} \\ &= \delta_{12} \cdot h_{12} + \delta_{13} \cdot h_{11}\end{aligned} \tag{8}$$

i_{21}的偏导：

$$\begin{aligned}\frac{\partial E}{\partial i_{21}} &= \frac{\partial E}{\partial net_{o_{11}}} \cdot \frac{\partial net_{o_{11}}}{\partial i_{21}} + \frac{\partial E}{\partial net_{o_{21}}} \cdot \frac{\partial net_{o_{21}}}{\partial i_{21}} \\ &= \delta_{11} \cdot h_{21} + \delta_{21} \cdot h_{11}\end{aligned} \tag{9}$$

i_{22}的偏导：

$$\begin{aligned}\frac{\partial E}{\partial i_{22}} &= \frac{\partial E}{\partial net_{o_{11}}} \cdot \frac{\partial net_{o_{11}}}{\partial i_{22}} + \frac{\partial E}{\partial net_{o_{12}}} \cdot \frac{\partial net_{o_{12}}}{\partial i_{22}} \\ &+ \frac{\partial E}{\partial net_{o_{21}}} \cdot \frac{\partial net_{o_{21}}}{\partial i_{22}} + \frac{\partial E}{\partial net_{o_{22}}} \cdot \frac{\partial net_{o_{22}}}{\partial i_{22}} \\ &= \delta_{11} \cdot h_{22} + \delta_{12} \cdot h_{21} + \delta_{21} \cdot h_{12} + \delta_{22} \cdot h_{11}\end{aligned} \tag{10}$$

观察一下上面几个式子的规律，归纳一下，可以得到如下表达式：

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \delta_{11} & \delta_{12} & \delta_{13} & 0 \\ 0 & \delta_{21} & \delta_{22} & \delta_{23} & 0 \\ 0 & \delta_{31} & \delta_{32} & \delta_{33} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} h_{22} & h_{21} \\ h_{12} & h_{11} \end{bmatrix} = \begin{bmatrix} \frac{\partial E}{\partial i_{11}} & \frac{\partial E}{\partial i_{12}} & \frac{\partial E}{\partial i_{13}} & \frac{\partial E}{\partial i_{14}} \\ \frac{\partial E}{\partial i_{21}} & \frac{\partial E}{\partial i_{22}} & \frac{\partial E}{\partial i_{23}} & \frac{\partial E}{\partial i_{24}} \\ \frac{\partial E}{\partial i_{31}} & \frac{\partial E}{\partial i_{32}} & \frac{\partial E}{\partial i_{33}} & \frac{\partial E}{\partial i_{34}} \\ \frac{\partial E}{\partial i_{41}} & \frac{\partial E}{\partial i_{42}} & \frac{\partial E}{\partial i_{43}} & \frac{\partial E}{\partial i_{44}} \end{bmatrix} \tag{11}$$

图中的卷积核进行了180°翻转，与这一层的误差敏感项矩阵$delta_{i,j}$)周围补零后的矩阵做卷积运算后，就可以得到$\frac{\partial E}{\partial i_{11}}$，即

$\frac{\partial E}{\partial i_{i,j}} = \sum_m \cdot \sum_n h_{m,n}\delta_{i+m,j+n}$

第一项求完后，我们来求第二项$\frac{\partial i_{11}}{\partial net_{i_{11}}}$

$$\begin{aligned} \because i_{11} &= out_{i_{11}} \\ &= activators(net_{i_{11}}) \\ \therefore \frac{\partial i_{11}}{\partial net_{i_{11}}} &= f'(net_{i_{11}}) \\ \therefore \delta_{11} &= \frac{\partial E}{\partial net_{i_{11}}} \\ &= \frac{\partial E}{\partial i_{11}} \cdot \frac{\partial i_{11}}{\partial net_{i_{11}}} \\ &= \sum_m \cdot \sum_n h_{m,n}\delta_{i+m,j+n} \cdot f'(net_{i_{11}}) \end{aligned} \tag{12}$$

此时我们的误差敏感矩阵就求完了，得到误差敏感矩阵后，即可求权重的梯度。

由于上面已经写出了卷积层的输入$net_{o_{11}}$与权重$h_{i,j}$之间的表达式，所以可以直接求出：

$$\begin{aligned} \frac{\partial E}{\partial h_{11}} &= \frac{\partial E}{\partial net_{o_{11}}} \cdot \frac{\partial net_{o_{11}}}{\partial h_{11}} + \dots \\ &+ \frac{\partial E}{\partial net_{o_{33}}} \cdot \frac{\partial net_{o_{33}}}{\partial h_{11}} \\ &= \delta_{11} \cdot h_{11} + \dots + \delta_{33} \cdot h_{11} \end{aligned} \tag{13}$$

推论出**权重的梯度**：

$$\frac{\partial E}{\partial h_{i,j}} = \sum_m \sum_n \delta_{m,n} out_{o_{i+m,j+n}} \tag{14}$$

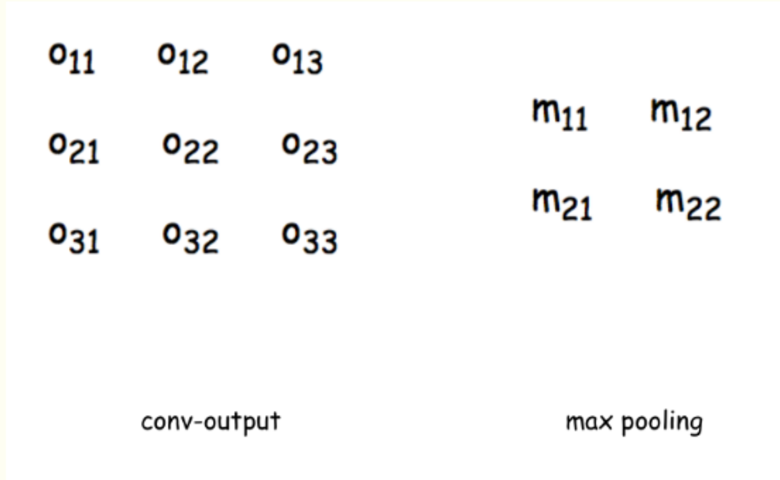
偏置项的梯度：

$$\begin{aligned} \frac{\partial E}{\partial b} &= \frac{\partial E}{\partial net_{o_{11}}} \frac{\partial net_{o_{11}}}{\partial w_b} + \frac{\partial E}{\partial net_{o_{12}}} \frac{\partial net_{o_{12}}}{\partial w_b} \\ &+ \frac{\partial E}{\partial net_{o_{21}}} \frac{\partial net_{o_{21}}}{\partial w_b} + \frac{\partial E}{\partial net_{o_{22}}} \frac{\partial net_{o_{22}}}{\partial w_b} \\ &= \delta_{11} + \delta_{12} + \delta_{21} + \delta_{22} \\ &= \sum_i \sum_j \delta_{i,j} \end{aligned} \tag{15}$$

可以看出，偏置项的偏导等于这一层所有误差敏感项之和。得到了权重和偏置项的梯度后，就可以根据梯度下降法更新权重和梯度了。

池化层的反向传播

池化层的反向传播就比较好求了，看着下面的图，左边是上一层的输出，也就是卷积层的输出feature_map，右边是池化层的输入，还是先根据前向传播，把式子都写出来，方便计算：



假设上一层这个滑动窗口的最大值是$out_{o_{11}}$

$$
\begin{aligned}
&\because net_{m_{11}} = max(out_{o_{11}}, out_{o_{12}}, out_{o_{21}}, out_{o_{22}}) \\
&\therefore \frac{\partial net_{m_{11}}}{\partial out_{o_{11}}} = 1 \\
&\frac{\partial net_{m_{11}}}{\partial out_{o_{12}}} = \frac{\partial net_{m_{11}}}{\partial out_{o_{21}}} = \frac{\partial net_{m_{11}}}{\partial out_{o_{22}}} = 0 \\
&\therefore \delta_{11}^{l-1} = \frac{\partial E}{\partial out_{o_{11}}} = \frac{\partial E}{\partial net_{m_{11}}} \cdot \frac{\partial net_{m_{11}}}{\partial out_{o_{11}}} = \delta_{11}^{l} \\
&\delta_{12}^{l-1} = \delta_{21}^{l-1} = \delta_{22}^{l-1} = 0
\end{aligned}
\tag{16}
$$

这样就求出了池化层的误差敏感项矩阵。同理可以求出每个神经元的梯度并更新权重。