

# 在CNN中的反向传播

## 1. 链式法则

先理解一下在反向传播中的链式法则。

假设有下面这个等式：

$$f(x, y, z) = (x + y)z$$

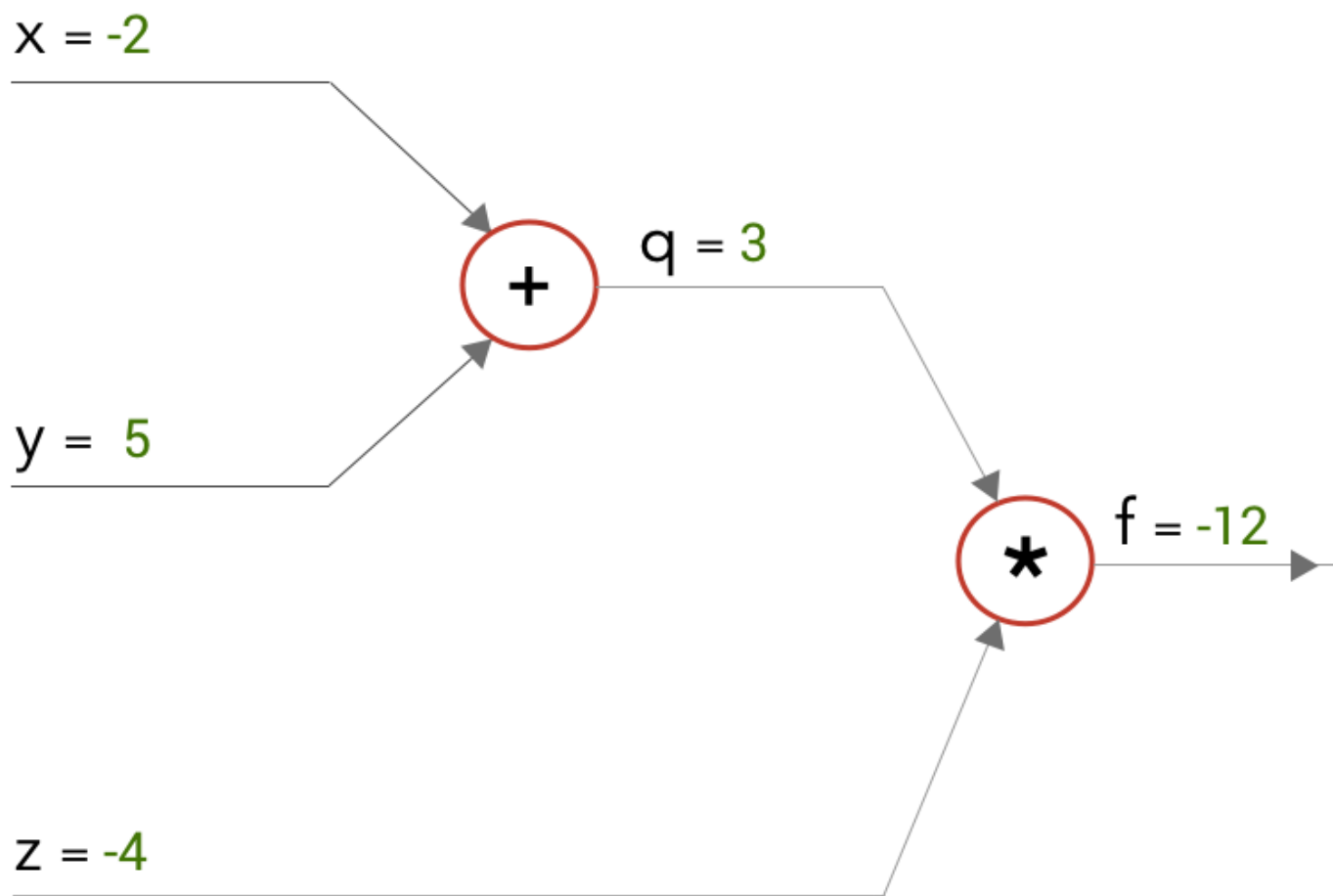
我们可以把它划分成两个等式：

$$f(x, y, z) = (x + y)z$$

$$q = x + y$$

$$f = q * z$$

下面让我们画出关于 $x, y, z$ 的计算图，其中 $x = -2, y = 5, z = 4$ ：



当我们按照上图从左到右进行计算时（前向传播），可以得到  $f = -12$ 。

现在让我们回到反向传播阶段。我们计算梯度从右往左，因此最后，我们可以得到关于我们的输入  $x, y, z$  的梯度： $\partial f / \partial x$ 、 $\partial f / \partial y$  和  $\partial f / \partial z$ 。

在从右往左进行计算时，在乘积门，我们可以得到  $\partial f / \partial q$  和  $\partial f / \partial z$ ，

在加和门我们可以得到  $\partial q / \partial x$  和  $\partial q / \partial y$ 。

$$x = -2$$

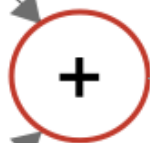
$$\frac{\partial f}{\partial x} = ?$$

$$y = 5$$

$$\frac{\partial f}{\partial y} = ?$$

$$z = -4$$

$$\frac{\partial f}{\partial z} = 3$$



$$q = 3$$

$$\frac{\partial f}{\partial q} = -4$$



$$f = -12$$

$$\frac{\partial f}{\partial f} = 1$$

$$f = q * z$$

$$\frac{\partial f}{\partial q} = z \mid z = -4$$

$$\frac{\partial f}{\partial z} = q \mid q = 3$$

$$q = x + y$$

$$\frac{\partial q}{\partial x} = 1 \quad \frac{\partial q}{\partial y} = 1$$

我们希望获得 $\partial f / \partial x$ 和 $\partial f / \partial y$ ，但是目前我们只获得了 $\partial q / \partial x$ 和 $\partial q / \partial y$ ，那么我们达到我们的目标

呢。

这里就可以使用链式法则来进行推导，通过链式法则，我们可以计算 $\partial f / \partial x$ ：

*Using chain rule:*

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} * \frac{\partial q}{\partial x}$$

那么我们可以计算得到 $\partial f / \partial x$ 和 $\partial f / \partial y$ 如下：

$$x = -2$$

$$\frac{\partial f}{\partial x} = -4$$

$$y = 5$$

$$\frac{\partial f}{\partial y} = -4$$

$$z = -4$$

$$\frac{\partial f}{\partial z} = 3$$



$$q = 3$$

$$\frac{\partial f}{\partial q} = -4$$



$$f = -12$$

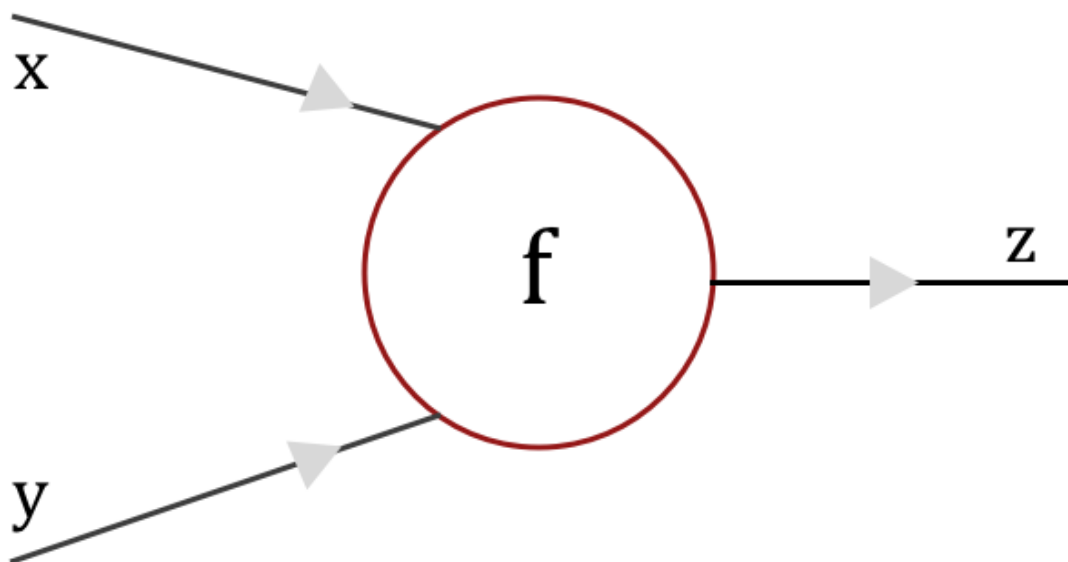
$$\frac{\partial f}{\partial f} = 1$$

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} * \frac{\partial q}{\partial x} = -4 * 1 = -4$$

$$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial q} * \frac{\partial q}{\partial y} = -4 * 1 = -4$$

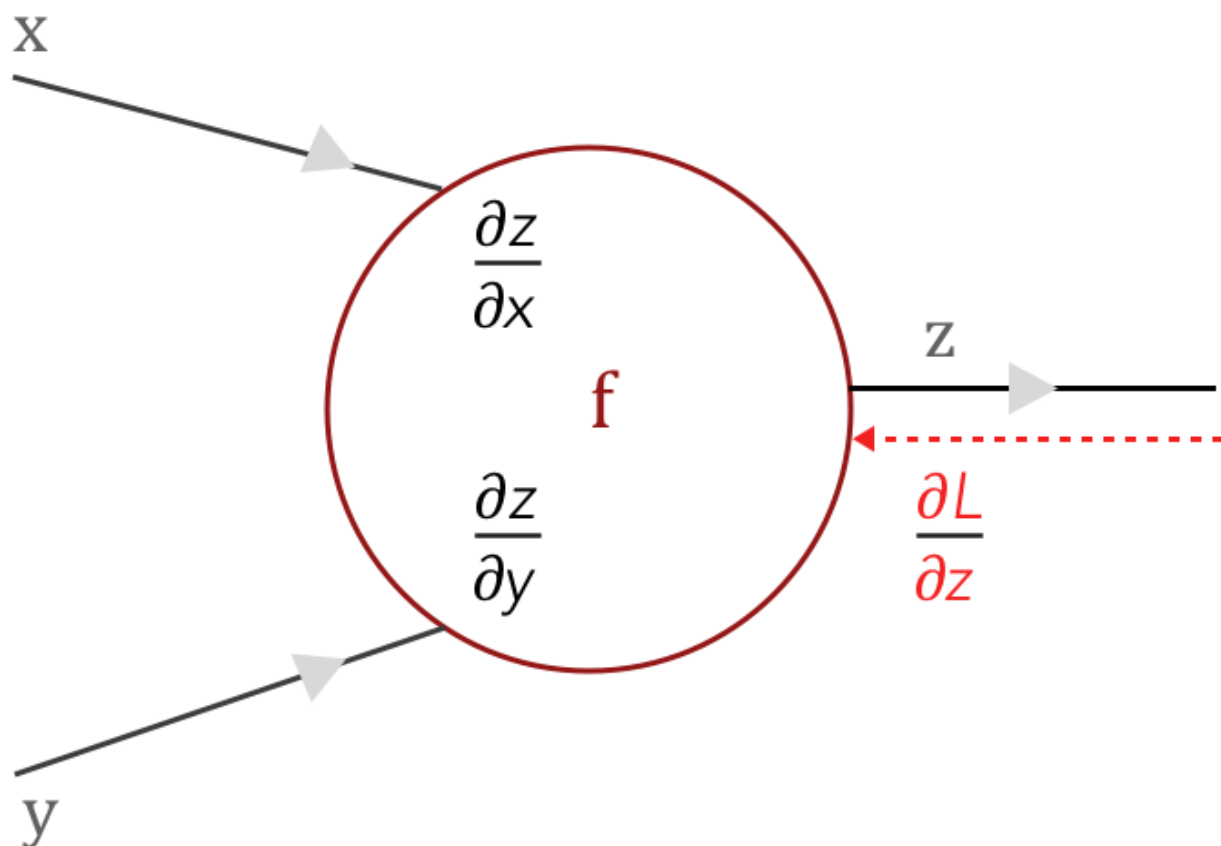
## 2. 在卷积层中的链式法则

如下所示，我们有一个门函数 $f$ ，它的输入是 $x$ 和 $y$ ，输出是 $z$ ：



我们可以很容易的计算出局部梯度 $\partial z / \partial x$  和  $\partial z / \partial y$ 。

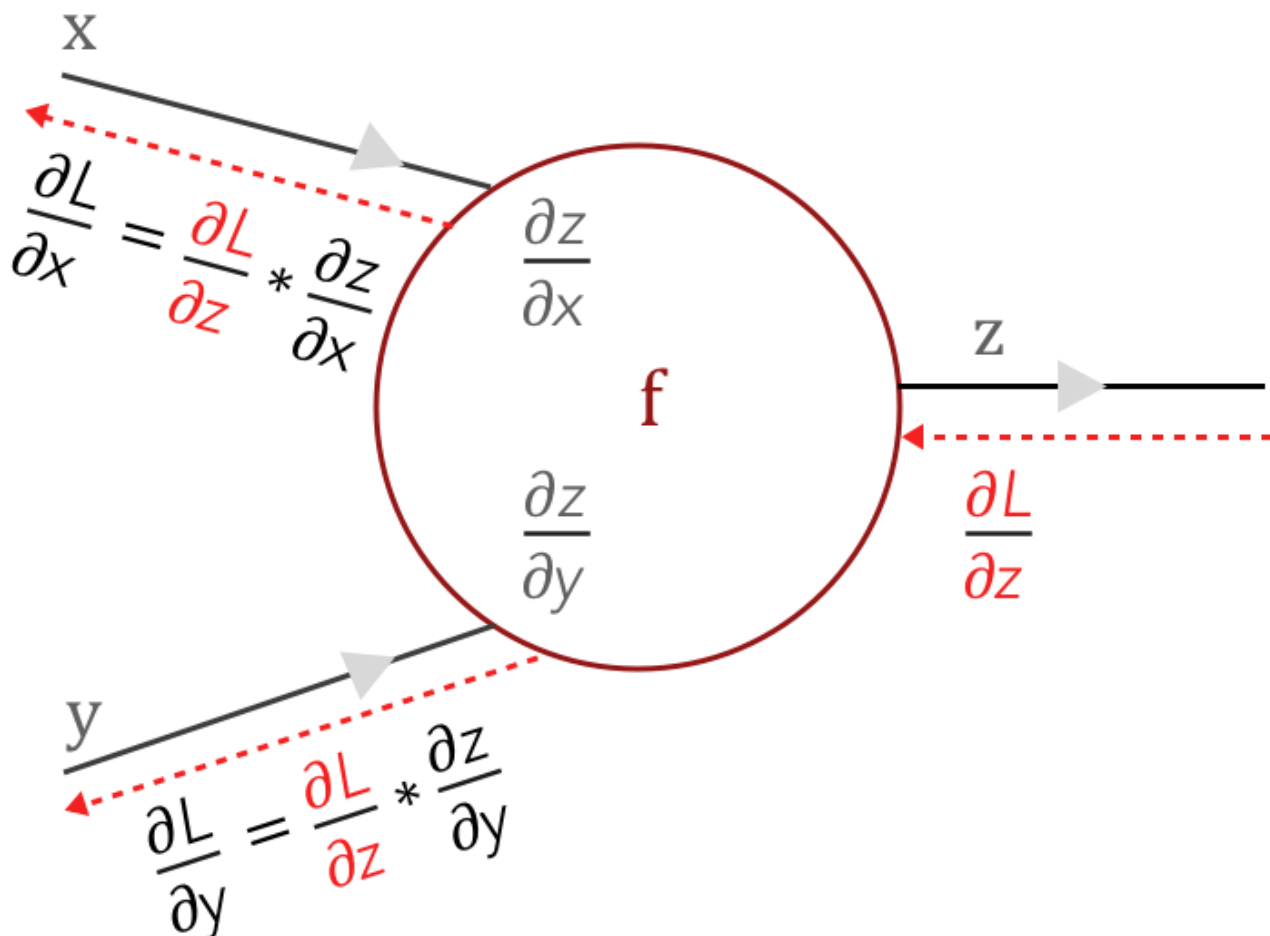
对于前向传播阶段，我们可以通过一个CNN层，然后一直往后传，直到用损失函数计算出它的损失。然后当我们计算损失反向传播时，一层一层的往前传，我们获得了对于 $z$ 的梯度 $\partial L / \partial z$ 。此时为了继续往前传，我们需要计算出 $\partial L / \partial x$ 和 $\partial L / \partial y$ 。



$\frac{\partial z}{\partial x}$  &  $\frac{\partial z}{\partial y}$  are local gradients

$\frac{\partial L}{\partial z}$  is the loss from the previous layer which has to be backpropagated to other layers

此时使用链式法则，我们可以计算出 $\partial L / \partial x$ 和 $\partial L / \partial y$ ：



$\frac{\partial z}{\partial x}$  &  $\frac{\partial z}{\partial y}$  are local gradients

$\frac{\partial L}{\partial z}$  is the loss from the previous layer which has to be backpropagated to other layers

那么，具体的对于CNN中的卷积层是如何进行反向传播的呢？

现在，让我们假设  $f$  是一个卷积函数，对于输入  $X$  和卷积核  $F$  进行卷积计算，其中  $X$  是一个  $3 \times 3$  的矩阵，而  $F$  是一个  $2 \times 2$  的矩阵：



$X_{11}$	$X_{12}$	$X_{13}$
$X_{21}$	$X_{22}$	$X_{23}$
$X_{31}$	$X_{32}$	$X_{33}$

Input **X**

$F_{11}$	$F_{12}$
$F_{21}$	$F_{22}$

Filter **F**

在 $X$ 和 $F$ 的卷积操作的输出为 $O$ ，可以被表示为如下：

$$\begin{array}{|c|c|} \hline O_{11} & O_{12} \\ \hline O_{21} & O_{22} \\ \hline \end{array} = \text{Convolution} \left( \begin{array}{|c|c|c|} \hline X_{11} & X_{12} & X_{13} \\ \hline X_{21} & X_{22} & X_{23} \\ \hline X_{31} & X_{32} & X_{33} \\ \hline \end{array}, \begin{array}{|c|c|} \hline F_{11} & F_{12} \\ \hline F_{21} & F_{22} \\ \hline \end{array} \right)$$

Output **O**                      Input **X**                      Filter **F**