若l层为卷积层:

*为卷积

$$\delta^{l-1} = \delta^l * \mathrm{rot}180\left(w^l\right) \odot \sigma'\left(z^{l-1}\right) \tag{Conv-1}$$

偏置更新需要的梯度:　　　　偏置项的梯度就是sensitivity map所有误差项之和

$$\frac{\partial C}{\partial b^l} = \sum_{u,v} \delta_{w,h} \tag{Conv-2}$$

其中u,v代表卷积核输出的size的长宽

权重更新需要的梯度:

$$\frac{\partial C}{\partial w^l} = a^{l-1} * \delta^l \tag{Conv-3}$$

$$b^l = b^l - \eta \frac{\partial C}{\partial b^l} \tag{Conv-4}$$

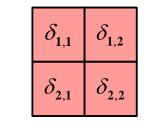$$w^l = w^l - \eta \frac{\partial C}{\partial w^l} \tag{Conv-5}$$

前向传播：

layer $l-1$

$$\begin{array}{|c|c|c|}\hline \delta_{1,1} & \delta_{1,2} & \delta_{1,3} \\ \hline \delta_{2,1} & \delta_{2,2} & \delta_{2,3} \\ \hline \delta_{3,1} & \delta_{3,2} & \delta_{3,3} \\ \hline \end{array}$$

input

3×3

$$\begin{array}{|c|c|}\hline W_{1,1} & W_{1,2} \\ \hline W_{2,1} & W_{2,2} \\ \hline \end{array}$$

W_b

filter

2×2

layer l

$$\begin{array}{|c|c|}\hline \delta_{1,1} & \delta_{1,2} \\ \hline \delta_{2,1} & \delta_{2,2} \\ \hline \end{array}$$

feature map

2×2

$$O = \frac{W - K + 2P}{S} + 1$$

反向传播：

layer l

sensitivity map

2×2

flipped filter

$$\begin{array}{|c|c|}\hline W_{2,2} & W_{2,1} \\ \hline W_{1,2} & W_{1,1} \\ \hline \end{array}$$

2×2

layer $l-1$

$$\begin{array}{|c|c|c|}\hline \delta_{1,1} & \delta_{1,2} & \delta_{1,3} \\ \hline \delta_{2,1} & \delta_{2,2} & \delta_{2,3} \\ \hline \delta_{3,1} & \delta_{3,2} & \delta_{3,3} \\ \hline \end{array}$$

input

3×3

YOLOv3当前层的参数的具体梯度计算过程

设当前层为第*l-1*层,那么计算其敏感度分两步:

1. 在l层的backward()函数的最后部分,会计算l-1层的

$$delta^{l-1} = \delta^l \frac{\partial z^l}{\partial a^{l-1}}$$

(P1)

2. 在l-1层调用backward函数开头部分,再计算:

$$\delta^{l-1} = delta^{l-1} \odot \sigma^{'}\left(z^{l-1}\right)$$

(P2)

参数更新

$$w = w - \alpha \nabla C \qquad \text{(update-1)}$$

引入动量的参数更新: $$v_t = \gamma v_{t-1} \qquad \text{(update-2)}$$

$$w = w - v_t - \alpha \nabla C \qquad \text{(update-3)}$$

$$w = w - \frac{\lambda}{m} w \qquad \text{(update-4)}$$