卷积神经网络（3）

深度学习

概览

1. 卷积核的来源。
2. 卷积神经网络中的参数。
3. 卷积神经网络的反向传播算法。
    1. 池化层的前一层的残差。
    2. 卷积层的前一层的残差。
    3. 卷积核参数的梯度。
4. 了解卷积神经网络中的经典模型。

1. 卷积核的来源

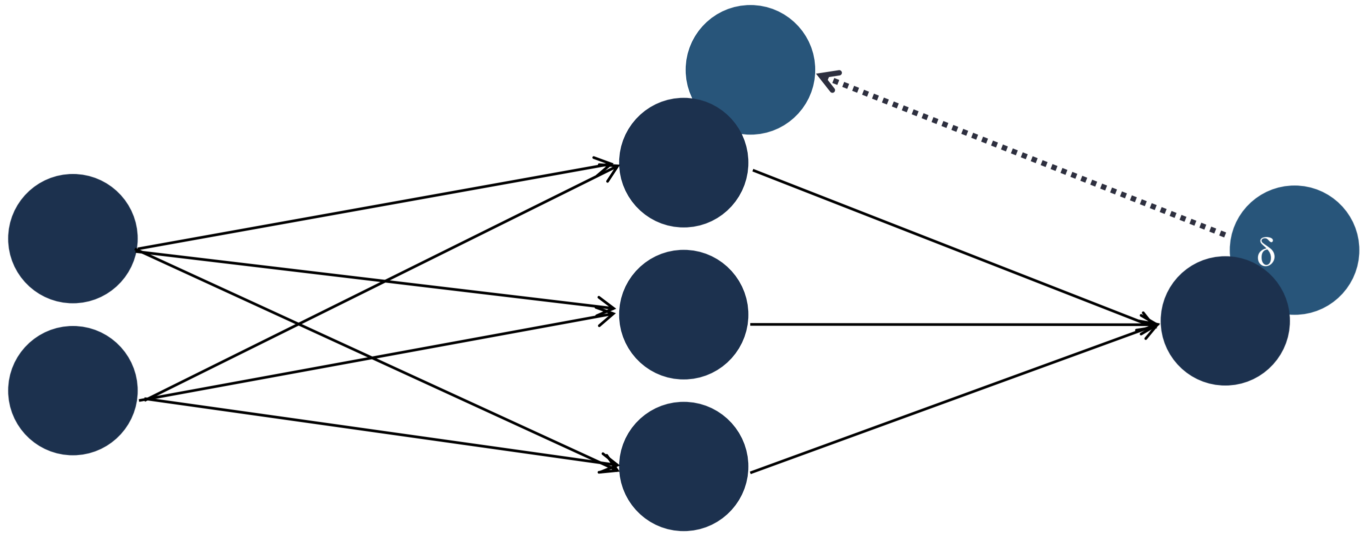
卷积核来源

1. 根据实践经验、统计学规律进行人工设定。

例如Sobel算子， Prewitt算子等。

卷积核来源

2. 通过训练得到：反向传播算法。



卷积核的来源

思考：两种卷积核（或算子）来源的优缺点?

2. 卷积神经网络中的参数

2.1 卷积中的参数

卷积的参数是什么

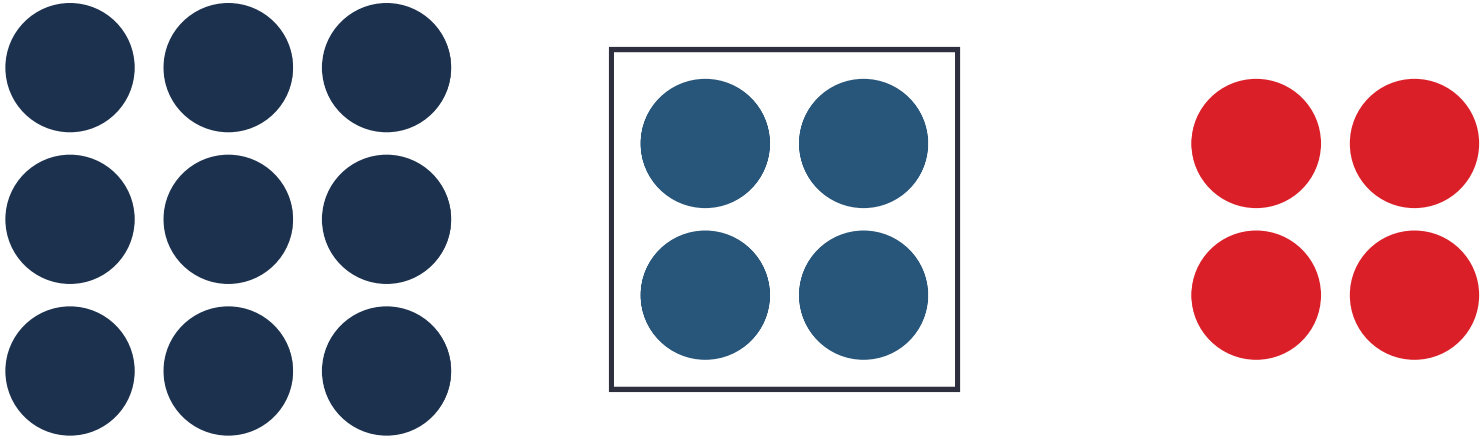
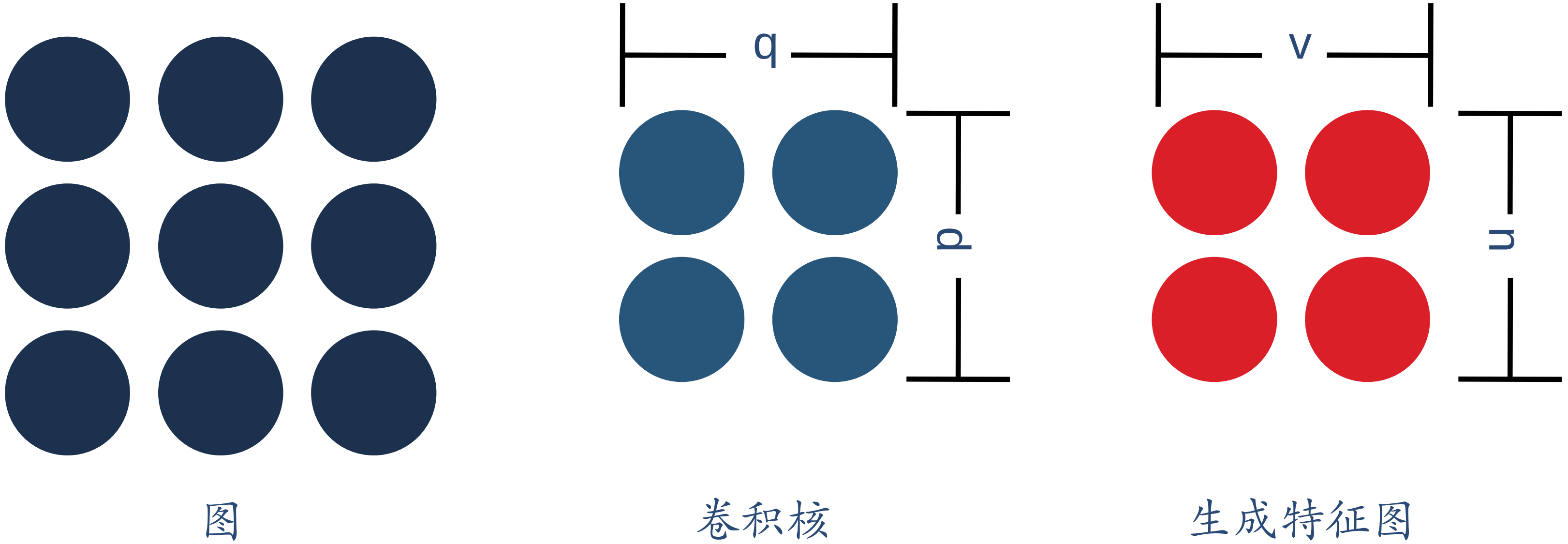


图　　　　卷积核　　　　生成特征图

卷积核是需要求解的参数。

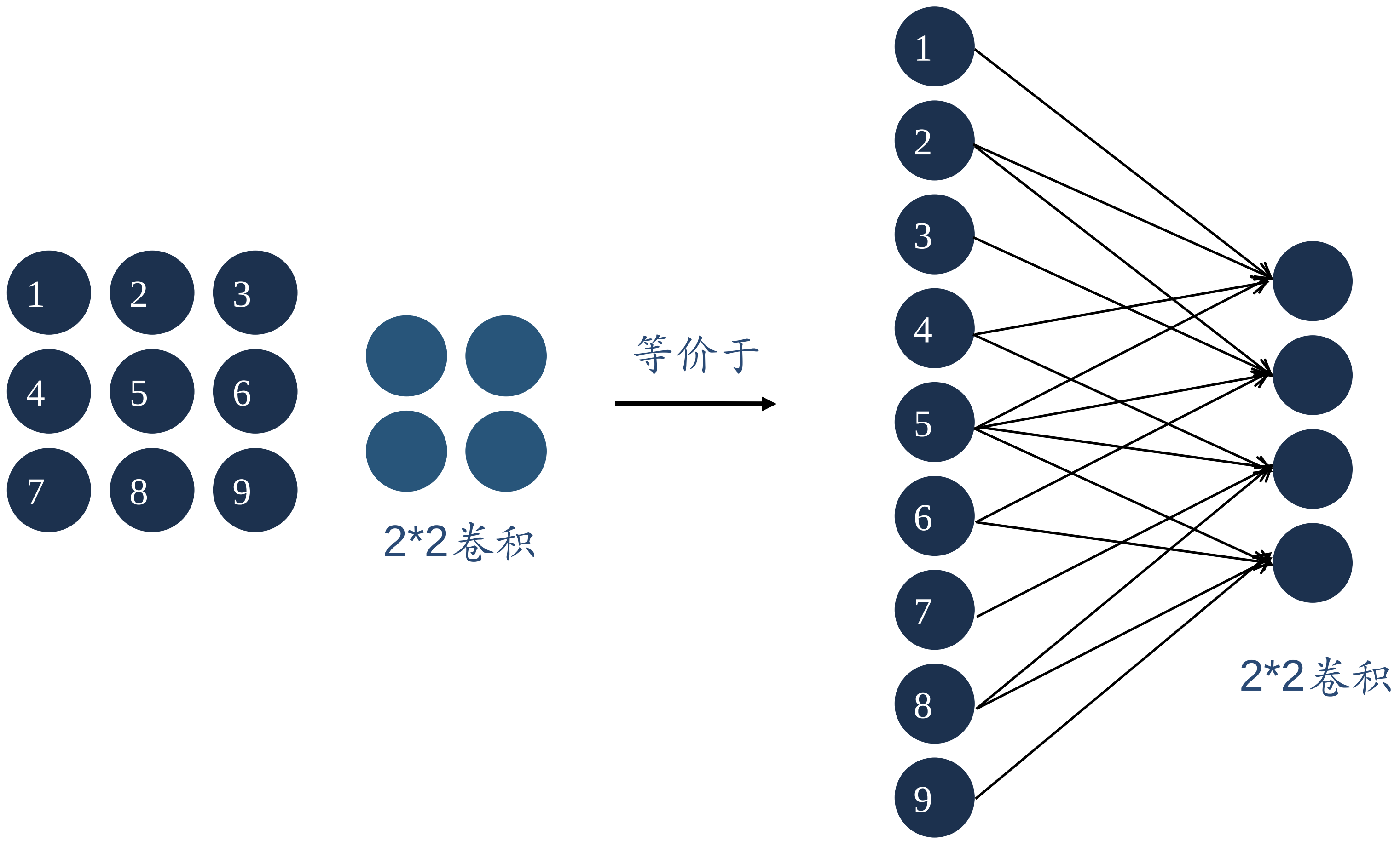
卷积运算



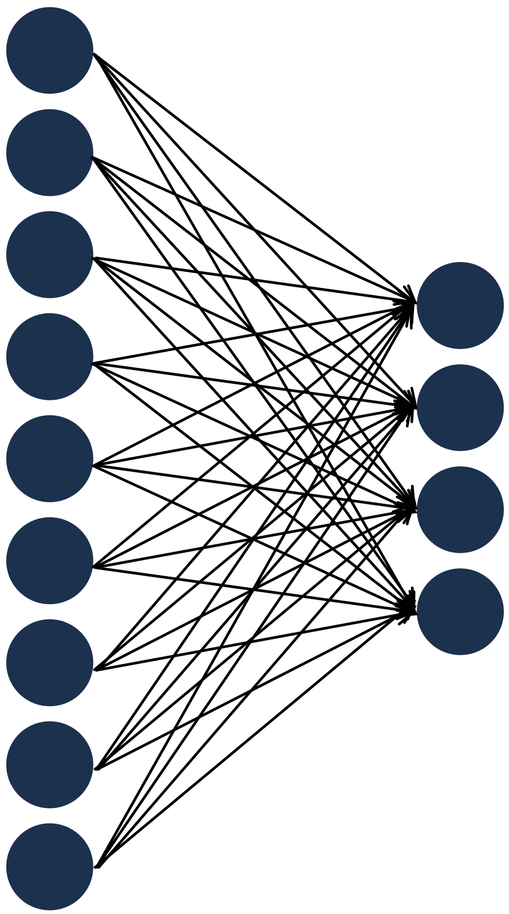
$$z_{u,v}^{(l+1)} = \sum_{p} \sum_{q} w_{p,q}^{(l+1)} a_{u+p-1,v+q-1}^{(l)} + b^{(l+1)} \qquad a_{u,v}^{(l+1)} = \sigma(z_{u,v}^{(l+1)})$$

2.2 卷积与全连接的关系

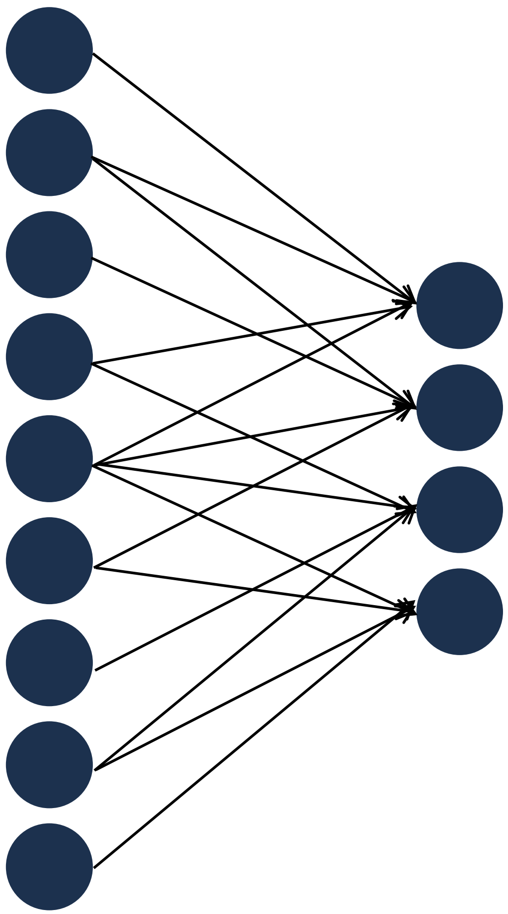
卷积展开



全连接与局部连接对比



全连接



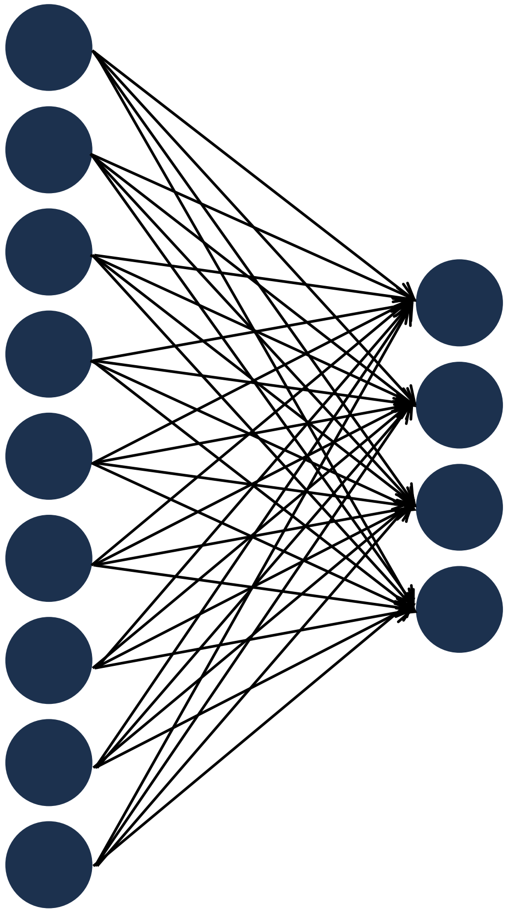
2*2卷积

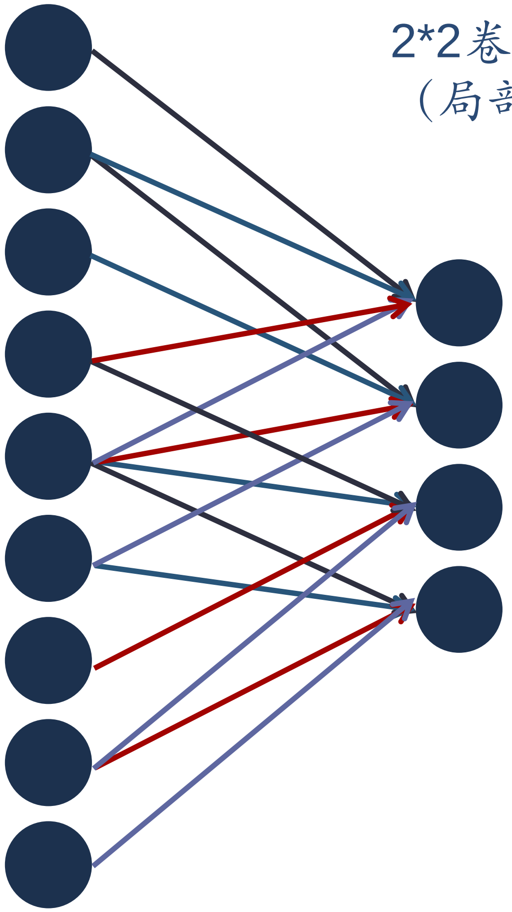
全连接与卷积的参数对比

全连接

连接数量：36

参数数量：36





2*2卷积
（局部连接，权值共享）

连接数量：16

参数数量： 4

3. CNN的反向传播算法

CNN反向传播算法

与全连接神经网络相比

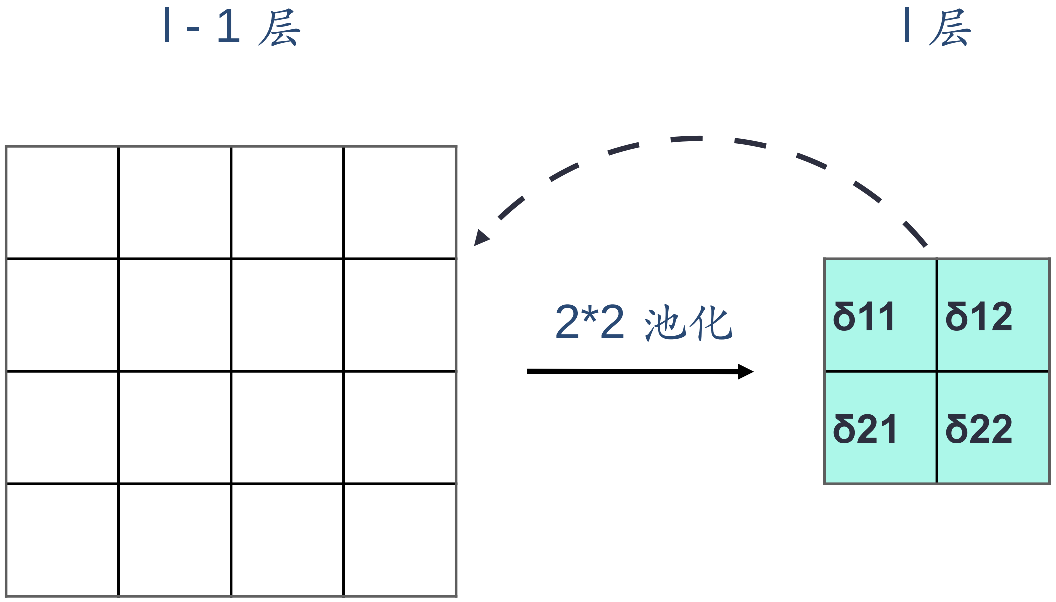
1. 池化层的前一层残差计算。

2. 卷积层的前一层残差计算。

3. 卷积核中的参数的偏导数计算。

3.1 池化层的前一层残差

池化层前向传播



$a_{u,v}^{(l-1)}$ 表示l-1层的输出

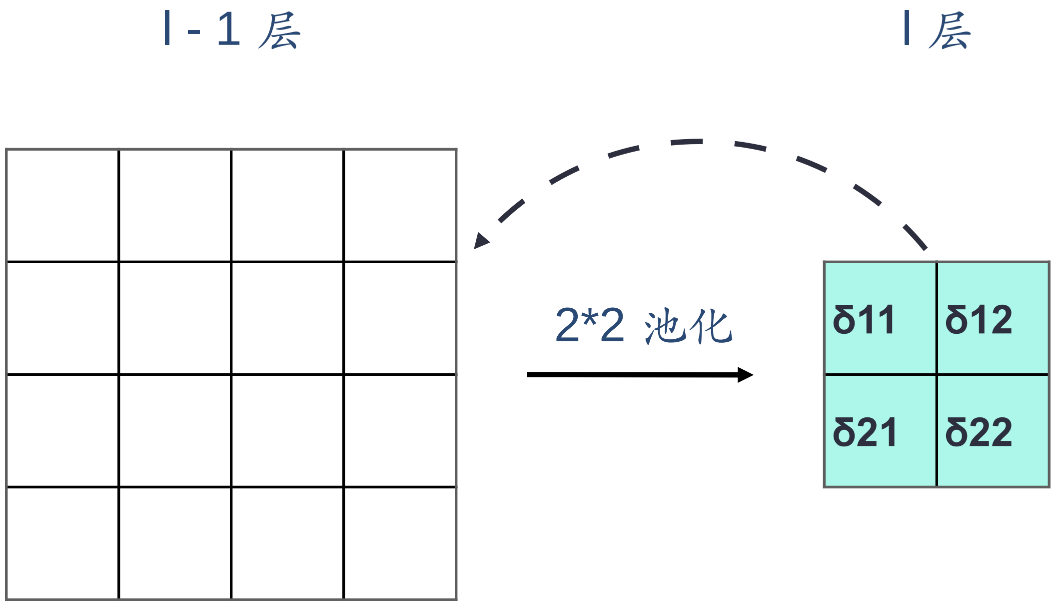
$z_{p,q}^{(l)}$ 表示l层的输入

$a_{p,q}^{(l)}$ 表示l层的输出

$$a_{p,q}^{(l)} = z_{p,q}^{(l)} = pooling(a_{u,v}^{(l-1)})$$

为了方便，此处使用步长等于滑窗大小的池化方法。且不考虑池化层激活函数。

池化层的前一层的残差



一般的，池化层没有参数需要学习，仅需要将残差传递给前一层即可。

最大池化层的前一层残差

最大池化层的前向传播

l - 1 层

1	8	4	3
7	2	1	9
6	1	2	4
8	7	6	0

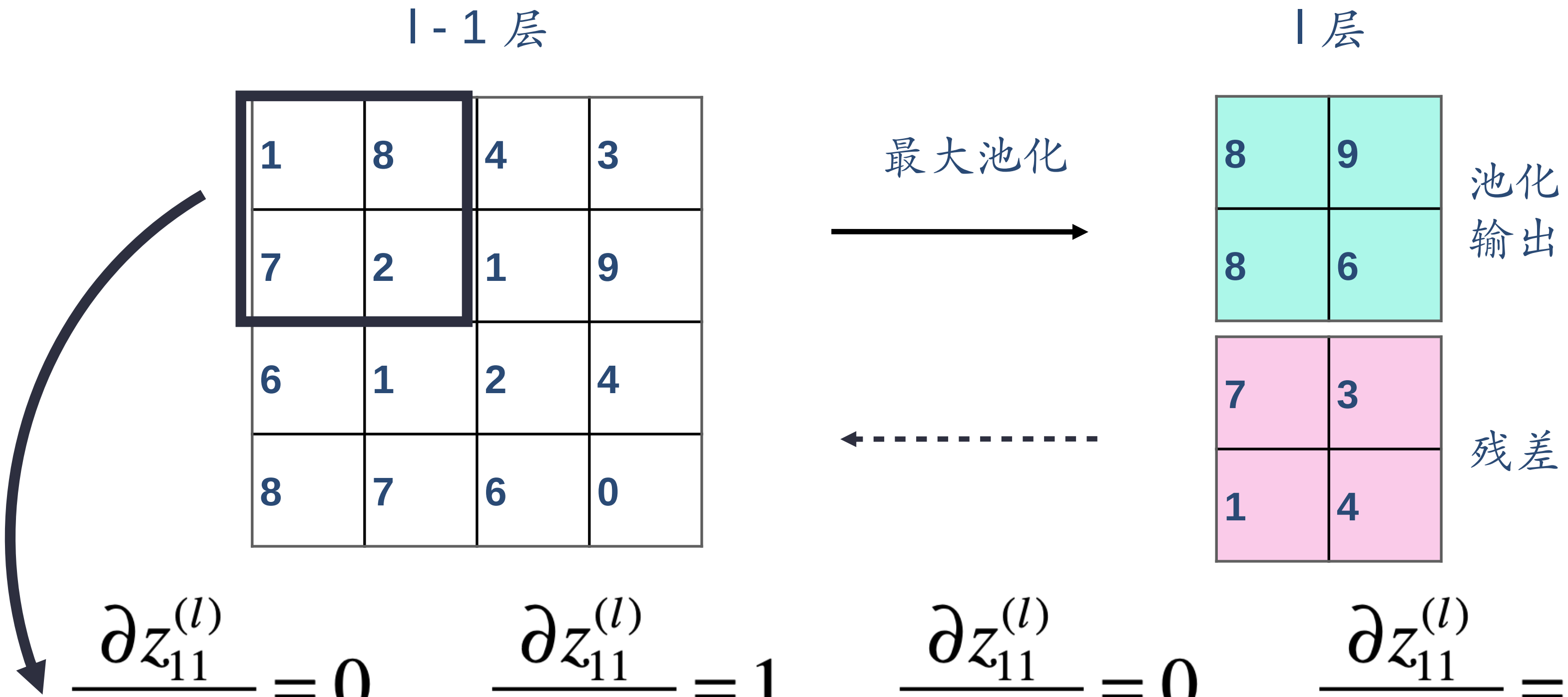
$\xrightarrow{\text{最大池化}}$

l 层

8	9
8	6

$$a_{p,q}^{(l)} = z_{p,q}^{(l)} = \max(a_{(p-1)*m+1,(q-1)*n+1}^{(l-1)}, \cdots, a_{p*m,q*n}^{(l-1)})$$ 此处的m、n指步长。

最大池化层的前一层残差



$$\frac{\partial z_{11}^{(l)}}{\partial a_{11}^{(l-1)}} = 0 \qquad \frac{\partial z_{11}^{(l)}}{\partial a_{12}^{(l-1)}} = 1 \qquad \frac{\partial z_{11}^{(l)}}{\partial a_{12}^{(l-1)}} = 0 \qquad \frac{\partial z_{11}^{(l)}}{\partial a_{12}^{(l-1)}} = 0$$

$$a_{12}^{(l-1)} = z_{11}^{(l)}$$

最大池化层的前一层残差

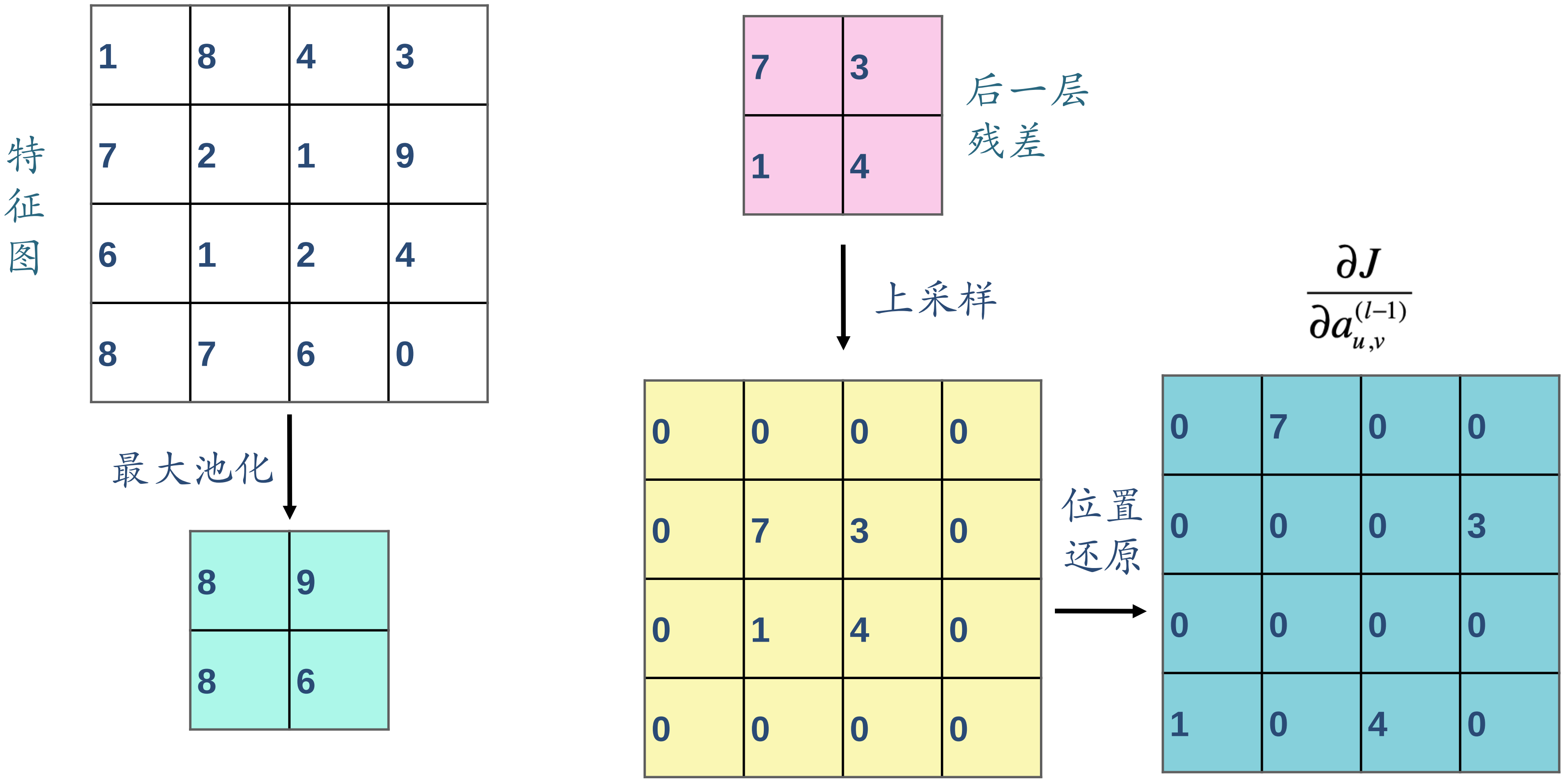
$$\delta_{u,v}^{(l-1)} = \frac{\partial J}{\partial a_{u,v}^{(l-1)}} \frac{\partial a_{u,v}^{(l-1)}}{\partial z_{u,v}^{(l-1)}} = \frac{\partial J}{\partial z_{p,q}^{(l)}} \frac{\partial z_{p,q}^{(l)}}{\partial a_{u,v}^{(l-1)}} \frac{\partial a_{u,v}^{(l-1)}}{\partial z_{u,v}^{(l-1)}} = \delta_{p,q}^{(l)} \frac{\partial z_{p,q}^{(l)}}{\partial a_{u,v}^{(l-1)}} \sigma'(z_{u,v}^{(l-1)})$$

当 $\frac{\partial z_{p,q}^{(l)}}{\partial a_{u,v}^{(l-1)}} = 0$ 则 $\delta_{u,v}^{(l-1)} = 0$

当 $\frac{\partial z_{p,q}^{(l)}}{\partial a_{u,v}^{(l-1)}} = 1$ 则 $\delta_{u,v}^{(l-1)} = \delta_{p,q}^{(l)} \sigma'(z_{u,v}^{(l-1)})$

$$\delta_{u,v}^{(l-1)} = upsample(\delta_{p,q}^{(l)}) \sigma'(z_{u,v}^{(l-1)})$$

最大池化层前向与反向传播



前向传播　　　　反向传播

平均池化层的前一层残差

平均池化层的前向传播

l - 1 层

1	8	4	3
7	2	1	9
6	1	2	4
8	7	6	0

平均池化 →

l 层

4.5	4.25
5.5	3

$$a_{p,q}^{(l)} = z_{p,q}^{(l)} = avg(sum(a_{(p-1)*m+1,(q-1)*n+1}^{(l-1)}, \cdots, a_{p*m,q*n}^{(l-1)}))$$

此处的m、n指步长。

平均池化层的前一层残差

l - 1 层

1	8	4	3
7	2	1	9
6	1	2	4
8	7	6	0

平均池化

l 层

池化输出

4.5	4.25
5.5	3

残差

7	3
1	4

$$\frac{\partial z_{11}^{(l)}}{\partial a_{11}^{(l-1)}} = \frac{1}{4} \qquad \frac{\partial z_{11}^{(l)}}{\partial a_{12}^{(l-1)}} = \frac{1}{4} \qquad \frac{\partial z_{11}^{(l)}}{\partial a_{12}^{(l-1)}} = \frac{1}{4} \qquad \frac{\partial z_{11}^{(l)}}{\partial a_{12}^{(l-1)}} = \frac{1}{4}$$

$$z_{11}^{(l)} = \frac{1}{4}(a_{11}^{(l-1)} + a_{12}^{(l-1)} + a_{21}^{(l-1)} + a_{22}^{(l-1)})$$

平均池化层的前一层残差

$$\delta_{u,v}^{(l-1)} = \frac{\partial J}{\partial a_{u,v}^{(l-1)}} \frac{\partial a_{u,v}^{(l-1)}}{\partial z_{u,v}^{(l-1)}} = \frac{\partial J}{\partial z_{p,q}^{(l)}} \frac{\partial z_{p,q}^{(l)}}{\partial a_{u,v}^{(l-1)}} \frac{\partial a_{u,v}^{(l-1)}}{\partial z_{u,v}^{(l-1)}} = \frac{1}{m*n} \delta_{p,q}^{(l)} \sigma'(z_{u,v}^{(l-1)})$$

其中m、n分别表示滑窗的高、宽。

平均池化层的前向与反向传播

特征图

1	8	4	3
7	2	1	9
6	1	2	4
8	7	6	0

平均池化

4.5	4.25
5.5	3

前向传播

后一层残差

8	4
16	4

2	2	1	1
2	2	1	1
4	4	1	1
4	4	1	1

$$\frac{\partial J}{\partial a_{u,v}^{(l-1)}}$$

前向传播

3.2 卷积层的前一层残差

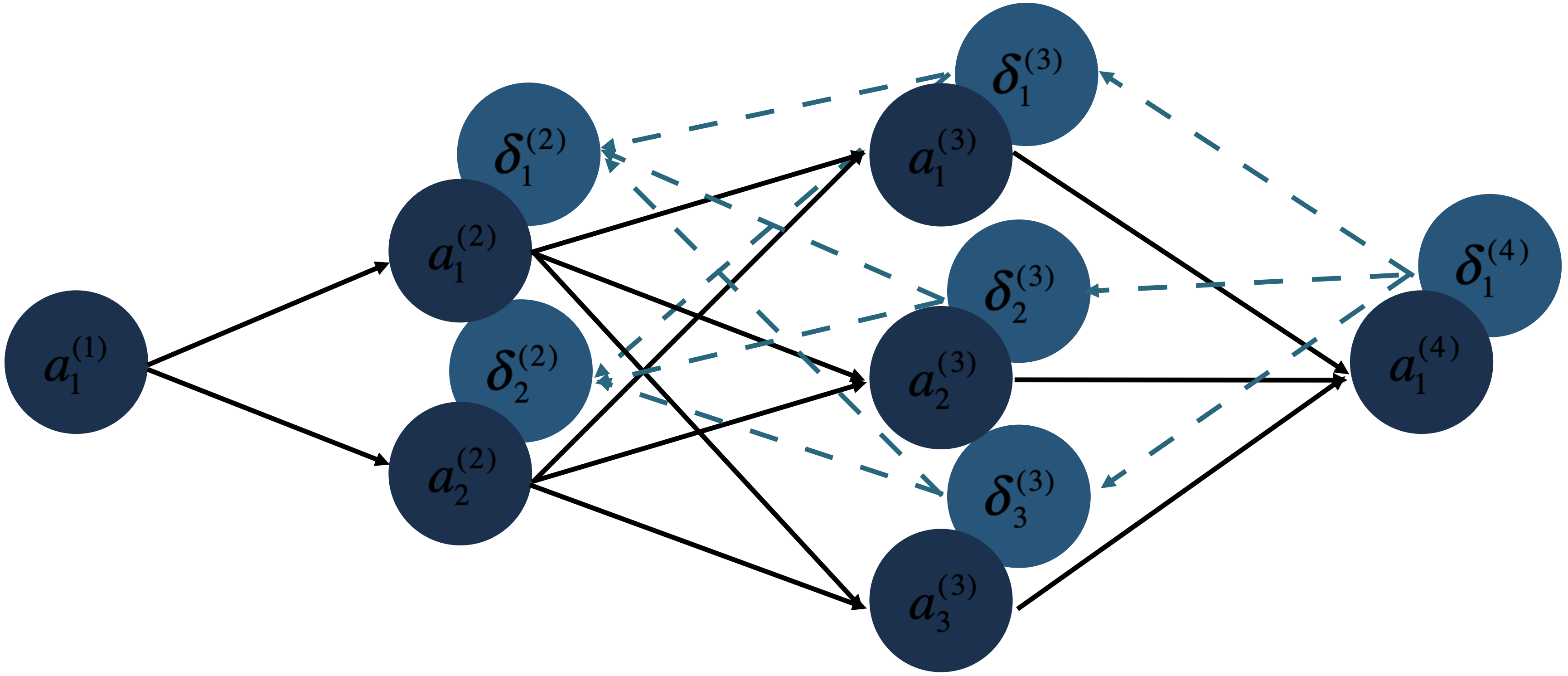
全连接ANN反向传播算法

全连接神经网络隐藏层的残差：

$$\delta_j^{(l)} = (w_j^{(l+1)})^{\mathrm{T}} \cdot \delta^{(l+1)} \cdot \sigma'(z_j^{(l)})$$

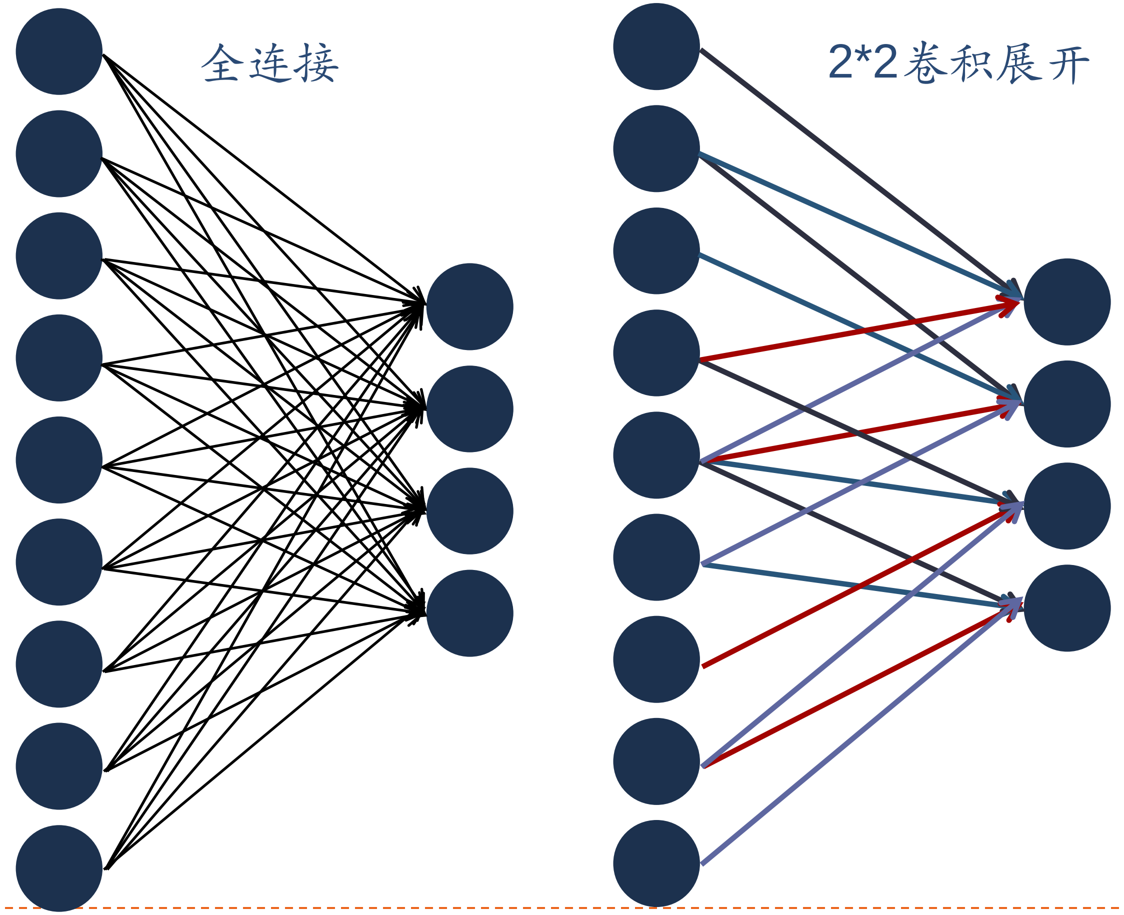
即 $\delta_j^{(l)} \sim (w_j^{(l+1)})^{\mathrm{T}} \cdot \delta^{(l+1)}$

全连接ANN反向传播算法



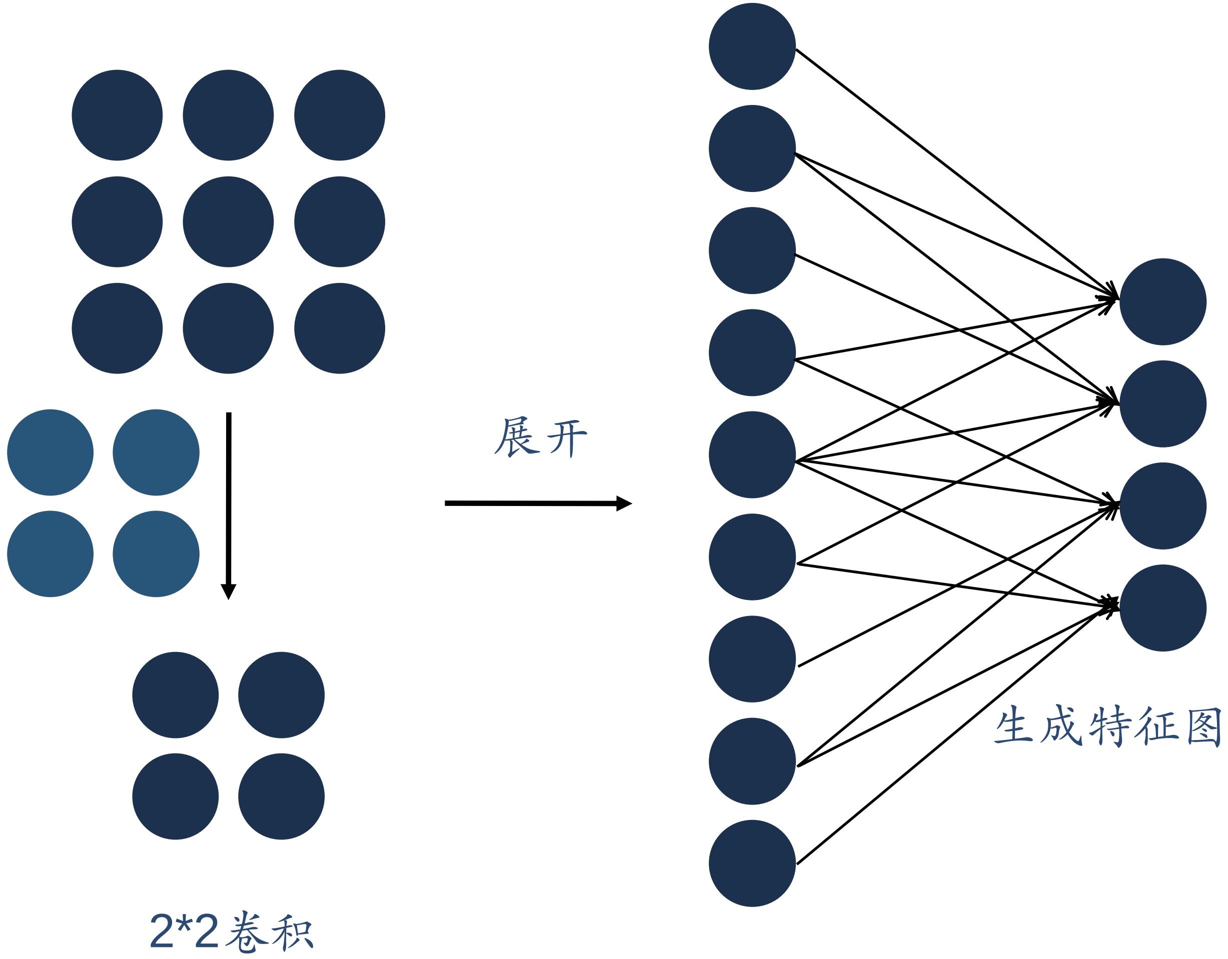
$$\delta_1^{(2)} \sim w_{11}^{(3)} \cdot \delta_1^{(3)} + w_{21}^{(3)} \cdot \delta_2^{(3)} + w_{31}^{(3)} \cdot \delta_3^{(3)}$$

全连接与卷积层的连接关系



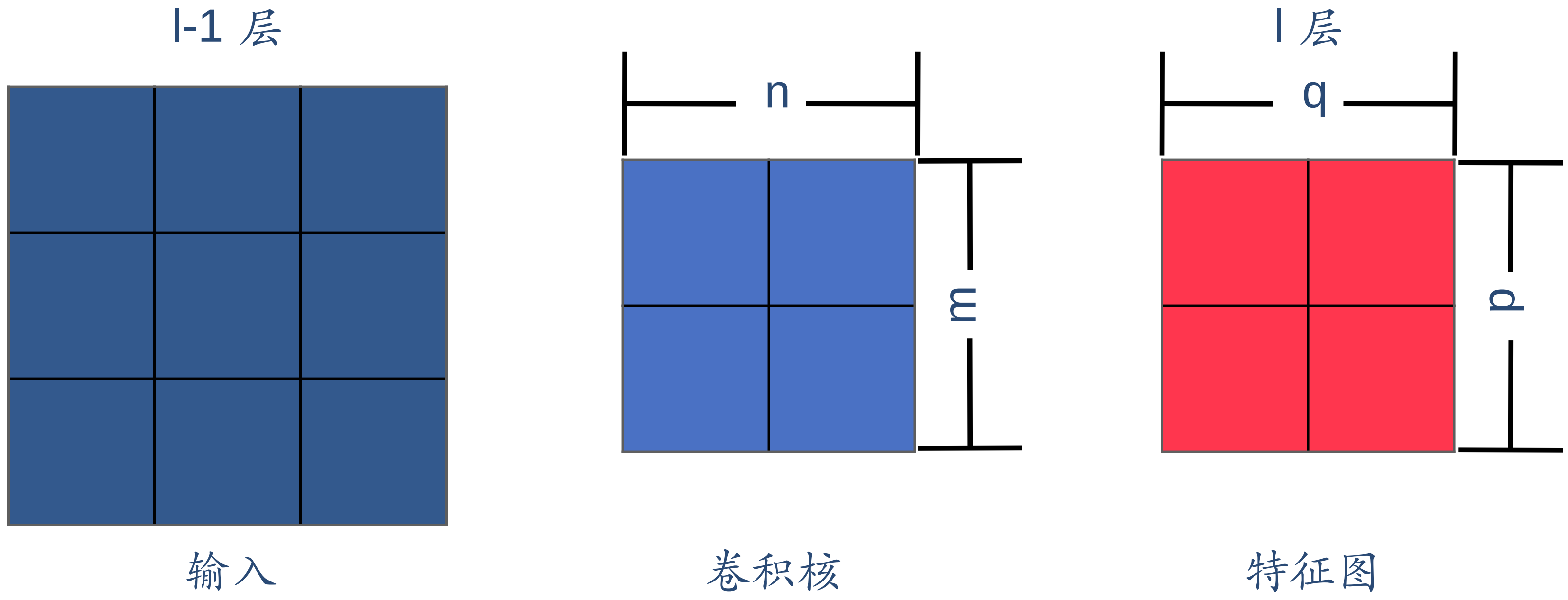
结合全连接层的残差求取公式，猜想卷积前一层的残差？

卷积的前一层的残差



卷积在正向传播时，卷积参数与输入运算，输出的是特征图。反向传播时，卷积参数与卷积层残差运算，输出的是前一层的残差。

卷积的前向传播

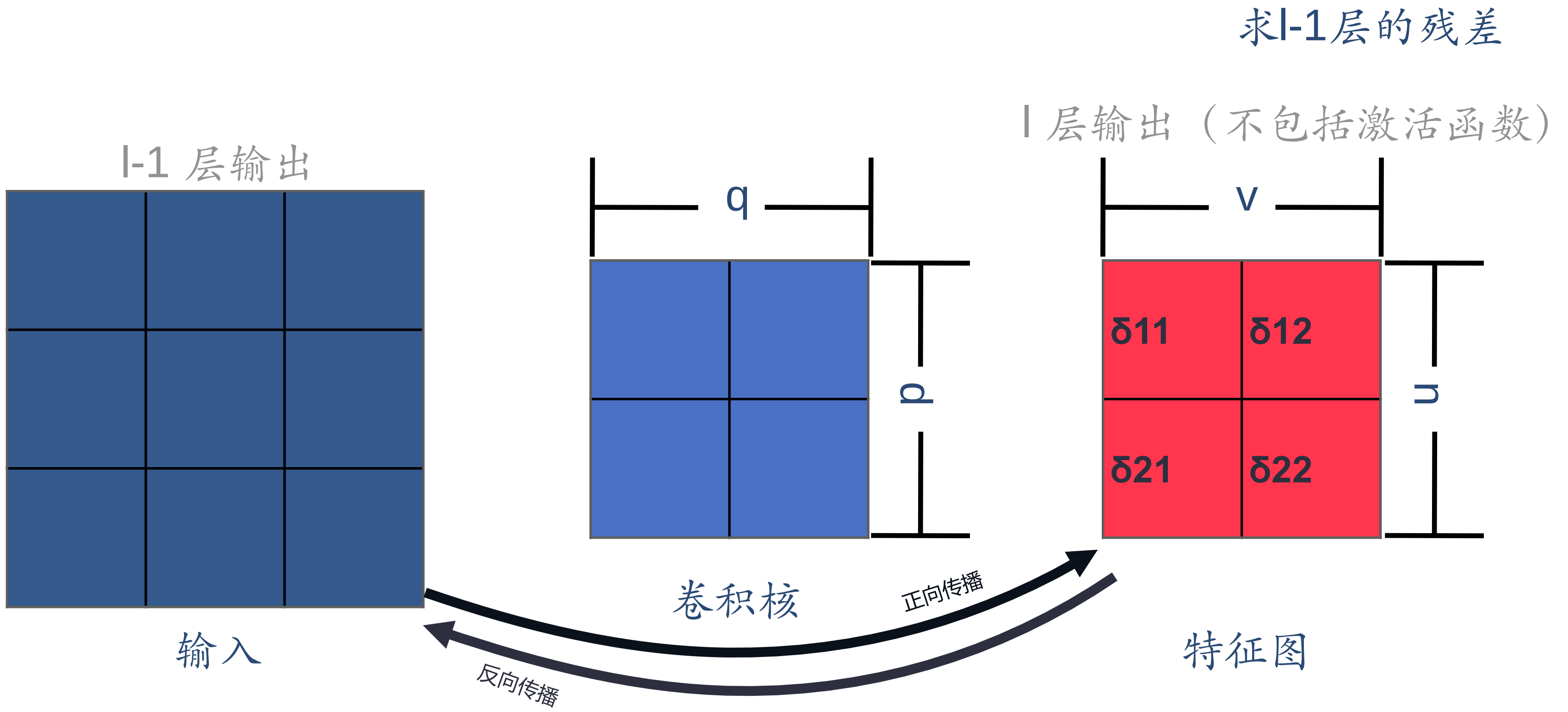


$$z_{p,q}^{(l)} = \sum_{m} \sum_{n} w_{m,n}^{(l)} a_{p+m-1,q+n-1}^{(l-1)} + b^{(l)} \quad a_{p,q}^{(l)} = \sigma(z_{p,q}^{(l)})$$

为了计算方便，默认情况下步长为1，卷积核数量为1。

卷积层前一层残差——实例

实例：前向传播



卷积层输入和

$$z_{11}^{(l)} = a_{11}^{(l-1)}w_{11}^{(l)} + a_{12}^{(l-1)}w_{12}^{(l)} + a_{21}^{(l-1)}w_{21}^{(l)} + a_{22}^{(l-1)}w_{22}^{(l)} + b^{(l)}$$

$$z_{12}^{(l)} = a_{12}^{(l-1)}w_{11}^{(l)} + a_{13}^{(l-1)}w_{12}^{(l)} + a_{22}^{(l-1)}w_{21}^{(l)} + a_{23}^{(l-1)}w_{22}^{(l)} + b^{(l)}$$

$$z_{21}^{(l)} = a_{21}^{(l-1)}w_{11}^{(l)} + a_{22}^{(l-1)}w_{12}^{(l)} + a_{31}^{(l-1)}w_{21}^{(l)} + a_{32}^{(l-1)}w_{22}^{(l)} + b^{(l)}$$

$$z_{22}^{(l)} = a_{22}^{(l-1)}w_{11}^{(l)} + a_{23}^{(l-1)}w_{12}^{(l)} + a_{32}^{(l-1)}w_{21}^{(l)} + a_{33}^{(l-1)}w_{22}^{(l)} + b^{(l)}$$

实例：残差

$$z_{11}^{(l)} = \boxed{a_{11}^{(l-1)} w_{11}^{(l)}} + a_{12}^{(l-1)} w_{12}^{(l)} + a_{21}^{(l-1)} w_{21}^{(l)} + a_{22}^{(l-1)} w_{22}^{(l)} + b^{(l)}$$
$$z_{12}^{(l)} = a_{12}^{(l-1)} w_{11}^{(l)} + a_{13}^{(l-1)} w_{12}^{(l)} + a_{22}^{(l-1)} w_{21}^{(l)} + a_{23}^{(l-1)} w_{22}^{(l)} + b^{(l)}$$
$$z_{21}^{(l)} = a_{21}^{(l-1)} w_{11}^{(l)} + a_{22}^{(l-1)} w_{12}^{(l)} + a_{31}^{(l-1)} w_{21}^{(l)} + a_{32}^{(l-1)} w_{22}^{(l)} + b^{(l)}$$
$$z_{22}^{(l)} = a_{22}^{(l-1)} w_{11}^{(l)} + a_{23}^{(l-1)} w_{12}^{(l)} + a_{32}^{(l-1)} w_{21}^{(l)} + a_{33}^{(l-1)} w_{22}^{(l)} + b^{(l)}$$

$\delta_{11}^{(l-1)}$ 残差仅仅与z11输入有关 $\dfrac{\partial z_{11}^{(l)}}{\partial a_{11}^{(l-1)}} = \dfrac{\partial (a_{11}^{(l-1)} w_{11}^{(l)})}{\partial a_{11}^{(l-1)}} = w_{11}^{(l)}$

$$\delta_{11}^{(l-1)} = \frac{\partial J}{\partial z_{11}^{(l-1)}} = \frac{\partial J}{\partial z_{11}^{(l)}} \frac{\partial z_{11}^{(l)}}{\partial a_{11}^{(l-1)}} \frac{\partial a_{11}^{(l-1)}}{\partial z_{11}^{(l-1)}} = \delta_{11}^{(l)} w_{11}^{(l)} \sigma'(z_{11}^{(l-1)})$$

实例：残差

$$
\begin{aligned}
z_{11}^{(l)} &= a_{11}^{(l-1)}w_{11}^{(l)} + \boxed{a_{12}^{(l-1)}w_{12}^{(l)}} + a_{21}^{(l-1)}w_{21}^{(l)} + a_{22}^{(l-1)}w_{22}^{(l)} + b^{(l)} \\
z_{12}^{(l)} &= \boxed{a_{12}^{(l-1)}w_{11}^{(l)}} + a_{13}^{(l-1)}w_{12}^{(l)} + a_{22}^{(l-1)}w_{21}^{(l)} + a_{23}^{(l-1)}w_{22}^{(l)} + b^{(l)} \\
z_{21}^{(l)} &= a_{21}^{(l-1)}w_{11}^{(l)} + a_{22}^{(l-1)}w_{12}^{(l)} + a_{31}^{(l-1)}w_{21}^{(l)} + a_{32}^{(l-1)}w_{22}^{(l)} + b^{(l)} \\
z_{22}^{(l)} &= a_{22}^{(l-1)}w_{11}^{(l)} + a_{23}^{(l-1)}w_{12}^{(l)} + a_{32}^{(l-1)}w_{21}^{(l)} + a_{33}^{(l-1)}w_{22}^{(l)} + b^{(l)}
\end{aligned}
$$

$\delta_{12}^{(l-1)}$ 残差与Z11、Z12输入有关

$$\frac{\partial z_{11}^{(l)}}{\partial a_{12}^{(l-1)}} = \frac{\partial(a_{12}^{(l-1)}w_{12}^{(l)})}{\partial a_{12}^{(l-1)}} = w_{12}^{(l)}$$

$$\frac{\partial z_{12}^{(l)}}{\partial a_{12}^{(l-1)}} = \frac{\partial(a_{12}^{(l-1)}w_{11}^{(l)})}{\partial a_{12}^{(l-1)}} = w_{11}^{(l)}$$

$$\delta_{12}^{(l-1)} = \frac{\partial J}{\partial z_{12}^{(l-1)}} = \frac{\partial J}{\partial z_{11}^{(l)}}\frac{\partial z_{11}^{(l)}}{\partial a_{12}^{(l-1)}}\frac{\partial a_{12}^{(l-1)}}{\partial z_{12}^{(l-1)}} + \frac{\partial J}{\partial z_{12}^{(l)}}\frac{\partial z_{12}^{(l)}}{\partial a_{12}^{(l-1)}}\frac{\partial a_{12}^{(l-1)}}{\partial z_{12}^{(l-1)}} = \delta_{11}^{(l)}w_{12}^{(l)}\sigma'(z_{12}^{(l-1)}) + \delta_{12}^{(l)}w_{11}^{(l)}\sigma'(z_{12}^{(l-1)})$$

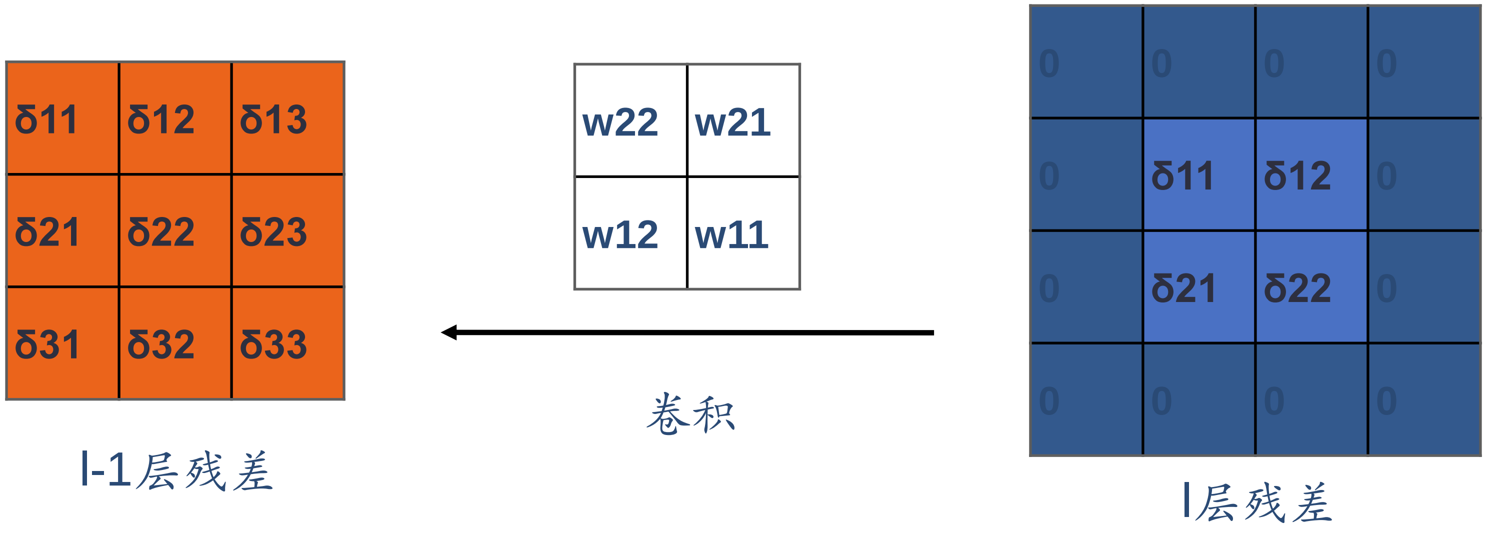
实例：残差规律

$$
\begin{aligned}
z_{11}^{(l)} &= \boxed{a_{11}^{(l-1)} w_{11}^{(l)}} + \boxed{a_{12}^{(l-1)} w_{12}^{(l)}} + \boxed{a_{21}^{(l-1)} w_{21}^{(l)}} + \boxed{a_{22}^{(l-1)} w_{22}^{(l)}} + b^{(l)} \\
z_{12}^{(l)} &= \boxed{a_{12}^{(l-1)} w_{11}^{(l)}} + \boxed{a_{13}^{(l-1)} w_{12}^{(l)}} + \boxed{a_{22}^{(l-1)} w_{21}^{(l)}} + \boxed{a_{23}^{(l-1)} w_{22}^{(l)}} + b^{(l)} \\
z_{21}^{(l)} &= \boxed{a_{21}^{(l-1)} w_{11}^{(l)}} + \boxed{a_{22}^{(l-1)} w_{12}^{(l)}} + \boxed{a_{31}^{(l-1)} w_{21}^{(l)}} + \boxed{a_{32}^{(l-1)} w_{22}^{(l)}} + b^{(l)} \\
z_{22}^{(l)} &= \boxed{a_{22}^{(l-1)} w_{11}^{(l)}} + \boxed{a_{23}^{(l-1)} w_{12}^{(l)}} + \boxed{a_{32}^{(l-1)} w_{21}^{(l)}} + \boxed{a_{33}^{(l-1)} w_{22}^{(l)}} + b^{(l)}
\end{aligned}
$$

一种颜色代表对前一层某个单元求残差时的相关项。

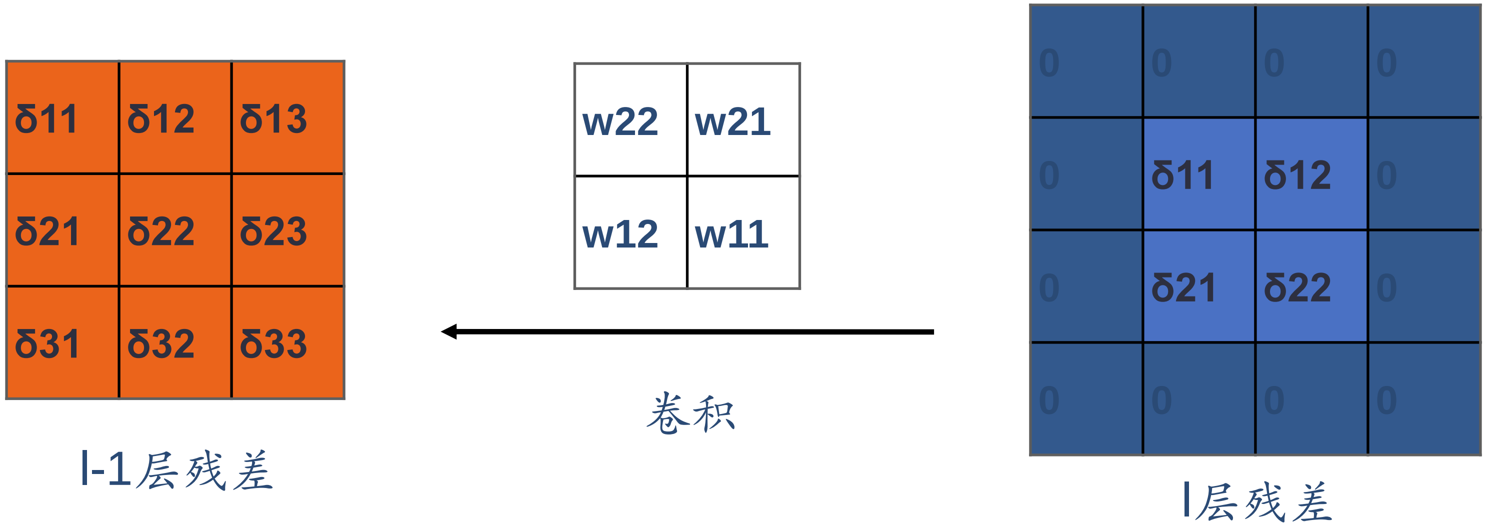
思考：前一层的残差与当前层的残差有什么关系？

前一层残差计算方式



1. 将卷积权重翻转180°；
2. 给当前层残差矩阵边界补0;
3. 使用翻转后的卷积核对pad后的残差做卷积。

前一层残差计算公式

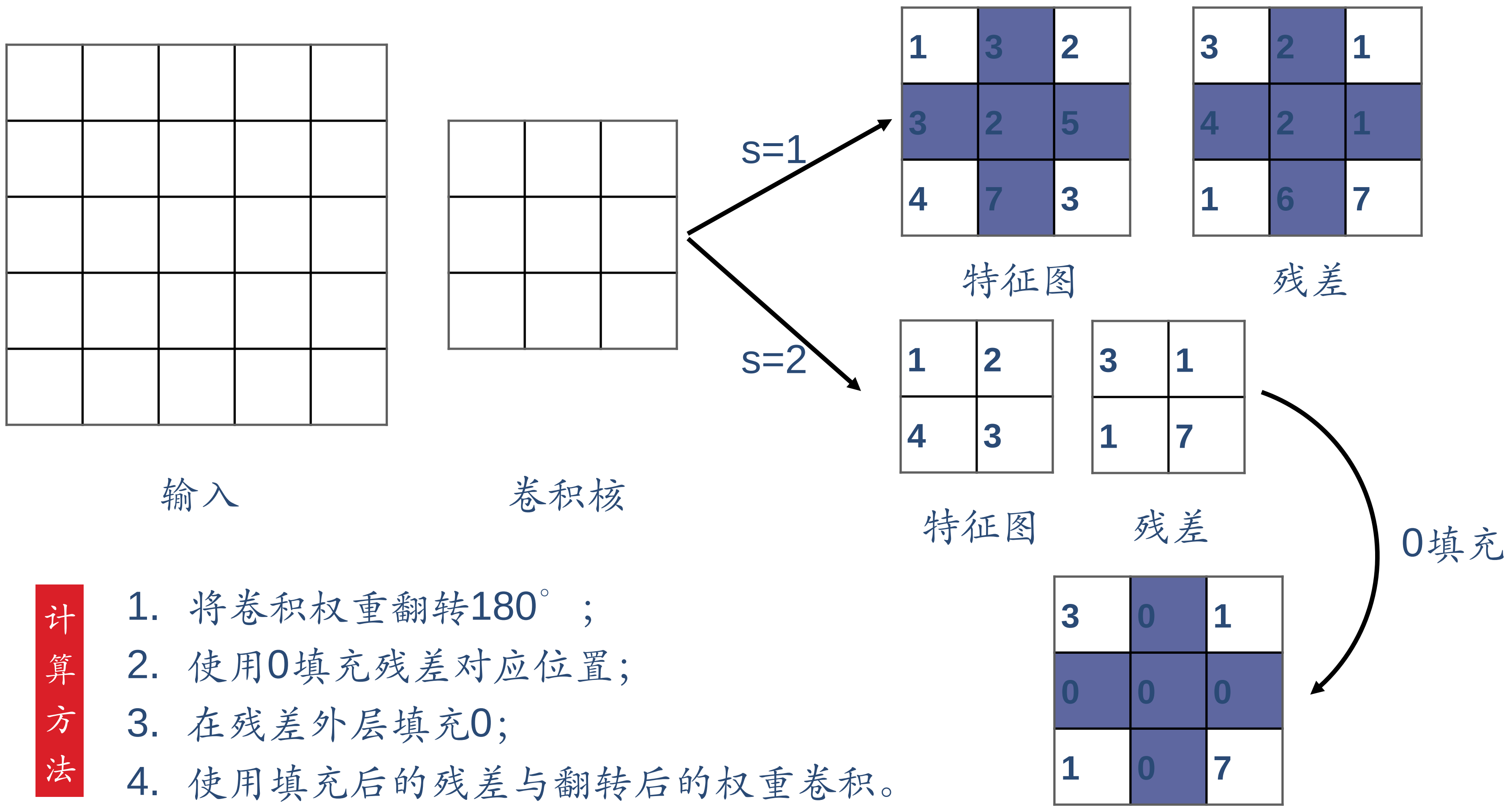


$$\delta_{u,v}^{(l-1)}=\frac{\partial J}{\partial z_{u,v}^{(l-1)}}=\frac{\partial J}{\partial a_{u,v}^{(l-1)}}\frac{\partial a_{u,v}^{(l-1)}}{\partial z_{u,v}^{(l-1)}}=\sum_{m}\sum_{n}w_{m,n}^{(l)}\delta_{m+u-1,n+v-1}^{l}\sigma'(z_{u,v}^{(l-1)})$$

注意：此处u,v表示l-1层的行、列下标；m,n表示翻转后卷积核的行列下标。

其它情况下卷积前一层的残差

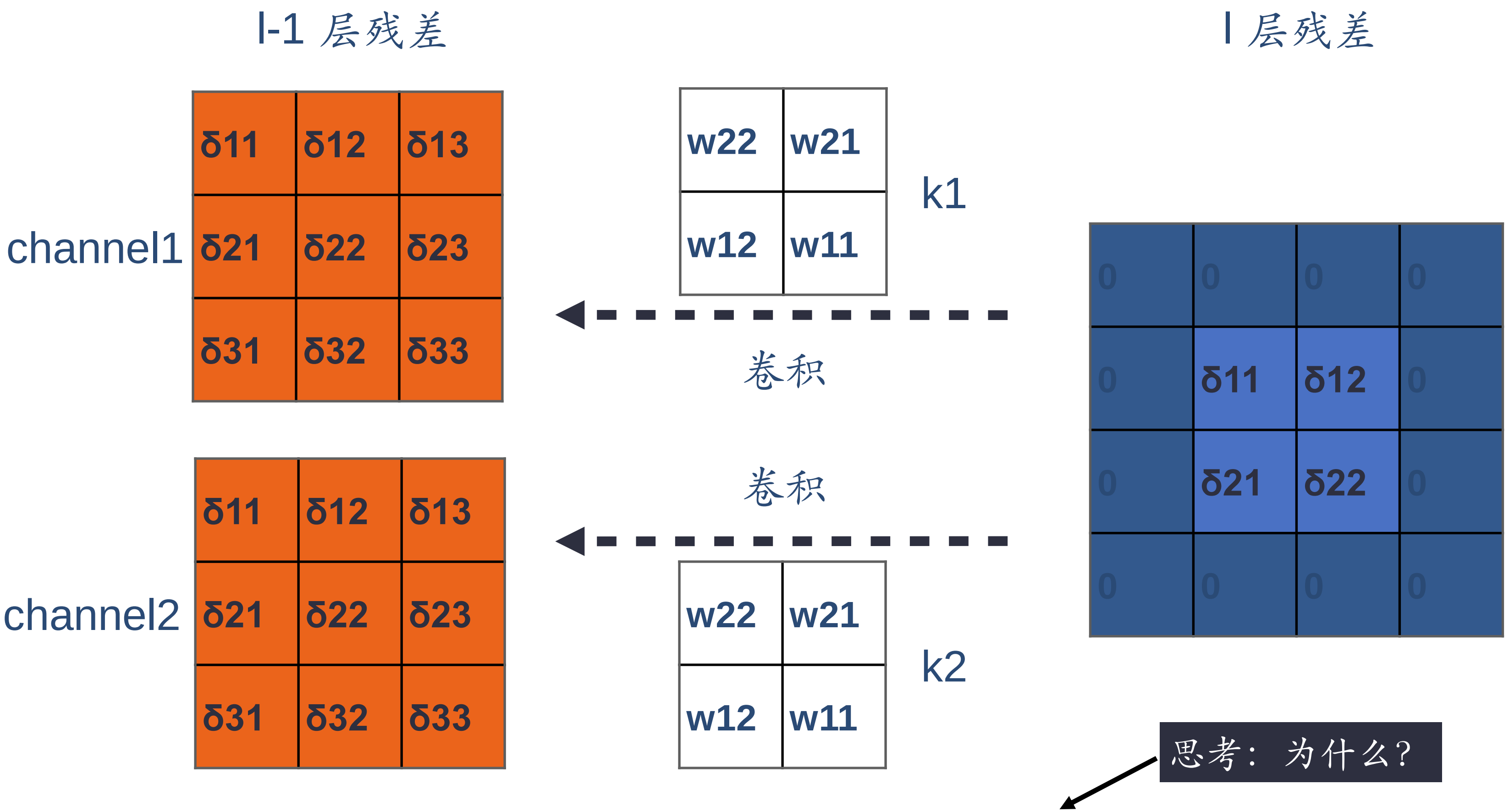
当步长不为1时



计算方法

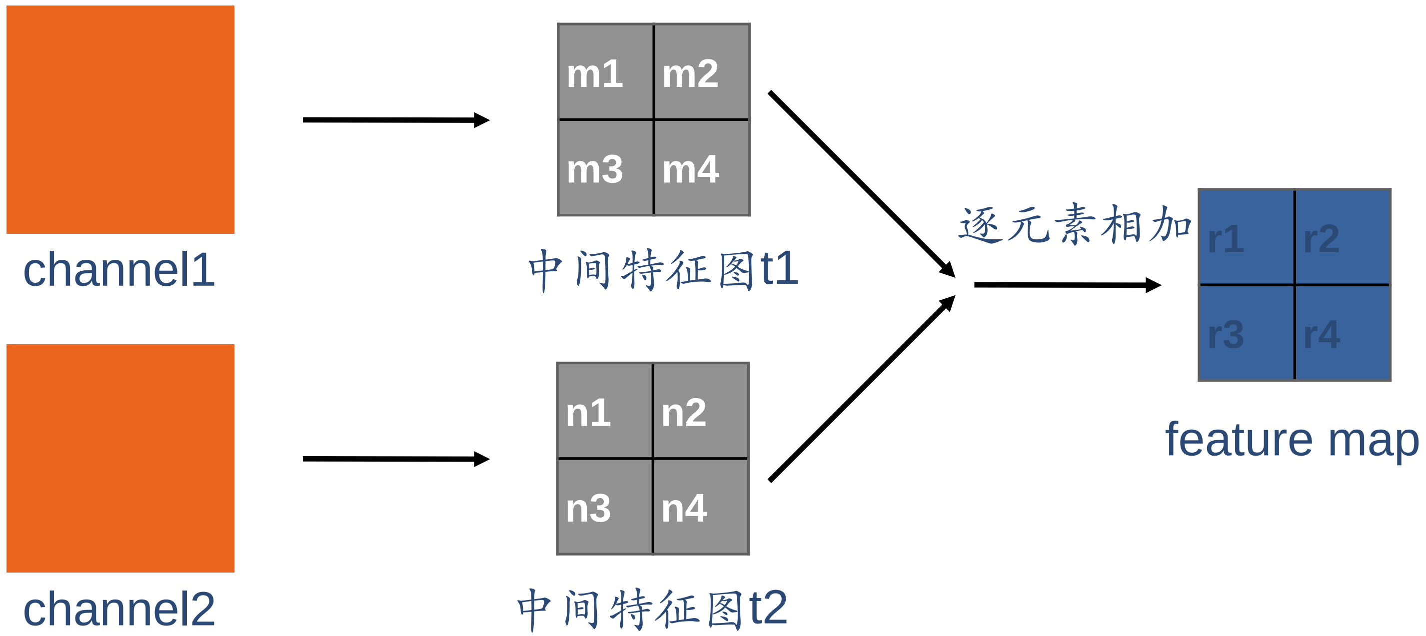
1. 将卷积权重翻转180°；
2. 使用0填充残差对应位置;
3. 在残差外层填充0;
4. 使用填充后的残差与翻转后的权重卷积。

当卷积输入通道不为1时



k1、k2分别表示每个通道的卷积参数，求残差时分别按照单个通道进行计算即可。

当卷积输入通道不为1时

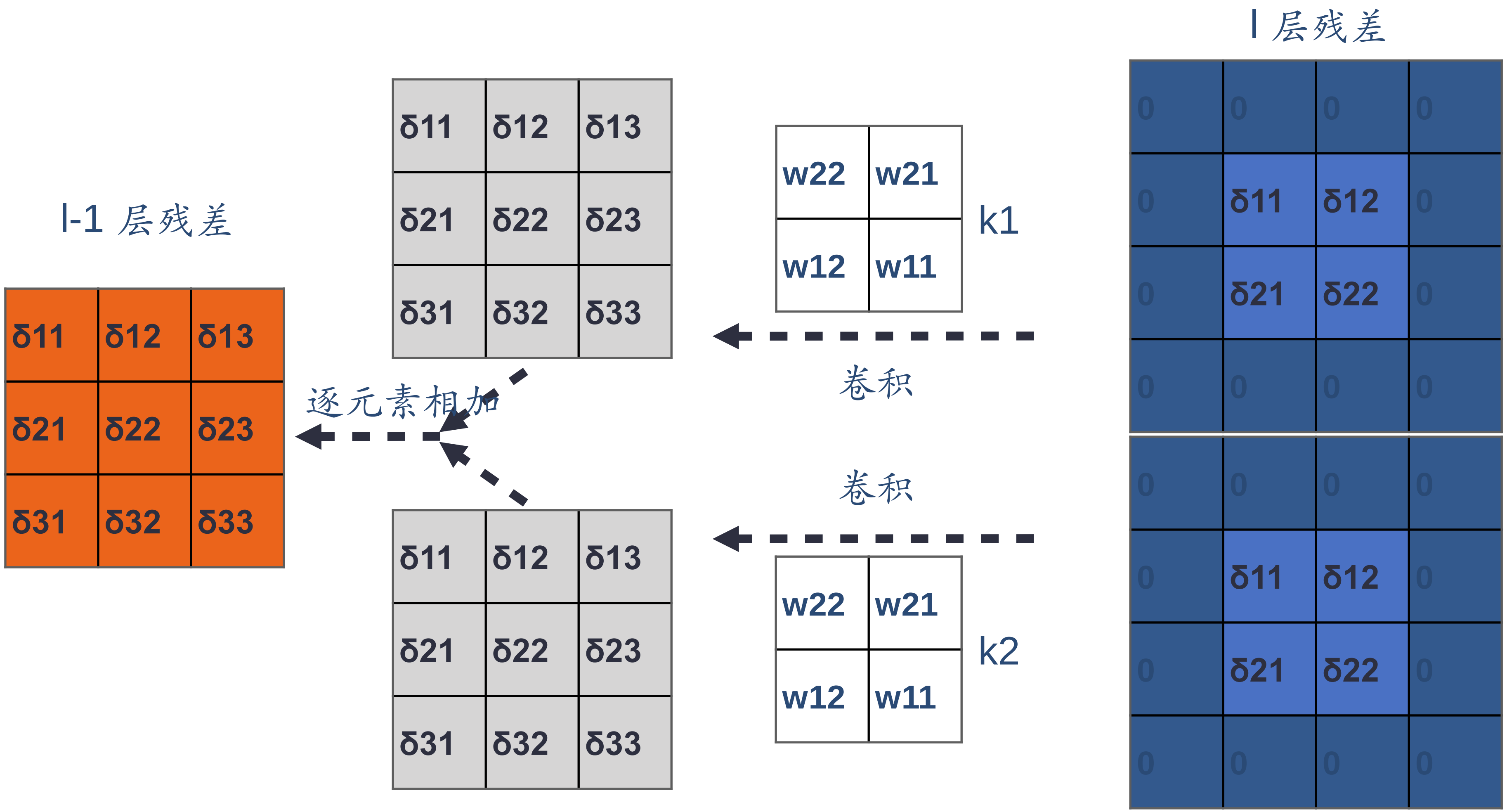


$$\frac{\partial(x_1+x_2)}{\partial x_1}=\frac{\partial(x_1+x_2)}{\partial x_2}=1$$

$$\frac{\partial J}{\partial r}=\frac{\partial J}{\partial(t_1+t_2)}=\frac{\partial J}{\partial r}\frac{\partial r}{\partial x_1}=\frac{\partial J}{\partial r}\frac{\partial r}{\partial x_2}$$

可以看到：中间特征图的残差等于特征图的残差。

当卷积核数量不为1时



k1、k2分别表示两个卷积核。

当输入的边界有0填充时

1. 卷积输入的边界处理不影响反向传播过程。

2. 计算填充0的边界的残差是冗余的，没有意义的。

3.3 卷积核参数的梯度

卷积核参数

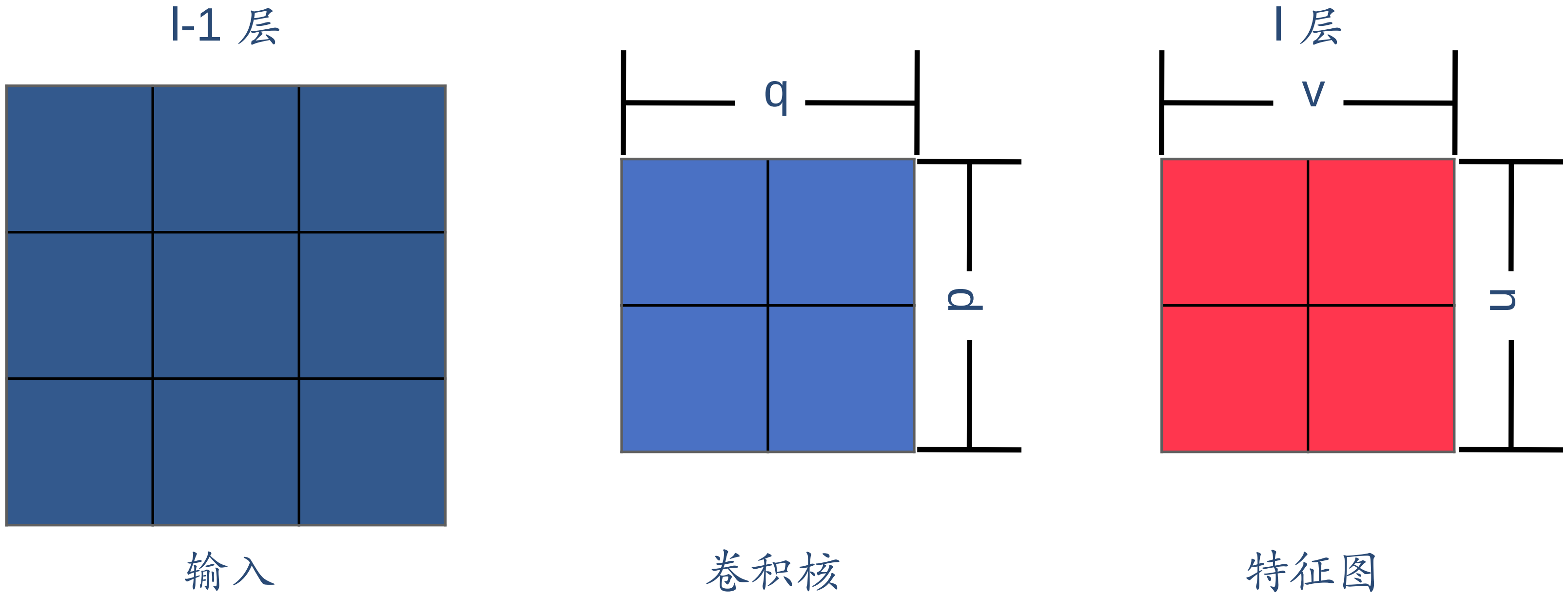
卷积中参数:

1. 卷积中的参数都来自于卷积核。

2. 卷积核的参数参与了当前层输入的计算。

实例

实例



$$z_{11}^{(l)}=a_{11}^{(l-1)}w_{11}^{(l)}+a_{12}^{(l-1)}w_{12}^{(l)}+a_{21}^{(l-1)}w_{21}^{(l)}+a_{22}^{(l-1)}w_{22}^{(l)}+b^{(l)}$$
$$z_{12}^{(l)}=a_{12}^{(l-1)}w_{11}^{(l)}+a_{13}^{(l-1)}w_{12}^{(l)}+a_{22}^{(l-1)}w_{21}^{(l)}+a_{23}^{(l-1)}w_{22}^{(l)}+b^{(l)}$$
$$z_{21}^{(l)}=a_{21}^{(l-1)}w_{11}^{(l)}+a_{22}^{(l-1)}w_{12}^{(l)}+a_{31}^{(l-1)}w_{21}^{(l)}+a_{32}^{(l-1)}w_{22}^{(l)}+b^{(l)}$$
$$z_{22}^{(l)}=a_{22}^{(l-1)}w_{11}^{(l)}+a_{23}^{(l-1)}w_{12}^{(l)}+a_{32}^{(l-1)}w_{21}^{(l)}+a_{33}^{(l-1)}w_{22}^{(l)}+b^{(l)}$$

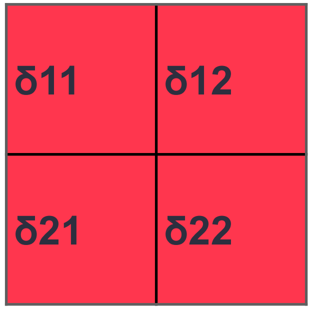
例子—求权重的梯度

$$z_{11}^{(l)} = a_{11}^{(l-1)}w_{11}^{(l)} + a_{12}^{(l-1)}w_{12}^{(l)} + a_{21}^{(l-1)}w_{21}^{(l)} + a_{22}^{(l-1)}w_{22}^{(l)} + b^{(l)}$$

$$z_{12}^{(l)} = a_{12}^{(l-1)}w_{11}^{(l)} + a_{13}^{(l-1)}w_{12}^{(l)} + a_{22}^{(l-1)}w_{21}^{(l)} + a_{23}^{(l-1)}w_{22}^{(l)} + b^{(l)}$$

$$z_{21}^{(l)} = a_{21}^{(l-1)}w_{11}^{(l)} + a_{22}^{(l-1)}w_{12}^{(l)} + a_{31}^{(l-1)}w_{21}^{(l)} + a_{32}^{(l-1)}w_{22}^{(l)} + b^{(l)}$$

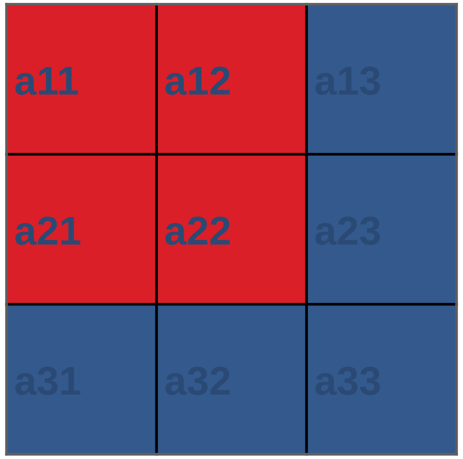
$$z_{22}^{(l)} = a_{22}^{(l-1)}w_{11}^{(l)} + a_{23}^{(l-1)}w_{12}^{(l)} + a_{32}^{(l-1)}w_{21}^{(l)} + a_{33}^{(l-1)}w_{22}^{(l)} + b^{(l)}$$



残差

$$\frac{\partial J}{\partial w_{11}^{(l)}} = \frac{\partial J}{\partial z_{11}^{(l)}}\frac{\partial z_{11}^{(l)}}{\partial w_{11}^{(l)}} + \frac{\partial J}{\partial z_{12}^{(l)}}\frac{\partial z_{12}^{(l)}}{\partial w_{11}^{(l)}} + \frac{\partial J}{\partial z_{21}^{(l)}}\frac{\partial z_{21}^{(l)}}{\partial w_{11}^{(l)}} + \frac{\partial J}{\partial z_{22}^{(l)}}\frac{\partial z_{22}^{(l)}}{\partial w_{11}^{(l)}}$$

$$= \delta_{11}^{(l)}a_{11}^{(l-1)} + \delta_{12}^{(l)}a_{12}^{(l-1)} + \delta_{21}^{(l)}a_{21}^{(l-1)} + \delta_{22}^{(l)}a_{22}^{(l-1)}$$



输入（前一层输出）

例子—求权重的梯度

$$z_{11}^{(l)}=a_{11}^{(l-1)}w_{11}^{(l)}+a_{12}^{(l-1)}w_{12}^{(l)}+a_{21}^{(l-1)}w_{21}^{(l)}+a_{22}^{(l-1)}w_{22}^{(l)}+b^{(l)}$$

$$z_{12}^{(l)}=a_{12}^{(l-1)}w_{11}^{(l)}+a_{13}^{(l-1)}w_{12}^{(l)}+a_{22}^{(l-1)}w_{21}^{(l)}+a_{23}^{(l-1)}w_{22}^{(l)}+b^{(l)}$$

$$z_{21}^{(l)}=a_{21}^{(l-1)}w_{11}^{(l)}+a_{22}^{(l-1)}w_{12}^{(l)}+a_{31}^{(l-1)}w_{21}^{(l)}+a_{32}^{(l-1)}w_{22}^{(l)}+b^{(l)}$$

$$z_{22}^{(l)}=a_{22}^{(l-1)}w_{11}^{(l)}+a_{23}^{(l-1)}w_{12}^{(l)}+a_{32}^{(l-1)}w_{21}^{(l)}+a_{33}^{(l-1)}w_{22}^{(l)}+b^{(l)}$$

δ11	δ12
δ21	δ22

残差

$$\frac{\partial J}{\partial w_{12}^{(l)}}=\frac{\partial J}{\partial z_{11}^{(l)}}\frac{\partial z_{11}^{(l)}}{\partial w_{12}^{(l)}}+\frac{\partial J}{\partial z_{12}^{(l)}}\frac{\partial z_{12}^{(l)}}{\partial w_{12}^{(l)}}+\frac{\partial J}{\partial z_{21}^{(l)}}\frac{\partial z_{21}^{(l)}}{\partial w_{12}^{(l)}}+\frac{\partial J}{\partial z_{22}^{(l)}}\frac{\partial z_{22}^{(l)}}{\partial w_{12}^{(l)}}$$

$$=\delta_{11}^{(l)}a_{12}^{(l-1)}+\delta_{12}^{(l)}a_{13}^{(l-1)}+\delta_{21}^{(l)}a_{22}^{(l-1)}+\delta_{22}^{(l)}a_{23}^{(l-1)}$$

a11	a12	a13
a21	a22	a23
a31	a32	a33

输入（前一层输出）

以此类推可以得到所有J对w的梯度。

例子—求偏置值的梯度

$$z_{11}^{(l)} = a_{11}^{(l-1)} w_{11}^{(l)} + a_{12}^{(l-1)} w_{12}^{(l)} + a_{21}^{(l-1)} w_{21}^{(l)} + a_{22}^{(l-1)} w_{22}^{(l)} + b^{(l)}$$
$$z_{12}^{(l)} = a_{12}^{(l-1)} w_{11}^{(l)} + a_{13}^{(l-1)} w_{12}^{(l)} + a_{22}^{(l-1)} w_{21}^{(l)} + a_{23}^{(l-1)} w_{22}^{(l)} + b^{(l)}$$
$$z_{21}^{(l)} = a_{21}^{(l-1)} w_{11}^{(l)} + a_{22}^{(l-1)} w_{12}^{(l)} + a_{31}^{(l-1)} w_{21}^{(l)} + a_{32}^{(l-1)} w_{22}^{(l)} + b^{(l)}$$
$$z_{22}^{(l)} = a_{22}^{(l-1)} w_{11}^{(l)} + a_{23}^{(l-1)} w_{12}^{(l)} + a_{32}^{(l-1)} w_{21}^{(l)} + a_{33}^{(l-1)} w_{22}^{(l)} + b^{(l)}$$

δ11	δ12
δ21	δ22

残差

$$\frac{\partial J}{\partial b^{(l)}} = \frac{\partial J}{\partial z_{11}^{(l)}} \frac{\partial z_{11}^{(l)}}{\partial b^{(l)}} + \frac{\partial J}{\partial z_{12}^{(l)}} \frac{\partial z_{12}^{(l)}}{\partial b^{(l)}} + \frac{\partial J}{\partial z_{21}^{(l)}} \frac{\partial z_{21}^{(l)}}{\partial b^{(l)}} + \frac{\partial J}{\partial z_{22}^{(l)}} \frac{\partial z_{22}^{(l)}}{\partial b^{(l)}}$$
$$= \delta_{11}^{(l)} + \delta_{12}^{(l)} + \delta_{21}^{(l)} + \delta_{22}^{(l)}$$

卷积核的梯度

权重的梯度： $$\frac{\partial J}{\partial w_{p,q}^{(l)}} = \sum_u \sum_v \delta_{u,v}^{(l)} a_{u+p-1,v+q-1}^{(l-1)}$$

将当前层的残差与前一层的输出做卷积。

偏置值的梯度： $$\frac{\partial J}{\partial b^{(l)}} = \sum_u \sum_v \delta_{u,v}^{(l)}$$

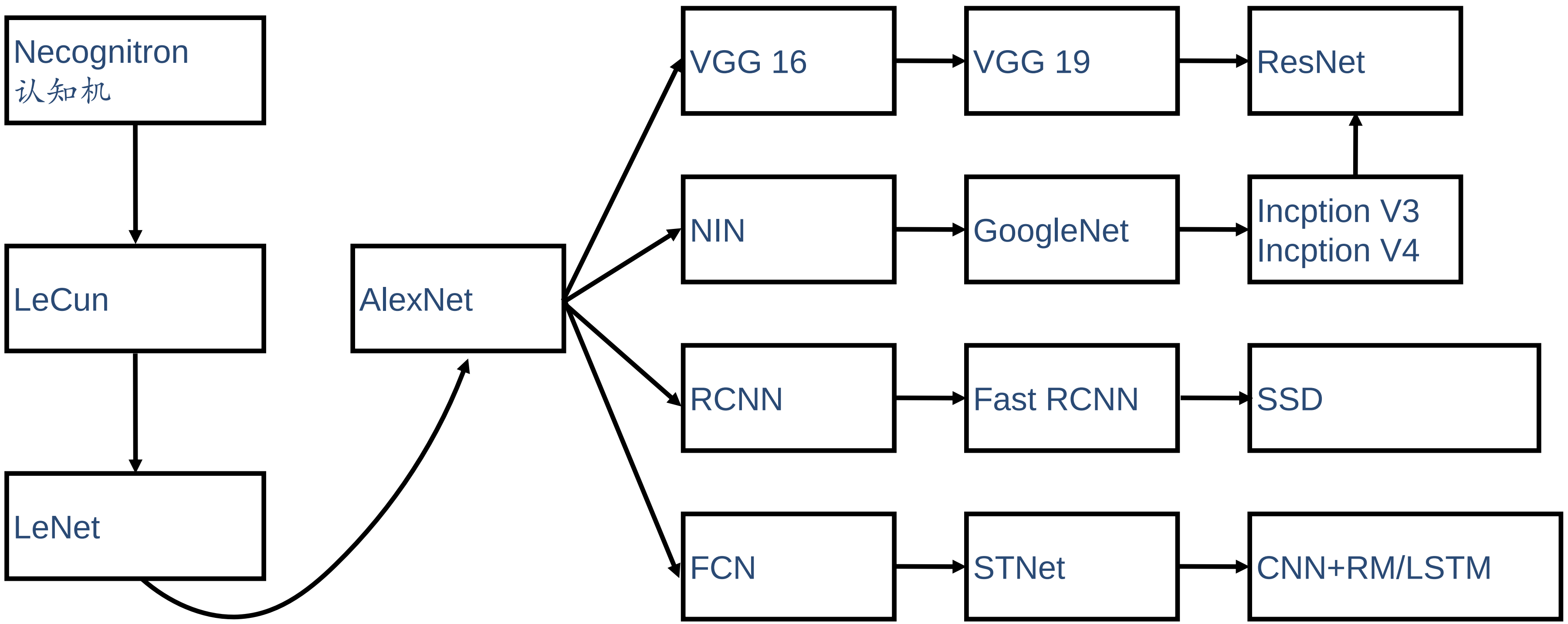
将当前层当前特征图的残差求和。

CNN中的参数

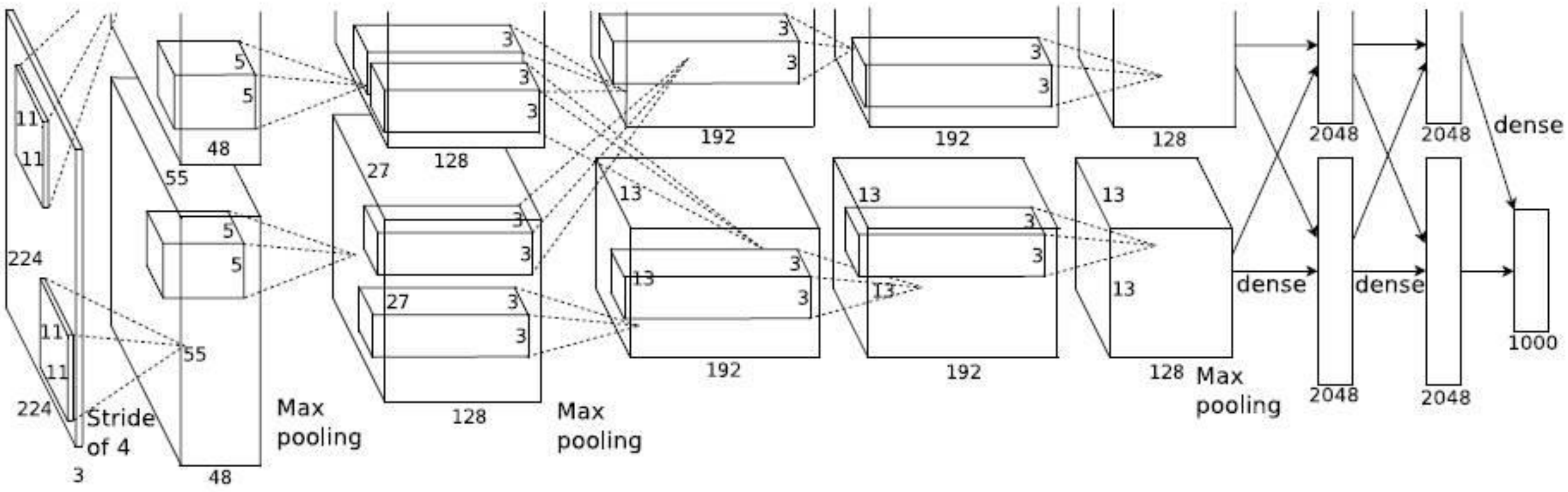
- 卷积层的参数包括卷积核参数与偏置值。偏置值数量等于生成的特征图的激活值数量（同一个特征图中的偏置值相同）。
- 池化层通常没有连接权重、偏置值和激活函数。
- 全连接层的参数。

4. 卷积神经网络中的重要模型

卷积神经网络结构演化史



AlexNet

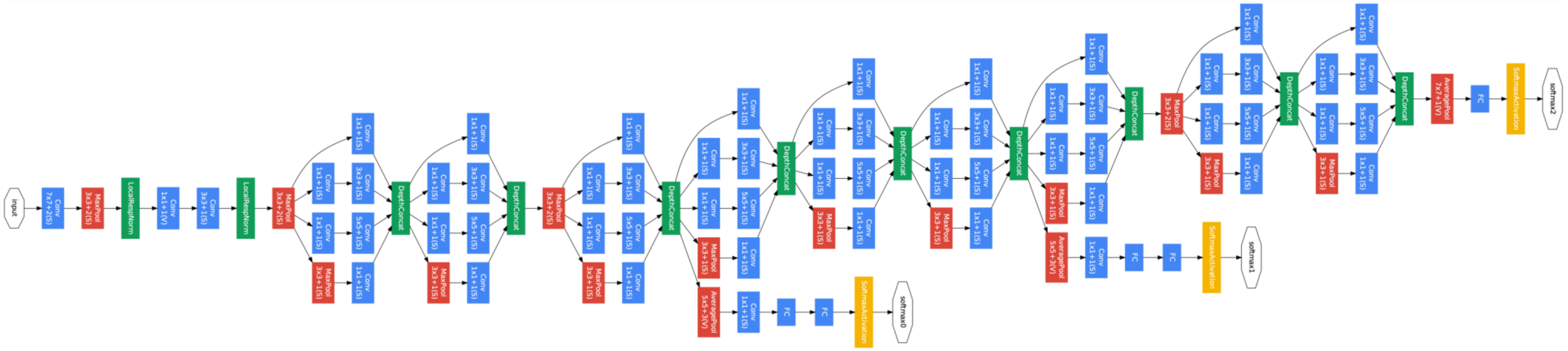


2012，Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton

特点：

使用ReLU激活函数　使用Dropout技术缓解了过拟合　使用了重叠最大池化

GoogLeNet



2014，Christian Szegedy, Wei Liu, Yangqing Jia. et al

特点：使用Inception结构

ResNet

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

2015，Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun

特点：使用Deeper Bottleneck Architectures（DBA）结构

小结

- 局部连接与权值共享降低了连接数量与参数数量。
- 卷积层反向传播残差到前一层的计算方法是：翻转卷积核与补0的当前层残差做卷积。
- 卷积层连接权重的梯度计算方法是：使用当前层的残差与前一层的输出做卷积。
- 卷积层偏置值的梯度计算方法是：将当前某个特征图的残差累加作为此特征图偏置值的残差。
- 卷积神经网络中的重要模型。

THANKS