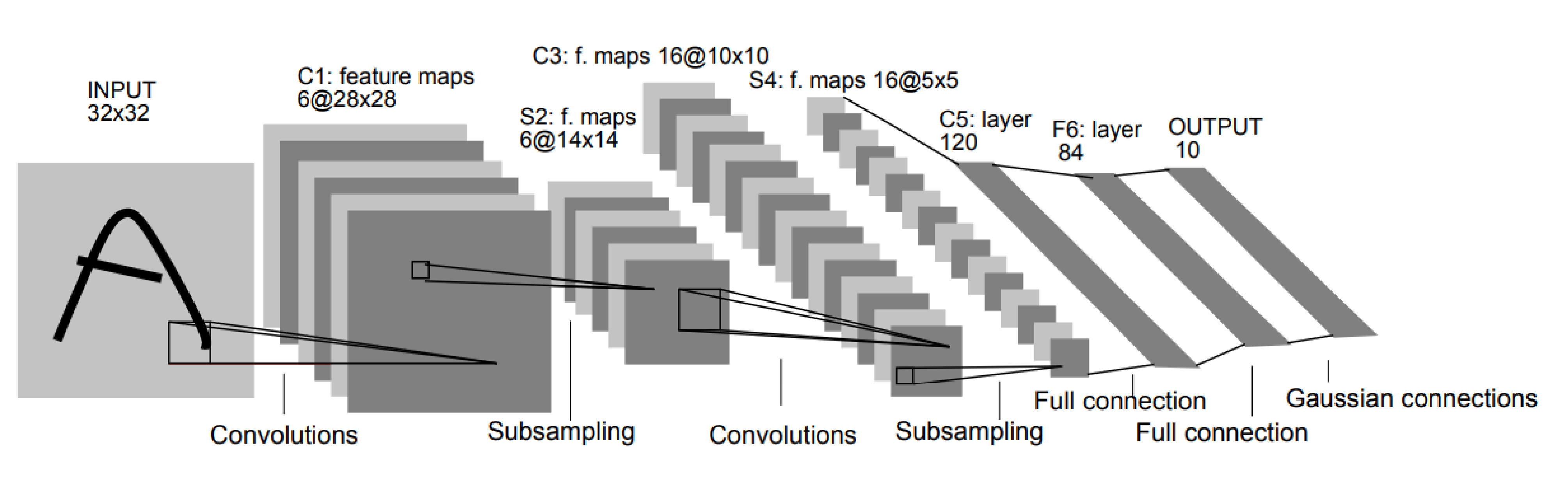
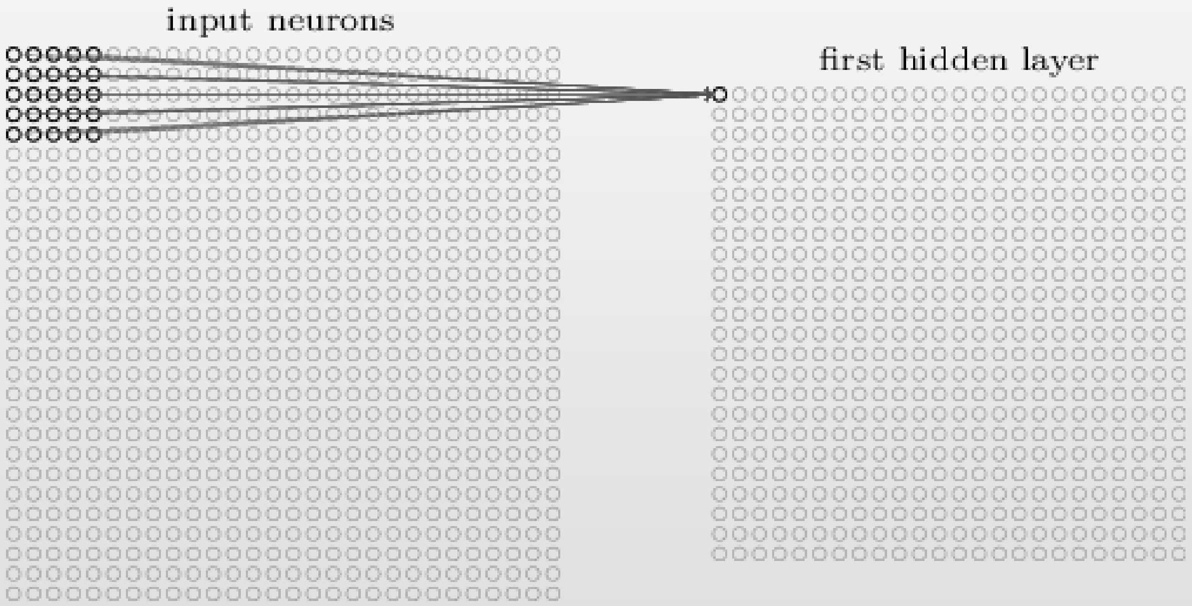
The BP Algorithm on CNN

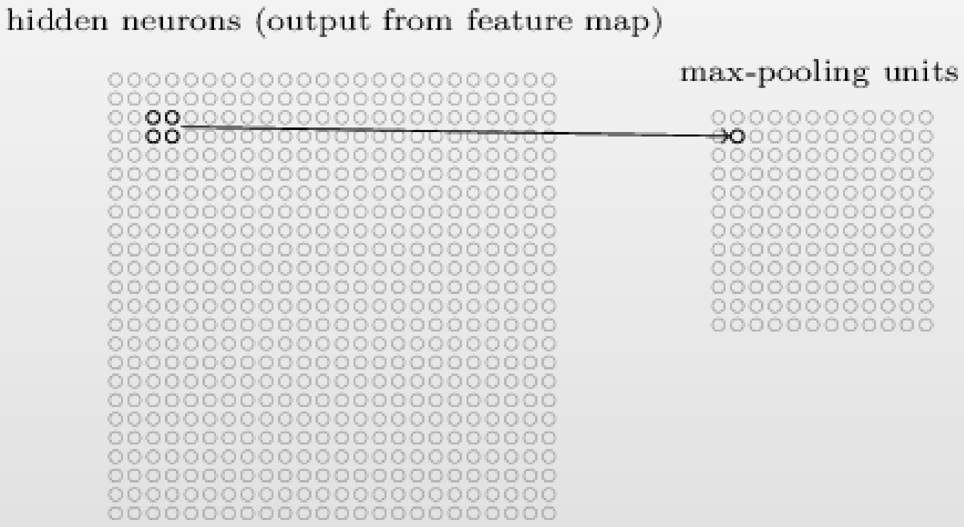


1 . Review the convolutional network.

- -Convolutional layer
- -Pooling layer
- -Fully connected layer

The connection of convolutional layer.





The formula in Convolutional layer.

$$z(u,v) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} x_{i+u,j+v} \cdot k_{rot\,i,j} \cdot \chi(i,j)$$

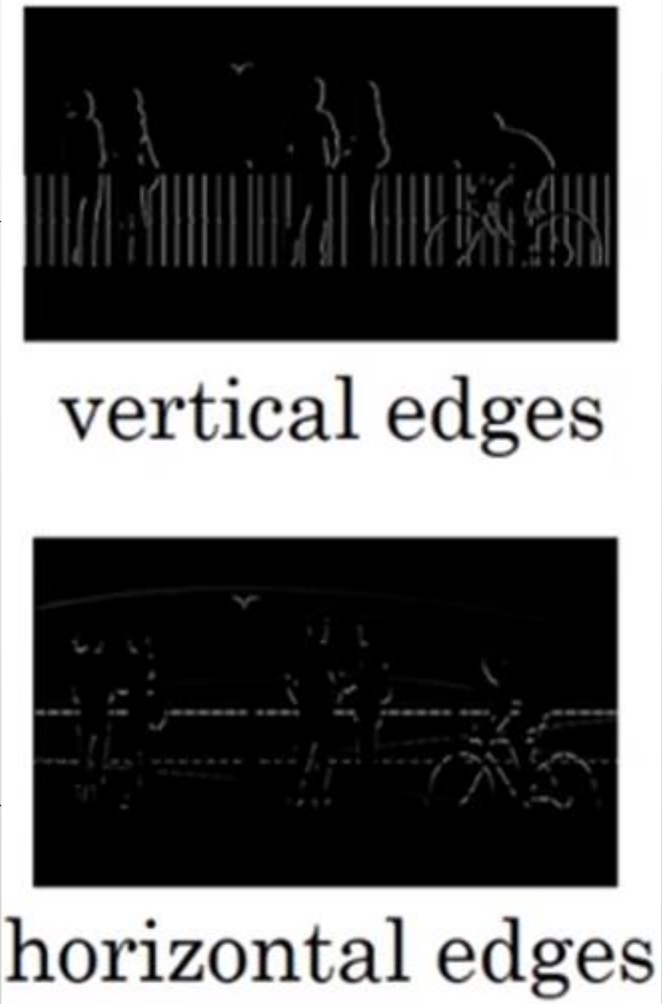
$$\chi(i,j) = \begin{matrix} 1, 0 \leqslant i,j \leqslant n \\ 0, others \end{matrix}$$

2 Convolution in Neural Network

---An example of edge detection

filter0

1	0	-1
1	0	-1
1	0	-1

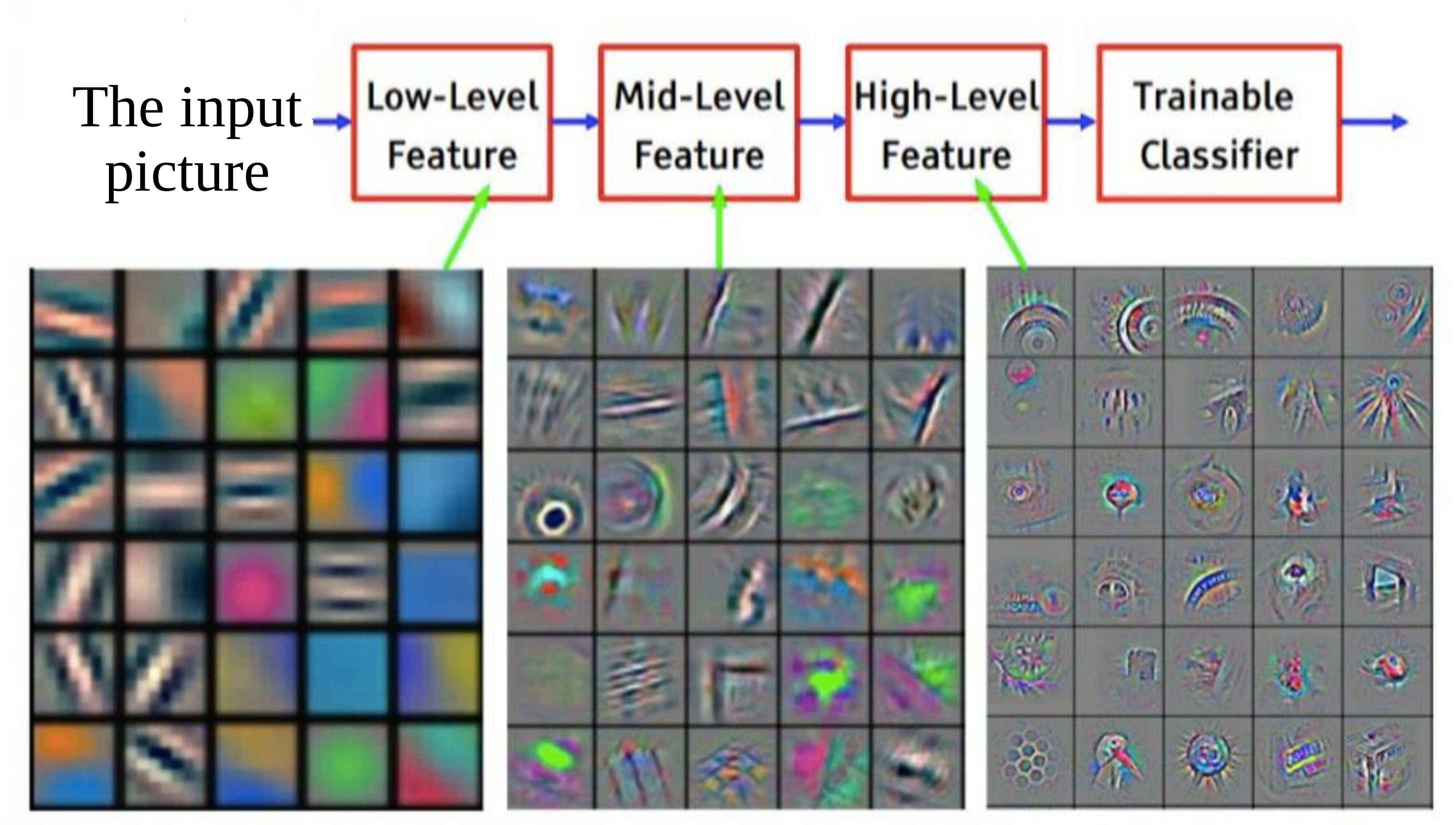


filter1

1	1	1
0	0	0
-1	-1	-1

The picture is from Andrew Ng.

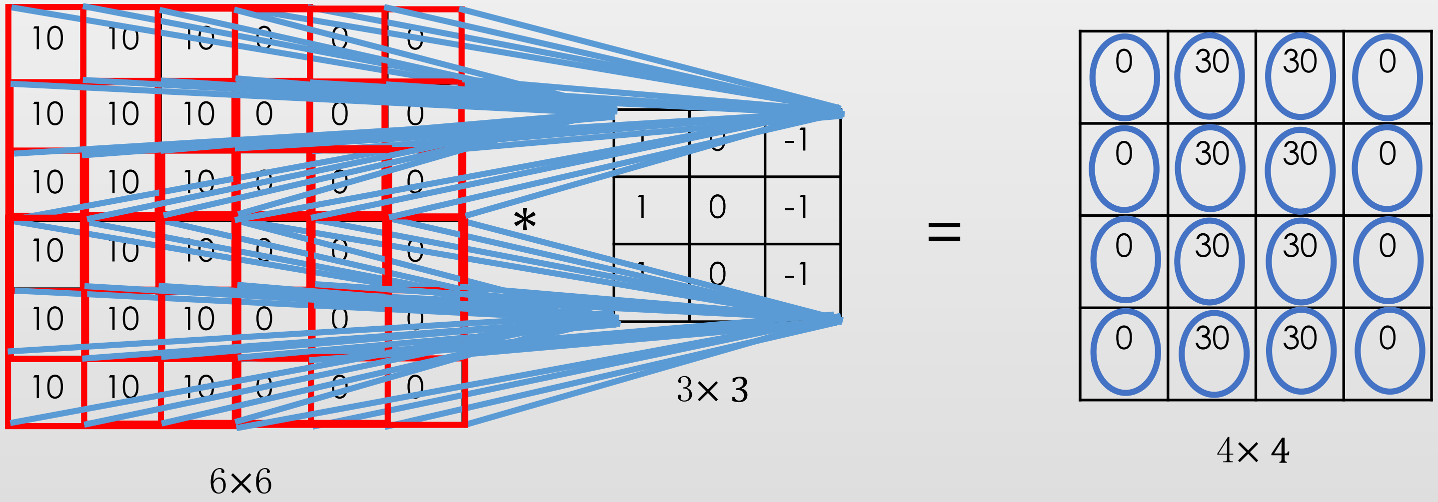
The filters can become more intricate as they start incorporating information from an increasingly larger spatial extent.



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

这意味着在卷积网络中尽管直接连接是很稀疏的，但处在更深的层中的单元可以直接地连接到全部或者大部分输入图像。

2.1 Convolution in Neural Network



Notation:

Image: n×n   filter: f×f

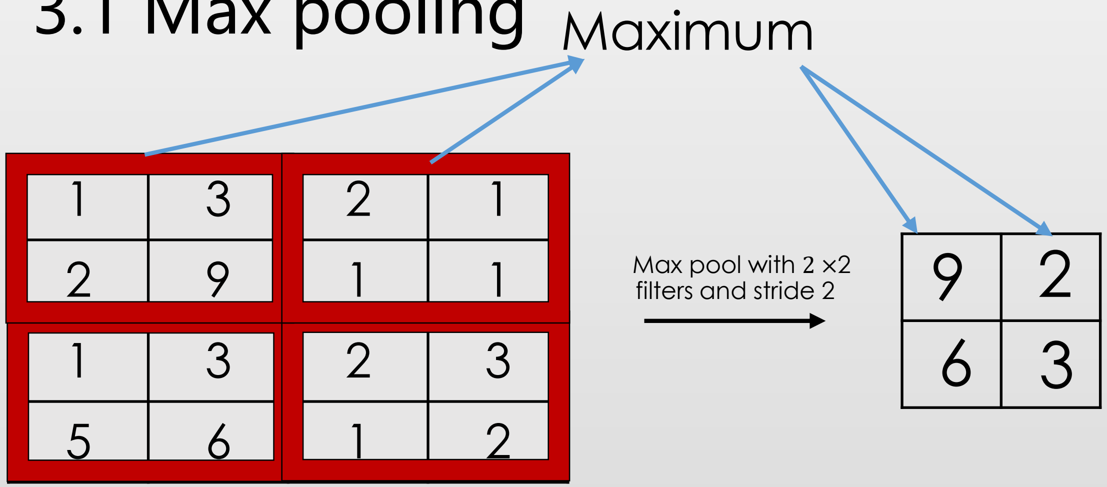
padding : p   stride : s

Then output matrix:

$$\left\lfloor \frac{n+2p-f}{s}+1 \right\rfloor \times \left\lfloor \frac{n+2p-f}{s}+1 \right\rfloor$$

3 . Pooling layers ---Shrinking the image stack

- 3.1 Max pooling



Pooling:

1.Pick a window size(usually 2 or 3)

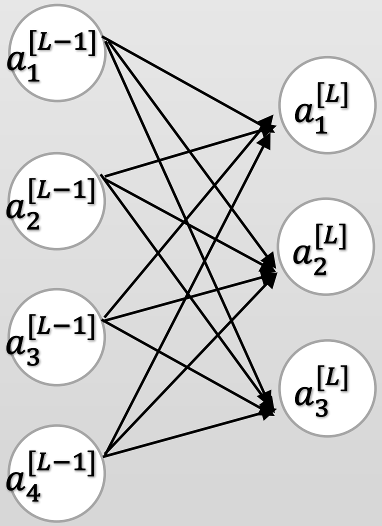
2.Pick a stride(usually 2)

3.Walk your window across your filtered images.

4.From each window , take the maximum value.

4. Full connection layer

The CNNs help extract certain features from the image , then fully connected layer is able to generalize from these features to the output-space.



2. How the BP Algorithm work on CNN?

2.1 已知池化层的δ^l,求上一隐藏层的δ^{l-1}

Feedforward:

1	3	2	1
2	9	1	1
1	3	2	3
5	6	1	2

Max pool with 2 ×2 filters and stride 2 →

9	2
6	3

Backpropagate：

δ^{l-1}

0	0	8	0
0	2	0	0
0	0	0	6
0	4	0	0

←

δ^l

0	0	0	0
0	2	8	0
0	4	6	0
0	0	0	0

2.2 已知卷积层的δ^{l},求上一隐藏层的δ^{l-1}

卷积层的前向传播公式：

$$a^{l} = g\left(z^{l}\right) = g(a^{l-1} * W^{l} + b^{l})$$

在DNN中， δ^{l-1} 和δ^{l}的递推公式：

$$\delta^{l-1} = \frac{\partial \mathrm{J}(W,b)}{\partial z^{l-1}} = \frac{\partial \mathrm{J}(W,b)}{\partial z^{l}}\frac{\partial z^{l}}{\partial z^{l-1}} = \delta^{l}\ \frac{\partial z^{l}}{\partial z^{l-1}}$$

z^{l} 和z^{l-1}的关系：

$$z^{l} = a^{l-1} * W^{l} + b^{l} = g\left(z^{l-1}\right) * W^{l} + b^{l}$$

Thus：

$$\delta^{l-1} = \delta^{l}\,\frac{\partial z^{l}}{\partial z^{l-1}} = \delta^{l} * rot180(W^{l}) \odot g^{'}\left(z^{l-1}\right)$$

2.2.1 Why we need $rot180(W^l)$?

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} * \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} = \begin{pmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{pmatrix}$$

Using the definition of convolution to calculate Z_{ij}:

$$z_{11} = a_{11}w_{11} + a_{12}w_{12} + a_{21}w_{21} + a_{22}w_{22}$$

$$z_{12} = a_{12}w_{11} + a_{13}w_{12} + a_{22}w_{21} + a_{23}w_{22}$$

$$z_{21} = a_{21}w_{11} + a_{22}w_{12} + a_{31}w_{21} + a_{32}w_{22}$$

$$z_{22} = a_{22}w_{11} + a_{23}w_{12} + a_{32}w_{21} + a_{33}w_{22}$$

$$\delta^{l-1} = \frac{\partial J(W,b)}{\partial z^{l-1}} = \frac{\partial J(W,b)}{\partial a^{l-1}} \frac{\partial a^{l-1}}{\partial z^{l-1}} = \nabla a^{l-1} \odot g'(z^{l-1})$$

$$\nabla a^{l-1} = \frac{\partial J(W,b)}{\partial a^{l-1}} = \frac{\partial J(W,b)}{\partial z^{l}} \frac{\partial z^{l}}{\partial a^{l-1}} = \delta^{l} \frac{\partial z^{l}}{\partial a^{l-1}}$$

$$\nabla a_{11} = \delta_{11}w_{11}$$

$$\nabla a_{12} = \delta_{11}w_{12} + \delta_{12}w_{11}$$

$$\nabla a_{13} = \delta_{12}w_{12}$$

$$\nabla a_{21} = \delta_{11}w_{21} + \delta_{21}w_{11}$$

$$\nabla a_{22} = \delta_{11}w_{22} + \delta_{12}w_{21} + \delta_{21}w_{12} + \delta_{22}w_{11}$$

$$\nabla a_{23} = \delta_{12}w_{22} + \delta_{22}w_{12}$$

$$\nabla a_{31} = \delta_{21}w_{21}$$

$$\nabla a_{32} = \delta_{21}w_{22} + \delta_{22}w_{21}$$

$$\nabla a_{33} = \delta_{22}w_{22}$$

2.2.1 Why we need $rot180(W^l)$?

$$\nabla a_{11} = \delta_{11} w_{11}$$

$$\nabla a_{12} = \delta_{11} w_{12} + \delta_{12} w_{11}$$

$$\nabla a_{13} = \delta_{12} w_{12}$$

$$\nabla a_{21} = \delta_{11} w_{21} + \delta_{21} w_{11}$$

$$\nabla a_{22} = \delta_{11} w_{22} + \delta_{12} w_{21} + \delta_{21} w_{12} + \delta_{22} w_{11}$$

$$\nabla a_{23} = \delta_{12} w_{22} + \delta_{22} w_{12}$$

$$\nabla a_{31} = \delta_{21} w_{21}$$

$$\nabla a_{32} = \delta_{21} w_{22} + \delta_{22} w_{21}$$

$$\nabla a_{33} = \delta_{22} w_{22}$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \delta_{11} & \delta_{12} & 0 \\ 0 & \delta_{21} & \delta_{22} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} * \begin{pmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{pmatrix} = \begin{pmatrix} \nabla a_{11} & \nabla a_{12} & \nabla a_{13} \\ \nabla a_{21} & \nabla a_{22} & \nabla a_{23} \\ \nabla a_{31} & \nabla a_{32} & \nabla a_{33} \end{pmatrix}$$

$$= \delta^l * rot180(W^l)$$

$$\delta^{l-1} = \delta^l \frac{\partial z^l}{\partial z^{l-1}} = \nabla a^{l-1} \odot g^{'}\left(z^{l-1}\right) = \delta^l * rot180(W^l) \odot g^{'}\left(z^{l-1}\right)$$

2.3 已知卷积层的δ^l,推导该层的w, b的梯度

卷积层z和w, b的关系为：

$$z^l = a^{l-1} * W^l + b$$

Thus：

$$\frac{\partial \mathrm{J}(W,b)}{\partial W^l} = \frac{\partial \mathrm{J}(W,b)}{\partial z^l}\frac{\partial z^l}{\partial W^l} = a^{l-1} * \delta^l$$

$$\frac{\partial J(w,b)}{\partial W^l} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} * \begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{pmatrix}$$

$$\frac{\partial J(w,b)}{\partial b^l} = \sum_{u,v} \left(\delta^l\right)_{u,v}$$

2.3.1 Don't need $rot180(W^l)$

$$\frac{\partial \mathrm{J}(W,b)}{\partial W^l} = \frac{\partial \mathrm{J}(W,b)}{\partial z^l}\frac{\partial z^l}{\partial W^l} = a^{l-1} * \delta^l$$

$$z_{11} = a_{11}w_{11} + a_{12}w_{12} + a_{21}w_{21} + a_{22}w_{22}$$

$$z_{12} = a_{12}w_{11} + a_{13}w_{12} + a_{22}w_{21} + a_{23}w_{22}$$

$$z_{21} = a_{21}w_{11} + a_{22}w_{12} + a_{31}w_{21} + a_{32}w_{22}$$

$$z_{22} = a_{22}w_{11} + a_{23}w_{12} + a_{32}w_{21} + a_{33}w_{22}$$

$$\frac{\partial \mathrm{J}(W,b)}{\partial W_{11}^l} = a_{11}\delta_{11} + a_{12}\delta_{12} + a_{21}\delta_{21} + a_{22}\delta_{22}$$

$$\frac{\partial \mathrm{J}(W,b)}{\partial W_{12}^l} = a_{12}\delta_{11} + a_{13}\delta_{12} + a_{22}\delta_{21} + a_{23}\delta_{22}$$

$$\frac{\partial \mathrm{J}(W,b)}{\partial W_{21}^l} = a_{21}\delta_{11} + a_{22}\delta_{12} + a_{31}\delta_{21} + a_{32}\delta_{22}$$

$$\frac{\partial \mathrm{J}(W,b)}{\partial W_{22}^l} = a_{22}\delta_{11} + a_{23}\delta_{12} + a_{32}\delta_{21} + a_{33}\delta_{22}$$

$$\frac{\partial J(w,b)}{\partial W^l} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} * \begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{pmatrix}$$

$$\frac{\partial J(w,b)}{\partial b^l} = \sum_{u,v} \left(\delta^l\right)_{u,v}$$