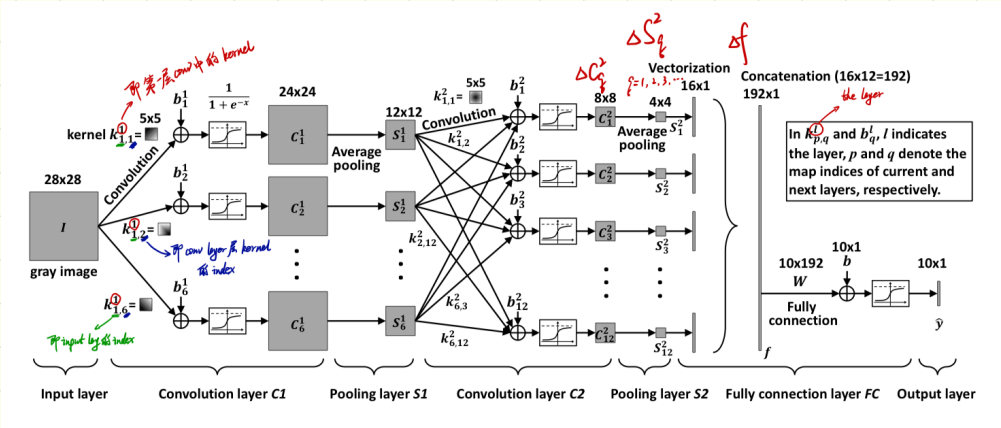


* Forward

* Input : 28 x 28 (2D)

* C1 layer: k^1_p (size 5x5), b^1_p (size 1x1), $p=1,2,\cdots,6$ 即 C1 layer 共有 6 个 filter, 每个 5x5

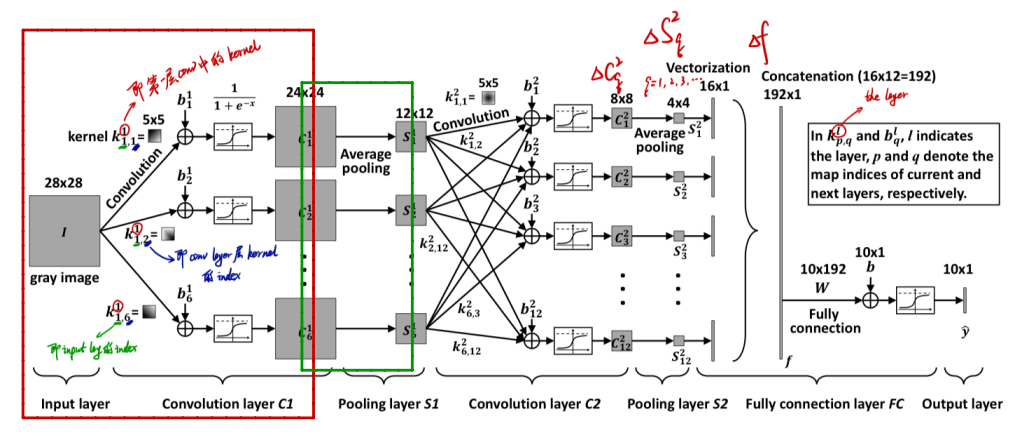
5x5x6

* C2 layer: k^2_{pq} (size 5x5), b^1_q (size 1x1), $q=1,2,\cdots,12$, 即 C2 layer 共有 6个 filter 每个 5x5x6

5x5x6 x12

* FC layer : W (size 10x192), b (size 10x1)

output 有10个 class



* Conv 1

$$\sigma(x) = \frac{1}{1+e^{-x}} \Rightarrow \sigma'(x) = \sigma(x)\left(1-\sigma(x)\right)$$

$$C^1_p(i,j) = \sigma\left(\sum_{u=-2}^{2}\sum_{v=-2}^{2} I(i-u, j-v)\cdot k^1_p(u,v) + b^1_p\right) \Rightarrow \text{element}$$

$$\Downarrow$$

$$\underbrace{C^1_p}_{24\times 24} = \sigma\left(\underbrace{I}_{28\times 28} * \underbrace{k^1_p}_{5\times 5} + \underbrace{b^1_p}_{1\times 1}\right)$$

p-th kernel

p-th bais

* Pooling Layer S1

$$S^1_p(i,j) = \frac{1}{4}\sum_{u=0}^{1}\sum_{v=0}^{1} C^1_p(2i-u, 2j-v) \quad , \quad i,j = 1,2,\cdots,12$$

12×12 $\quad$ 24×24

Average Pooling

$k'_{ip}(u,v)$ $\quad -2, -1, 0, 1, 2$

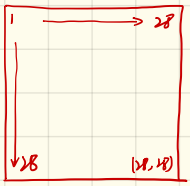
回去的时候，带美元现金，
别换成人民币，
也别跟别人说你有美金.

$$G'_{(1,1)} = \sigma\left(\sum_{u=-2}^{2}\sum_{v=-2}^{2} I_{(1-u,1-v)} \cdot k'_p(u,v) + k'_p\right)$$

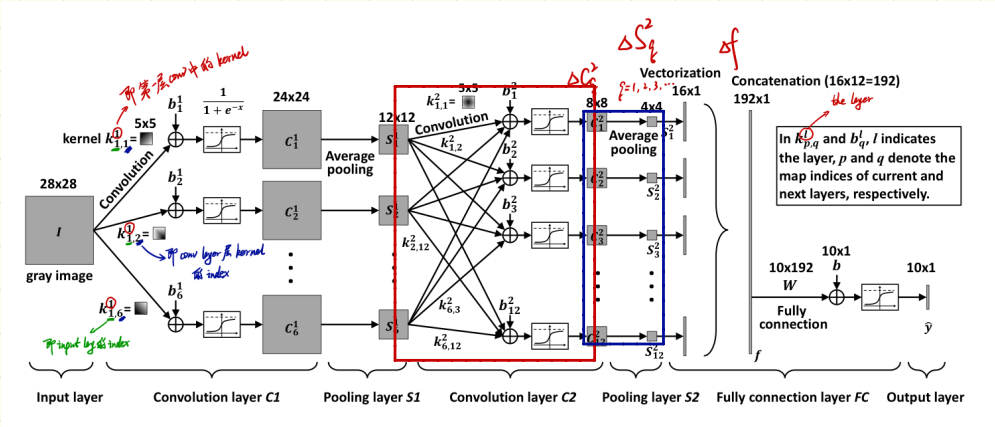
$$= \sigma\Big(I_{(3,3)} \cdot k'_p(-2,-2)$$
$$+ I_{(3,2)} \cdot k'_p(-2,-1)$$
$$+ I_{(3,1)} \cdot k'_p(-2,0)$$
$$+ I_{(3,0)} \cdot k'_p(-2,1)$$
$$+ I_{(3,-1)} \cdot k'_p(-2,2)$$

$$\left\{\begin{array}{l} + I_{(2,3)} \cdot k'_p(-1,-2) \\ + I_{(2,2)} \cdot k'_p(-1,-1) \\ + I_{(2,1)} \cdot k'_p(-1,0) \\ + I_{(2,0)} \cdot k'_p(-1,1) \\ + I_{(2,-1)} \cdot k'_p(-1,2) \end{array}\right.$$

I



k'_{ip}



* $$C^2_q(i,j) = \sigma\left(\sum_{p=1}^{6}\sum_{u=-2}^{2}\sum_{v=-2}^{2} S^1_p(i-u,j-v)\cdot k^2_{p,q}(u,v) + b^2_q\right)$$ ⇒ element

input 的 kernel 的 p-th layer

↳ C2 的 q-th filter

$$C^2_q = \sigma\left(\sum_{p=1}^{6} S^1_p * k^2_{pq} + b^2_q\right)$$

8x8 12x12 $q=1,2,\cdots,12$

其中 $\sigma(x) = \frac{1}{1+e^{-x}}$

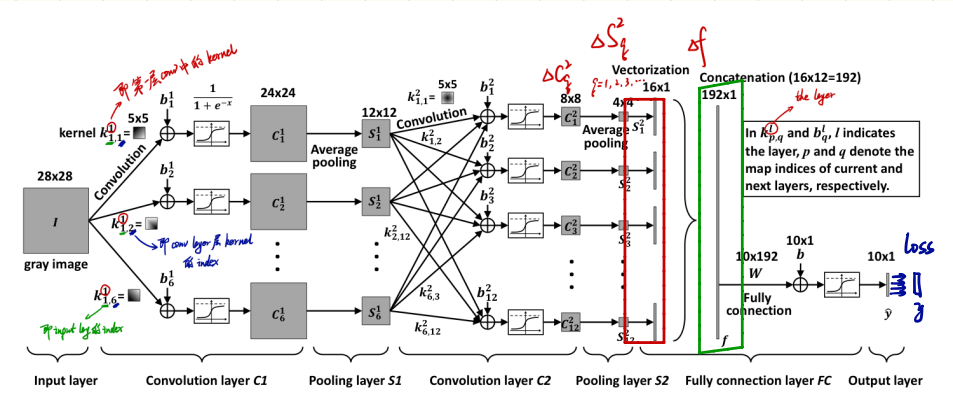
C^2: 8x8 x 12

12个 filter 的 output.

* $$S^2_q(i,j) = \underbrace{\frac{1}{4}\sum_{u=0}^{1}\sum_{v=0}^{1}}_{2\times 2 \text{ average pooling}} C^2_q(2i-u,2j-v), \quad i,j=1,2,\cdots,4$$

8x2

4x4



* $f = F(\{S_q^2\}_{q=1,2,\cdots,12})$

將 12 个 S_q^2 vectorization and concatenation

$4 \times 4 \times 12 \Rightarrow 192$

* Fully Connection layer FC

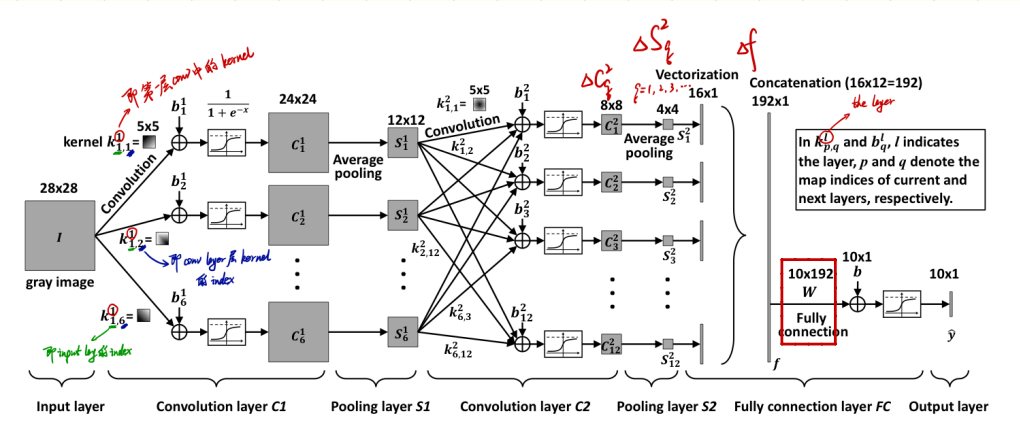
$$\hat{y} = \sigma(W \times f + b)$$

$10 \times 192 \quad 192 \times 1 \quad 1 \times 1$

10×1

* Loss Function

$$L = \frac{1}{2} \sum_{i=1}^{10} (\hat{y}_{(i)} - y_{(i)})^2$$



2.1 ΔW (size 10×192)

$$L = \frac{1}{2}\sum_{i=1}^{10}\left(\hat{y}_{(i)} - y_{(i)}\right)^2$$

$$\Delta W_{(i,j)} = \frac{\partial L}{\partial W_{(i,j)}} = \frac{\partial L}{\partial \hat{y}_{(i)}} \cdot \frac{\partial \hat{y}_{(i)}}{\partial W_{(i,j)}}$$

$$= \left(\hat{y}_{(i)} - y_{(i)}\right) \cdot \frac{\partial \hat{y}_{(i)}}{\partial W_{(i,j)}}$$

$$\hat{y}_{(i)} = \sigma\left(\sum_{j=1}^{192} W_{(i,j)} \times f_{(j)} + b_{(i)}\right)$$

$$= \sigma(x)$$

$$= \left(\hat{y}_{(i)} - y_{(i)}\right) \cdot \hat{y}_{(i)} \cdot \left(1 - \hat{y}_{(i)}\right) \cdot \frac{\partial}{\partial W_{(i,j)}}\left(\sum_{j=1}^{192} W_{(i,j)} \times f_{(j)} + b_{(i)}\right)$$

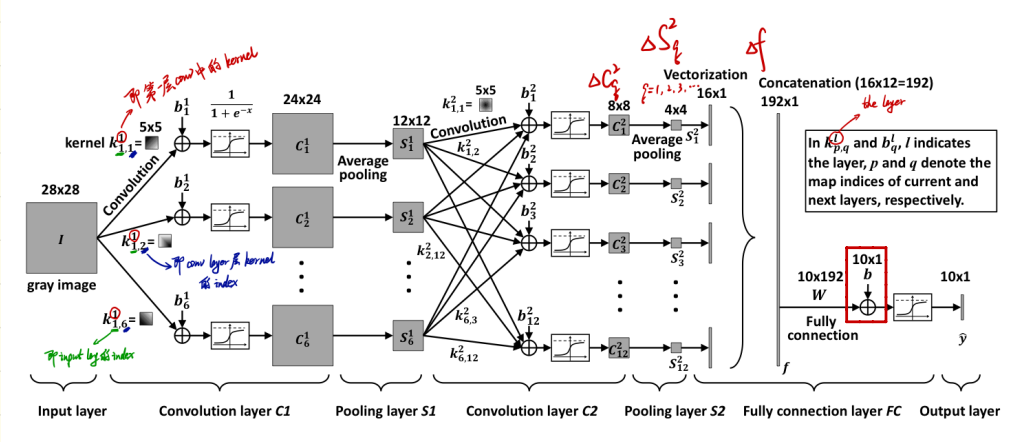
$$\sigma'(x) = \sigma(x)\left(1 - \sigma(x)\right)$$

$$\hat{y}'_{(i)} = \hat{y}_{(i)}\left(1 - \hat{y}_{(i)}\right)$$

$$= \left(\hat{y}_{(i)} - y_{(i)}\right) \cdot \hat{y}_{(i)} \cdot \left(1 - \hat{y}_{(i)}\right) \cdot f_{(j)}$$

def: $\Delta \hat{y}_{(i)} = \left(\hat{y}_{(i)} - y_{(i)}\right) \cdot \hat{y}_{(i)} \cdot \left(1 - \hat{y}_{(i)}\right)$

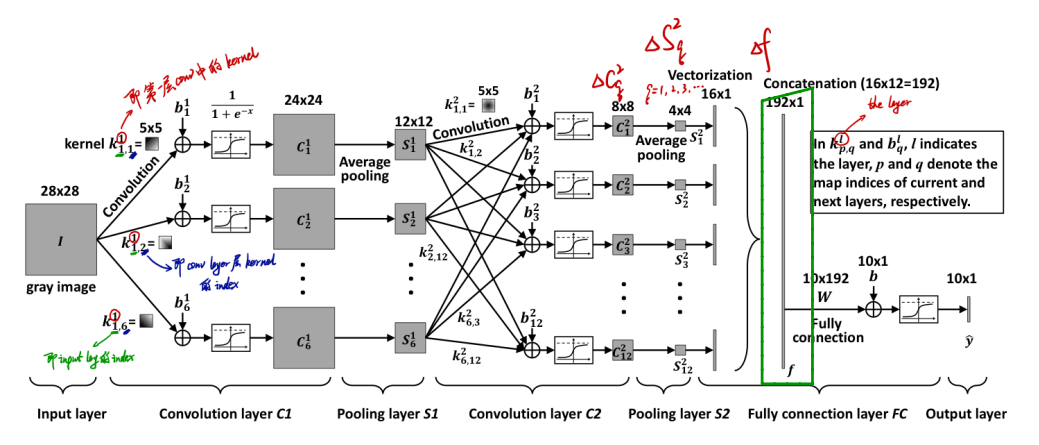
$$\Delta W_{(i,j)} = \Delta \hat{y}_{(i)} \cdot f_{(j)} \Rightarrow \Delta W = \underset{10 \times 1}{\Delta \hat{y}} \times \underset{1 \times 192}{f^T}$$



$$* \quad \Delta b_{(i)} = \frac{\partial L}{\partial b_{(i)}} = \frac{\partial L}{\partial \hat{y}_{(i)}} \cdot \frac{\partial \hat{y}_{(i)}}{\partial b_i}$$

$$= \underbrace{\left(\hat{y}_{(i)} - y_{(i)}\right) \cdot \hat{y}_{(i)} \cdot \left(1 - \hat{y}_{(i)}\right)}_{\Delta \hat{y}_{(i)}} \cdot 1$$

$$\Rightarrow \Delta b = \Delta \hat{y}$$



$$\Delta f_{(j)} = \frac{\partial L}{\partial f_{(i)}} \qquad L = \frac{1}{2}\sum_{i=1}^{10}\left(\hat{y}_{(i)} - y_{(i)}\right)^2$$

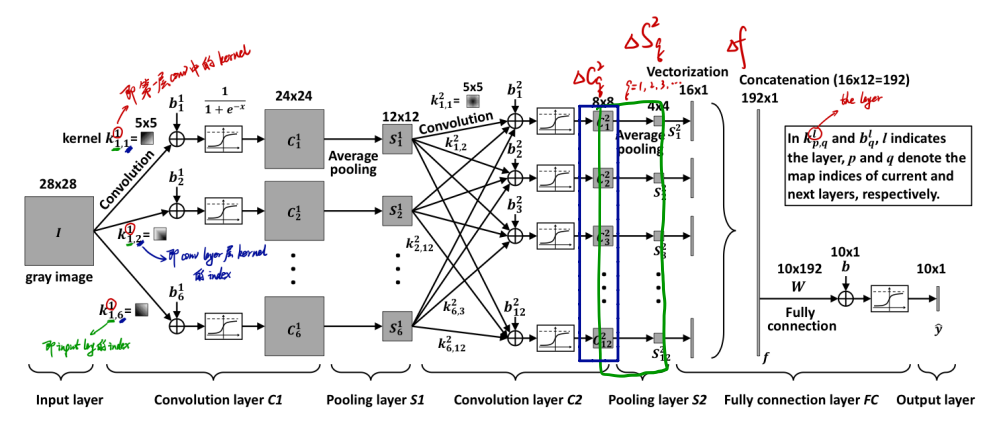
$$= \sum_{i=1}^{10} \frac{\partial L}{\partial \hat{y}_{(i)}} \cdot \frac{\partial \hat{y}_{(i)}}{\partial f_{(j)}}$$

$$= \sum_{i=1}^{10} \left(\hat{y}_{(i)} - y_{(i)}\right) \cdot \hat{y}_{(i)}\left(1-\hat{y}_{(i)}\right) \cdot \frac{\partial}{\partial f_{(j)}}\left(\sum_{j=1}^{192} W_{(i,j)} \times f_{(j)} + b_{(i)}\right)$$

$$= \sum_{i=1}^{10} \left(\hat{y}_{(i)} - y_{(i)}\right) \cdot \hat{y}_{(i)} \cdot \left(1-\hat{y}_{(i)}\right) \cdot W_{(i,j)}$$

$$= \sum_{i=1}^{10} \Delta \hat{y}_{(i)} \cdot W_{(i,j)} \quad \Rightarrow \text{element}$$

$$\Rightarrow \underbrace{\Delta f}_{192\times 1} = \underbrace{W^T}_{192\times 10} \times \underbrace{\Delta \hat{y}}_{10\times 1}$$



$$* \ \{\Delta S_q^2\}_{q=1,2,\cdots,12} = F^{-1}(\Delta f)$$

$$\Delta C_q^2(i,j) = \frac{1}{4}\Delta S_q^2(\lceil i/2\rceil, \lceil j/2\rceil), \qquad i=1,2,\cdots,8$$

8x8 ← 4x4 →

$\Downarrow$

ΔS_q^2 均分至 4个 C_q^2 elements.

$$* \ \Delta k_{p,q}^2(u,v) = \frac{\partial L}{\partial k_{p,q}^2(u,v)}$$

$$= \sum_{i=1}^{8}\sum_{j=1}^{8} \frac{\partial L}{\partial C_q^2(i,j)} \cdot \frac{\partial C_q^2(i,j)}{\partial k_{p,q}^2(u,v)}$$

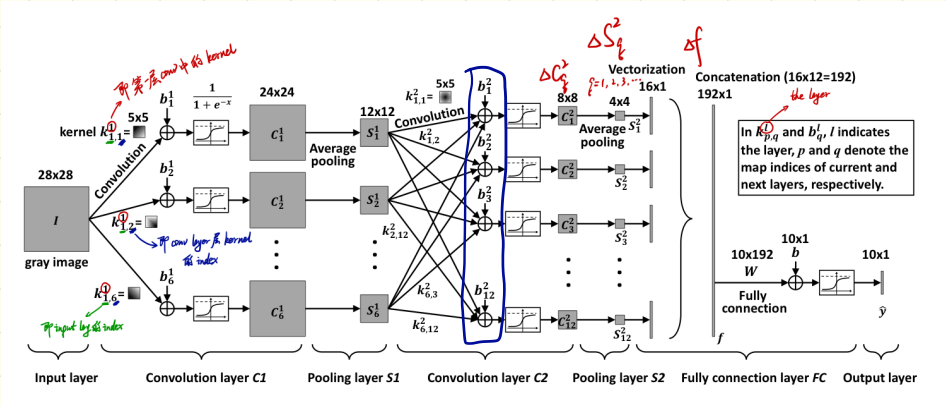
$$= \sum_{i=1}^{8}\sum_{j=1}^{8} \Delta C_q^2(i,j) \cdot \frac{\partial}{\partial k_{p,q}^2(u,v)} \sigma\left(\sum_{p=1}^{6}\sum_{u=-2}^{2}\sum_{v=-2}^{2} S_p^1(i-u,j-v)\cdot k_{p,q}^2(u,v) + b_q^2\right)$$

$$= \sum_{i=1}^{8}\sum_{j=1}^{8} \underbrace{\Delta C_q^2(i,j)\cdot\left(1-C_q^2(i,j)\right)\cdot C_q^2(i,j)}_{\Delta C_{q,\sigma}^2(i,j) = \Delta C_q^2(i,j)\cdot C_q^2(i,j)\cdot(1-C_q^2(i,j))} \cdot S_p^1(i-u,j-v)$$

$\hookrightarrow S_{p,rot}^1(i-u, v-j)$

$$\Delta k_{p,q}^2(u,v) = \sum_{i=1}^{8}\sum_{j=1}^{8} S_{p,rot180}^1(u-i, v-j)\cdot \Delta C_{q,\sigma}^2(i,j)$$

$$\Rightarrow \Delta k_{p,q}^2 = S_{p,rot180}^1 * C_{q,\sigma}^2$$

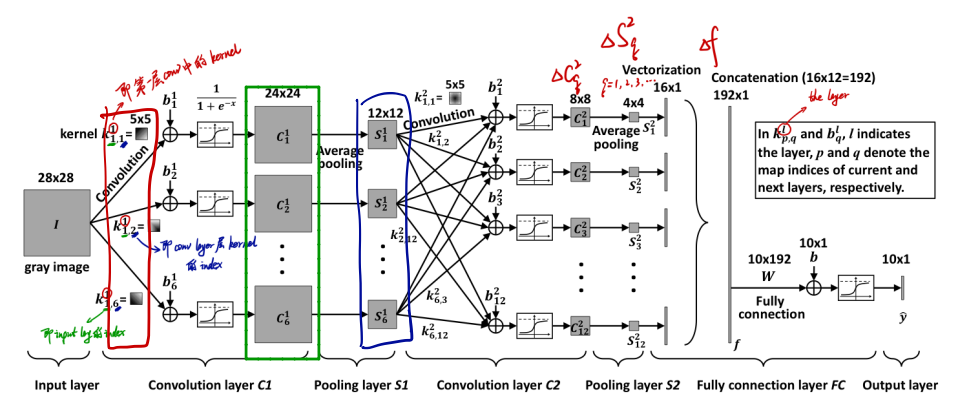


$$* \quad \Delta b_q^2 = \frac{\partial L}{\partial b_q^2} = \sum_{i=1}^{8}\sum_{j=1}^{8} \frac{\partial L}{\partial C_q^2(i,j)} \cdot \frac{\partial C_q^2(i,j)}{\partial b_q^2}$$

$$= \sum_{i=1}^{8}\sum_{j=1}^{8} \Delta C_q^2(i,j) \cdot C_q^2(i,j) \cdot \left(1 - C_q^2(i,j)\right) \cdot \frac{\partial}{\partial b_q^2}\left(\sum_{p=1}^{6}\sum_{u=-2}^{2}\sum_{v=-2}^{2} S_p^1(i-u, j-v) \cdot k_{p,q}^2(u,v) + b_q^2\right)$$

$$= \sum_{i=1}^{8}\sum_{j=1}^{8} \Delta C_q^2(i,j) \cdot C_q^2(i,j) \cdot \left(1 - C_q^2(i,j)\right) \cdot 1$$

$$= \sum_{i=1}^{8}\sum_{j=1}^{8} \Delta C_{q,\sigma}^2(i,j)$$



$$\Delta S^1_p(i,j) = \frac{\partial L}{\partial S^1_p(i,j)} = \sum_{q=1}^{12}\sum_{u=-2}^{2}\sum_{v=-2}^{2} \Delta G^2_{q,\sigma}(i+u, j+v)\cdot k^2_{p,q}(u,v)$$

$$\Downarrow$$

$$\Delta S^1_p = \sum_{q=1}^{12} \Delta G^2_{q,\sigma} * k^2_{p,q,\,rot180}$$

$$\Delta G^1_p(i,j) = \frac{1}{4}\Delta S^1_p(\lceil i/2\rceil, \lceil j/2\rceil) \qquad i,j = 1,2,\cdots,24$$

$\rightarrow 12\times 12$

$\swarrow$ 24×24

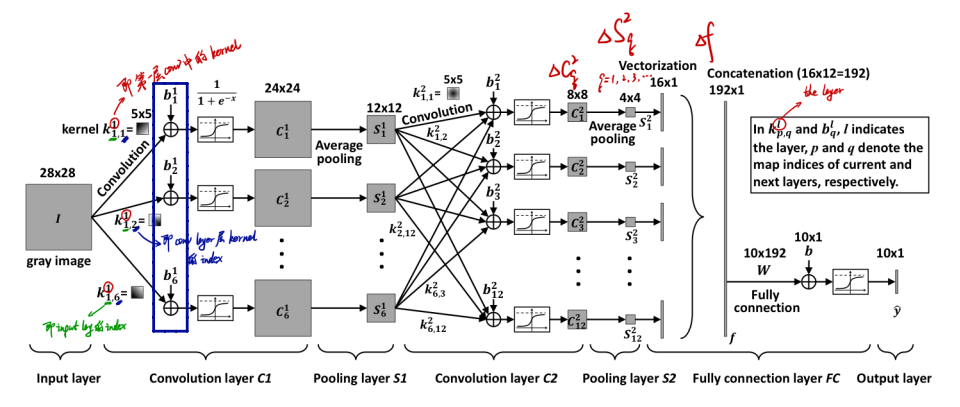
$$\Delta k^1_{1,p}(u,v) = \frac{\partial L}{\partial k^1_{1,p}(u,v)} = \sum_{i=1}^{24}\sum_{j=1}^{24} \frac{\partial L}{\partial G^1_p(i,j)}\cdot\frac{\partial G^1_p(i,j)}{\partial k^1_p(u,v)}$$

$$= \sum_{i=1}^{24}\sum_{j=1}^{24} \Delta G^1_p(i,j)\cdot G^1_p(i,j)\left(1-G^1_p(i,j)\right)\cdot\frac{\partial}{\partial k^1_p(u,v)}\left(\sum_{u=-2}^{2}\sum_{v=-2}^{2} I(i-u,j-v)\cdot k^1_p(u,v) + b^1_p\right)$$

$$= \sum_{i=1}^{24}\sum_{j=1}^{24} \underbrace{\Delta G^1_p(i,j)\cdot G^1_p(i,j)\cdot\left(1-G^1_p(i,j)\right)}_{\Delta G^1_{p,\sigma}(i,j)}\cdot I(i-u,j-v)$$

$$\Delta k^1_{1,p}(u,v) = \sum_{i=1}^{24}\sum_{j=1}^{24} I_{rot180}(u-i, v-j)\cdot \Delta G^1_{p,\sigma}(i,j) \qquad \text{element}$$

$$\Rightarrow \Delta k^1_{1,p} = I_{rot180} * \Delta G^1_{p,\sigma} \qquad \text{matrix}$$



$$\Delta b_p^1 = \frac{\partial L}{\partial b_p^1} = \sum_{i=1}^{24}\sum_{j=1}^{24} \frac{\partial L}{\partial G_p^1(i,j)} \cdot \frac{\partial G_p^1(i,j)}{\partial b_p^1}$$

$$= \sum_{i=1}^{24}\sum_{j=1}^{24} \Delta G_p^1(i,j) \cdot G_p^1(i,j) \cdot \left(1 - G_p^1(i,j)\right) \cdot \frac{\partial}{\partial b_p^1}\left(\sum_{u=-2}^{2}\sum_{v=-2}^{2} I(i-u, j-v) \cdot k_p^1(u,v) + b_p^1\right)$$

$$= \sum_{i=1}^{24}\sum_{j=1}^{24} \Delta G_p^1(i,j) \cdot G_p^1(i,j) \cdot \left(1 - G_p^1(i,j)\right)$$

$$= \sum_{i=1}^{24}\sum_{j=1}^{24} \Delta G_{p,\sigma}^1(i,j)$$

Parameter Update

$$\begin{cases} k^1_{1,p} = k^1_{1,p} - \underline{\underline{\alpha}} \cdot \Delta k^1_{1,p} \\ b^1_p = b^1_p - \underline{\underline{\alpha}} \cdot \Delta b^1_p \end{cases}$$

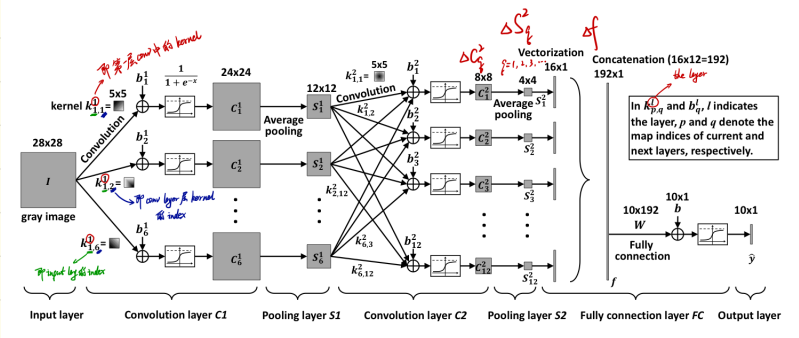
$$\begin{cases} k^2_{p,q} = k^2_{p,q} - \underline{\underline{\alpha}} \cdot \Delta k^2_{p,q} \\ b^2_q = b^2_q - \underline{\underline{\alpha}} \cdot \Delta b^2_q \end{cases}$$

$$\begin{cases} W = W - \underline{\underline{\alpha}} \cdot \Delta W \\ b = b - \underline{\underline{\alpha}} \cdot \Delta b \end{cases}$$

conv1 $\begin{cases} k^1_{1,p} \leftarrow k^1_{1,p} - \alpha \cdot \Delta k^1_{1,p} \\ b^1_p \leftarrow b^1_p - \alpha \cdot \Delta k^1_{1,p} \end{cases}$

conv2 $\begin{cases} k^2_{p,q} \leftarrow k^2_{p,q} - \alpha \cdot \Delta k^2_{p,q} \\ b^2_p \leftarrow b^2_p - \alpha \cdot \Delta b^2_q \end{cases}$

fc $\begin{cases} W \leftarrow W - \alpha \cdot \Delta W \\ b \leftarrow b - \alpha \cdot \Delta b \end{cases}$



1.2 Convolution Layer C1

$$C_p^1 = \sigma(I * k_{1,p}^1 + b_p^1), \text{ where } \sigma(x) = \frac{1}{1+\exp^{-x}} \tag{4}$$

$$C_p^1(i,j) = \sigma\left(\sum_{u=-2}^{2}\sum_{v=-2}^{2} I(i-u, j-v)\cdot k_{1,p}^1(u,v) + b_p^1\right) \tag{5}$$

where $p = 1, 2, \cdots, 6$ because there are 6 feature maps on C1 layer, $*$ denotes the convolution, and i, j are row and column indices of the feature map. Only keeping those parts of the convolution that are computed without the zero-padded edges, the size of C_p^1 is 24×24, rather than 28×28 like I.

2.1 ΔW (size 10×192)

$$\begin{aligned}
\Delta W(i,j) &= \frac{\partial L}{\partial W(i,j)} && (14)\\
&= \frac{\partial L}{\partial \hat{y}(i)} \cdot \frac{\partial \hat{y}(i)}{\partial W(i,j)} && (15)\\
&= (\hat{y}(i) - y(i)) \cdot \frac{\partial}{\partial W(i,j)} \sigma\left(\sum_{j=1}^{192} W(i,j) \times f(j) + b(i)\right) && (16)\\
&= (\hat{y}(i) - y(i)) \cdot \hat{y}(i)(1-\hat{y}(i)) \cdot f(j) && (17)
\end{aligned}$$

Let $\Delta\hat{y}(i) = (\hat{y}(i) - y(i)) \cdot \hat{y}(i)(1-\hat{y}(i))$, whose size is 10×1, then

$$\begin{aligned}
\Delta W(i,j) &= \Delta\hat{y}(i) \cdot f(j) && (18)\\
\Longrightarrow \Delta W &= \Delta\hat{y} \times f^T && (19)
\end{aligned}$$

2.2 Δb (size 10×1)

$$\begin{aligned}
\Delta b(i) &= \frac{\partial L}{\partial b(i)} && (20)\\
&= \frac{\partial L}{\partial \hat{y}(i)} \cdot \frac{\partial \hat{y}(i)}{\partial b(i)} && (21)\\
&= (\hat{y}(i) - y(i)) \cdot \frac{\partial}{\partial b(i)} \sigma\left(\sum_{j=1}^{192} W(i,j) \times f(j) + b(i)\right) && (22)\\
&= (\hat{y}(i) - y(i)) \cdot \hat{y}(i)(1-\hat{y}(i)) && (23)\\
\Longrightarrow \Delta b &= \Delta\hat{y} && (24)
\end{aligned}$$

fully connected layer

$$W \leftarrow W - \alpha \cdot \Delta W$$

$$b \leftarrow b - \alpha \cdot \Delta b$$

$$\frac{\partial L}{\partial \alpha} = \frac{\partial L}{\partial W} \cdot \frac{\partial W}{\partial \alpha} + \frac{\partial L}{\partial b} \cdot \frac{\partial b}{\partial \alpha}$$

$$= \sum_i \sum_j \frac{\partial L}{\partial W_{(i,j)}} \cdot \frac{\partial W_{(i,j)}}{\partial \alpha} + \sum_i \frac{\partial L}{\partial b_{(i)}} \cdot \frac{\partial b_{(i)}}{\partial \alpha}$$

$$= \sum_i \sum_j \Delta W_{(i,j)} \cdot \left(-\Delta W_{(i,j)}\right) + \sum_i \Delta b_{(i)} \cdot \left(-\Delta b_{(i)}\right)$$

$$= - \left(\sum_i \sum_j \Delta W_{(i,j)} \cdot \Delta W_{(i,j)} + \sum_i \Delta b_{(i)} \cdot \Delta b_{(i)} \right)$$

$$k^2_{p,q} \leftarrow k^2_{p,q} - \alpha \cdot \Delta k^2_{p,q}$$

$$b^2_q \leftarrow b^2_q - \alpha \cdot \Delta k^2_{p,q}$$

Conv 2.

q个 kernel. 每个 p层

$$\frac{\partial L}{\partial \alpha} = \sum_p \sum_q \frac{\partial L}{\partial k^2_{p,q}(u,v)} \cdot \frac{\partial k^2_{p,q}(u,v)}{\partial \alpha} + \sum_q \frac{\partial L}{\partial b^2_q} \cdot \frac{\partial b^2_q}{\partial \alpha}$$

$$= \sum_p \sum_q \Delta k^2_{p,q}(u,v) \cdot \left(-\Delta k^2_{p,q}(u,v)\right) + \sum_q \Delta b^2_q \cdot \left(-\Delta b^2_q\right)$$

$$= -\left(\sum_p \sum_q \Delta k^2_{p,q}(u,v) \cdot \Delta k^2_{p,q}(u,v) + \sum_q \Delta b^2_q \cdot \Delta b^2_q\right)$$

Conv 1.

$$k'_{ip} \leftarrow k'_{ip} - \alpha \cdot \Delta k'_{ip}$$

$$b'_p \leftarrow b'_p - \alpha \cdot \Delta b'_p$$

$$\frac{\partial L}{\partial \alpha} = \sum_p \frac{\partial L}{\partial k'_{ip}} \cdot \frac{\partial k'_{ip}}{\partial \alpha} + \sum_p \frac{\partial L}{\partial b'_p} \cdot \frac{\partial b'_p}{\partial \alpha}$$

$$= \sum_p \Delta k'_{ip} \cdot (-\Delta k'_{ip}) + \sum_p \Delta b'_p \cdot (-\Delta b'_p)$$

$$= -\left(\sum_p \Delta k'_{ip} \cdot \Delta k'_{ip} + \sum_p \Delta k'_p \cdot \Delta b'_p \right)$$

* 对 FC, conv1 and conv2, $\frac{\partial L}{\partial \alpha}$ 恒为负 $\Rightarrow$ α 朝一个方向改变

$\Downarrow$

* update: $\alpha \leftarrow \alpha - \beta \cdot \frac{\partial L}{\partial \alpha}$ $\qquad$ $\alpha \leftarrow \alpha - \alpha \cdot \frac{\partial L}{\partial \alpha}$