1 Layers

1. 2D Convolution:

$$\mathcal{C}(x^l_{ijkm}, w^l_{npmq}) = \sum_{\substack{n\in[0,R^l]\\ p\in[0,C^l]\\ m\in[0,N^l]}} x^l_{i,(s_1j+n),(s_2k+p),m} w^l_{(R^l-n),(C^l-p),m,q} \tag{1}$$

2. Max Pooling 2×2:

$$\mathcal{P}(x^l_{ijmn}) = \sum_{\substack{j\in[0,H^l]\\ m\in[0,W^l]}} x^l_{ijmn}\delta^l_{ijpmqn}, \quad \delta^l_{ijpmqn} = \begin{cases} 1 & p,q = \underset{\substack{\alpha\in[s_1j,s_1j+k_1]\\ \beta\in[s_2m,s_2m+k_2]}}{\arg\max} x^l_{i\alpha\beta n} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

3. Flatten:

$$\mathcal{F}(x^l_{ijkm}) = x^l_{i,(H^l\times W^l\times m+H^l\times k+j)} \tag{3}$$

4. Fully Connected (Dense):

$$\mathcal{D}(x^l_{ij}, w^l_{jk}) = \sum_{j\in[0,H^l\times W^l\times N^l]} x^l_{ij} w^l_{jk} \tag{4}$$

2 Activation Functions

1. ReLU:

$$\mathcal{R}(z^l_i) = \begin{cases} z^l_i & z^l_i \geq 0 \\ 0 & z^l_i < 0 \end{cases} \tag{5}$$

$$\partial_{z^l_i}\mathcal{R} = \begin{cases} 1 & z^l_i > 0 \\ 1/2 & z^l_i = 0 \\ 0 & z^l_i < 0 \end{cases} \tag{6}$$

2. Softmax:

$$\mathcal{S}(z_i^l) = \frac{e^{z_i^l}}{\sum\limits_k e^{z_k^l}} \tag{7}$$

$$\partial_{z_i^l}\mathcal{S} = \begin{cases} \frac{1}{\left(\sum\limits_k e^{z_k^l}\right)^2} e^{z_i^l} \sum\limits_{k\neq j} e^{z_k^l} & i = j \\ \frac{1}{\left(\sum\limits_k e^{z_k^l}\right)^2} e^{z_i^l} e^{z_j^l} & i \neq j \end{cases} \tag{8}$$

3 Cost Function

Softmax cross-entropy:

$$\mathcal{J}(z_{ij}^l) = \frac{1}{B}\sum_{i=0}^{B}\left(\sum_j -y_{ij}\log(x_{ij}^L)\right) \tag{9}$$

$$\partial_{z_{ij}^{L-1}}\mathcal{J} = \frac{1}{B}\sum_{k=0}^{B}\left(\sum_m -\frac{y_{km}}{x_{km}^L}\frac{\partial x_{km}^L}{\partial z_{ij}^{L-1}}\right) \tag{10}$$

4 Backpropagation

$$x^l = \sigma_{l-1}(z^{l-1}) \tag{11}$$
$$z^l = a^l + b^l \tag{12}$$
$$a^l = f^l(x^l, w^l) \tag{13}$$

$$\frac{\partial x^L}{\partial w^{L-1}} = \frac{\partial S}{\partial z^{L-1}} \frac{\partial \mathcal{D}^{L-1}}{\partial w^{L-1}} \tag{14}$$

$$\frac{\partial x^L}{\partial b^{L-1}} = \frac{\partial S}{\partial z^{L-1}} \tag{15}$$

$$\frac{\partial x^L}{\partial w^{L-2}} = \frac{\partial S}{\partial z^{L-1}} \frac{\partial \mathcal{D}^{L-1}}{\partial x^{L-1}} \frac{\partial R}{\partial z^{L-2}} \frac{\partial \mathcal{D}^{L-2}}{\partial w^{L-2}} \tag{16}$$

$$\frac{\partial x^L}{\partial b^{L-1}} = \frac{\partial S}{\partial z^{L-1}} \frac{\partial \mathcal{D}^{L-1}}{\partial x^{L-1}} \frac{\partial R}{\partial z^{L-2}} \tag{17}$$