1 Layers

1. 2D Convolution:

$$C_{ijk}(\mathbf{x}, \mathbf{w}) = \sum_{p,q,r} x_{(s_1 i+p)(s_2 j+q)r}\, w_{(P-p)(Q-q)rk} \tag{1}$$

$$\frac{\partial C_{ijk}}{\partial x_{tuv}} = w_{(P+s_1 i-t)(Q+s_2 j-u)vk} \tag{2}$$

$$\frac{\partial C_{ijk}}{\partial w_{ctuv}} = x_{(s_1 i+P-c)(s_2 j+Q-t)u}\, \delta_{kv} \tag{3}$$

2. Max Pooling 2×2:

$$P_{ijk}(\mathbf{x}) = \max_{p,q} x_{(s_1 i+p)(s_2 j+q)k} \tag{4}$$

3. Flatten:

$$F_{(iJK+jK+k)}(\mathbf{x}) = x_{ijk} \tag{5}$$

4. Fully Connected (Dense):

$$D_j(\mathbf{x}, \mathbf{w}) = \sum_i x_i w_{ij} \tag{6}$$

$$\frac{\partial D_j}{\partial x_k} = w_{kj} \tag{7}$$

$$\frac{\partial D_j}{\partial w_{ck}} = x_c\, \delta_{jk} \tag{8}$$

2 Activation Functions

1. ReLU:

$$R_i(\mathbf{z}) = \begin{cases} z_i & z_i \geq 0 \\ 0 & z_i < 0 \end{cases} \tag{9}$$

$$\frac{\partial R_i}{\partial z_j} = \begin{cases} \delta_{ij} & z_i > 0 \\ \\ \frac{1}{2}\delta_{ij} & z_i = 0 \\ \\ 0 & z_i < 0 \end{cases} \tag{10}$$

2. Softmax:

$$S_i(\mathbf{z}) = \frac{e^{z_i}}{\sum\limits_k e^{z_k}} \tag{11}$$

$$\frac{\partial S_i}{\partial z_j} = \begin{cases} S_i(\mathbf{z})(1 - S_j(\mathbf{z})) & i = j \\ \\ -S_i(\mathbf{z})S_j(\mathbf{z}) & i \neq j \end{cases} \tag{12}$$

3 Cost Function

Softmax cross-entropy:

$$J(\mathbf{z}^{(L-1)b}) = \frac{1}{B}\sum_{b=0}^{B} J^b(\mathbf{z}^{(L-1)b}) \tag{13}$$

$$J^b(\mathbf{z}^{(L-1)b}) = \sum_i -y_i \log(z_i^{(L-1)b}) \tag{14}$$

$$
\begin{aligned}
\frac{\partial J^b}{\partial z_j^{(L-1)b}} &= \frac{\partial J^b}{\partial x_k^{Lb}} \frac{\partial S_k}{\partial z_j^{(L-1)b}} \\
&= -\frac{y_j}{x_j^{Lb}} S_j(\mathbf{z}^{(L-1)b})(1 - S_j(\mathbf{z}^{(L-1)b})) + \sum_{k \neq j} \frac{y_k}{x_k^{Lb}} S_k(\mathbf{z}^{(L-1)b}) S_j(\mathbf{z}^{(L-1)b}) \\
&= -\frac{y_j}{x_j^{Lb}} x_j^{Lb}(1 - x_j^{Lb}) + \sum_{k \neq j} \frac{y_k}{x_k^{Lb}} x_k^{Lb} x_j^{Lb} \\
&= -y_j(1 - x_j^{Lb}) + \sum_{k \neq j} y_k x_j^{Lb} \\
&= -y_j + x_j^{Lb} \sum_{kj} y_k \\
&= x_j^{Lb} - y_j
\end{aligned} \tag{15}
$$

4 Backpropagation

4.1 General

$$
\mathbf{x}^l = \sigma^{l-1}(\mathbf{z}^{l-1}) \tag{16}
$$

$$
\mathbf{z}^l = \mathbf{f}^l(\mathbf{x}^l, \mathbf{w}^l) + \mathbf{b}^l \tag{17}
$$

$$
\begin{aligned}
\Delta b_i^{lb} &= -\eta \frac{\partial J^b}{\partial b_i^{lb}} = -\eta \frac{\partial J^b}{\partial z_j^{lb}} \frac{\partial z_j^{lb}}{\partial b_i^{lb}} = -\eta \frac{\partial J^b}{\partial z_i^{lb}} \\
&= -\eta \frac{\partial J^b}{\partial x_{j_0}^{Lb}} \frac{\partial x_{j_0}^{Lb}}{\partial x_{j_1}^{(L-1)b}} \cdots \frac{\partial x_{j_{L-l-2}}^{(l+2)b}}{\partial x_{j_{L-l-1}}^{(l+1)b}} \frac{\partial x_{j_{L-l-1}}^{(l+1)b}}{\partial z_i^{lb}} \\
&= -\eta \frac{\partial J^b}{\partial x_{j_0}^{Lb}} \left(\frac{\partial x_{j_0}^{Lb}}{\partial z_{k_0}^{(L-1)b}} \frac{\partial z_{k_0}^{(L-1)b}}{\partial x_{j_1}^{(L-1)b}} \right) \cdots \left(\frac{\partial x_{j_{L-l-2}}^{(l+2)b}}{\partial z_{k_{L-l-2}}^{(l+1)b}} \frac{\partial z_{k_{L-l-2}}^{(l+1)b}}{\partial x_{j_{L-l-1}}^{(l+1)b}} \right) \frac{\partial x_{j_{L-l-1}}^{(l+1)b}}{\partial z_i^{lb}} \\
&= -\eta \frac{\partial J^b}{\partial x_{j_0}^{Lb}} \left(\frac{\partial \sigma_{j_0}^{(L-1)b}}{\partial z_{k_0}^{(L-1)b}} \frac{\partial f_{k_0}^{(L-1)b}}{\partial x_{j_1}^{(L-1)b}} \right) \cdots \left(\frac{\partial \sigma_{j_{L-l-2}}^{(l+1)b}}{\partial z_{k_{L-l-2}}^{(l+1)b}} \frac{\partial f_{k_{L-l-2}}^{(l+1)b}}{\partial x_{j_{L-l-1}}^{(l+1)b}} \right) \frac{\partial \sigma_{j_{L-l-1}}^{lb}}{\partial z_i^{lb}}
\end{aligned} \tag{18}
$$

$$\Delta b_i^{(L-1)b} = -\eta \frac{\partial J^b}{\partial x_j^{Lb}} \frac{\partial \sigma_j^{(L-1)b}}{\partial z_i^{(L-1)b}} \tag{19}$$

$$\Delta b_i^{(l-1)b} = \Delta b_j^{lb} \frac{\partial f_j^{lb}}{\partial x_k^{lb}} \frac{\partial \sigma_k^{(l-1)b}}{\partial z_i^{(l-1)b}} \tag{20}$$

$$\Delta w_{ij}^{lb} = -\eta \frac{\partial J^b}{\partial w_{ij}^{lb}} = -\eta \frac{\partial J^b}{\partial z_k^{lb}} \frac{\partial f_k^l}{\partial w_{ij}^{lb}} = \Delta b_k^{lb} \frac{\partial f_k^l}{\partial w_{ij}^{lb}} \tag{21}$$

4.2 This Case

$$L = 4 \tag{22}$$

$$\begin{aligned}
\sigma^0 &= \mathbf{P}(\mathbf{R}(\mathbf{z}^0)), & \mathbf{f}^0 &= \mathbf{C}(\mathbf{x}^0, \mathbf{w}^0) \\
\sigma^1 &= \mathbf{F}(\mathbf{P}(\mathbf{R}(\mathbf{z}^1))), & \mathbf{f}^1 &= \mathbf{C}(\mathbf{x}^1, \mathbf{w}^1) \\
\sigma^2 &= \mathbf{R}(\mathbf{z}^2), & \mathbf{f}^2 &= \mathbf{D}(\mathbf{x}^2, \mathbf{w}^2) \\
\sigma^3 &= \mathbf{S}(\mathbf{z}^3), & \mathbf{f}^3 &= \mathbf{D}(\mathbf{x}^3, \mathbf{w}^3)
\end{aligned} \tag{23}$$

$$\begin{aligned}
\frac{\partial \sigma_i^{3b}}{\partial z_j^{3b}} &= \frac{\partial S_i^b}{\partial z_j^{3b}}, & \frac{\partial f_i^{3b}}{\partial x_j^{3b}} &= \frac{\partial D_i^b}{\partial x_j^{3b}}, & \frac{\partial f_i^{3b}}{\partial w_{jk}^{3b}} &= \frac{\partial D_i^b}{\partial w_{jk}^{3b}} \\
\frac{\partial \sigma_i^{2b}}{\partial z_j^{2b}} &= \frac{\partial R_i^b}{\partial z_j^{2b}}, & \frac{\partial f_i^{2b}}{\partial x_j^{2b}} &= \frac{\partial D_i^b}{\partial x_j^{2b}}, & \frac{\partial f_i^{2b}}{\partial w_{jk}^{2b}} &= \frac{\partial D_i^b}{\partial w_{jk}^{2b}} \\
\frac{\partial}{\partial z_{pqr}^{1b}} P_{ijk}^{-1}(F^{-1}(\sigma^{1b})) &= \frac{\partial R_{ijk}^b}{\partial z_{pqr}^{1b}}, & \frac{\partial f_{ijk}^{1b}}{\partial x_{pqr}^{1b}} &= \frac{\partial C_{ijk}^b}{\partial x_{pqr}^{1b}}, & \frac{\partial f_{ijk}^{1b}}{\partial w_{pqrt}^{1b}} &= \frac{\partial C_{ijk}^b}{\partial w_{pqrt}^{1b}} \\
\frac{\partial}{\partial z_{pqr}^{0b}} P_{ijk}^{-1}(\sigma^{0b}) &= \frac{\partial R_{ijk}^b}{\partial z_{pqr}^{0b}}, & \frac{\partial f_{ijk}^{0b}}{\partial x_{pqr}^{0b}} &= \frac{\partial C_{ijk}^b}{\partial x_{pqr}^{0b}}, & \frac{\partial f_{ijk}^{0b}}{\partial w_{pqrt}^{0b}} &= \frac{\partial C_{ijk}^b}{\partial w_{pqrt}^{0b}}
\end{aligned} \tag{24}$$

$$
\begin{aligned}
\delta_i^{3b} &= -\eta \frac{\partial J^b}{\partial z_i^{3b}} & \Delta w_{ij}^{3b} &= \delta_k^{3b} \frac{\partial D_k^b}{\partial w_{ij}^{3b}} & \Delta b \\
\delta_i^{2b} &= \delta_j^{3b} \frac{\partial D_j^b}{\partial x_k^{3b}} \frac{\partial R_k^b}{\partial z_i^{2b}} & \Delta w_{ij}^{2b} &= \delta_k^{2b} \frac{\partial D_k^b}{\partial w_{ij}^{2b}} \\
\delta_{ijk}^{1b} &= P_{tuv}^{-1}(F^{-1}(\delta_p^{2b} \frac{\partial D_p^b}{\partial x^{2b}})) \frac{\partial R_{tuv}^b}{\partial z_{ijk}^{1b}} & \Delta w_{ijkm}^{1b} &= \delta_{pqr}^{1b} \frac{\partial C_{pqr}^b}{\partial w_{ijkm}^{1b}} \\
\delta_{ijk}^{0b} &= P_{tuv}^{-1}(\delta_{pqr}^{1b} \frac{\partial C_{pqr}^b}{\partial x^{2b}}) \frac{\partial R_{tuv}^b}{\partial z_{ijk}^{1b}} & \Delta w_{ijkm}^{0b} &= \delta_{pqr}^{0b} \frac{\partial C_{pqr}^b}{\partial w_{ijkm}^{0b}}
\end{aligned} \tag{25}
$$