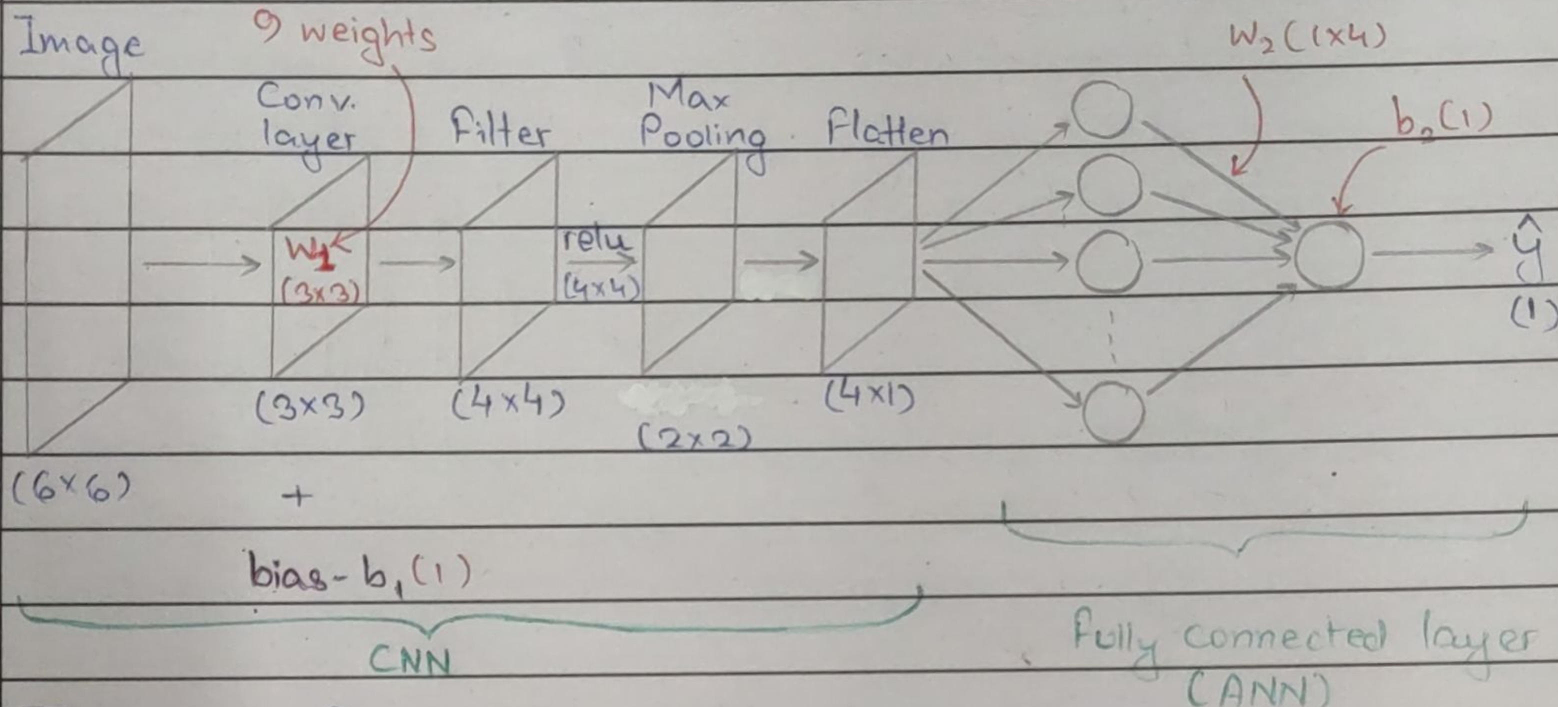
BACKPROPAGATION IN CNN



Trainable Parameters

$w_1 = (3\times3)$ $\quad$ $w_2 = (1\times4)$ $\quad \Rightarrow$ 15 trainable parameters

$b_1 = (1\times1)$ $\quad$ $b_2 = (1\times1)$

Loss → Binary Classification:

$$L = -y_i \log(\hat{y}_i) - (1-y_i)\log(1-\hat{y}_i)$$

Logical Flow

$$\cancel{X \longrightarrow \circledast \longrightarrow Z_1 \xrightarrow{relu} A_1 \xrightarrow[pool]{max} P_1 \xrightarrow{Flatten} f \xrightarrow{Sigmoid} Z_2}$$

$$\underbrace{X \longrightarrow \underset{\uparrow \atop w_1,\, b_1}{\circledast} \longrightarrow Z_1 \xrightarrow{relu} A_1 \xrightarrow[pool]{max} P_1 \xrightarrow{Flatten}}_{\text{CNN}} \underbrace{f \xrightarrow[b_2]{w_2} Z_2 \xrightarrow{Sigmoid} A_2 \longrightarrow L}_{\text{Fully connected (ANN)}}$$

RAISONI GROUP
— a vision beyond —

Forward Propagation

$Z_1 = Conv(X, w_1) + b_1$ $\quad$ $f = Flatten(P_1)$

$A_1 = relu(Z_1)$ $\quad$ $Z_2 = w_2 f + b_2$

$P_1 = MaxPool(A_1)$ $\quad$ $A_2 = \sigma(Z_2)$

Gradient Descent

$$\underbrace{w_1}_{\text{matrix}} = w_1 - \eta \underbrace{\boxed{\frac{\partial L}{\partial w_1}}}_{\text{matrix}} \qquad \underbrace{w_2}_{\text{matrix}} = w_2 - \eta \underbrace{\boxed{\frac{\partial L}{\partial w_2}}}_{\text{matrix}}$$

$$b_1 = b_1 - \eta \frac{\partial L}{\partial b_1} \qquad b_2 = b_2 - \eta \frac{\partial L}{\partial b_2}$$

- ANN Portion Backpropagation

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial w_2} \qquad Z_2 = w_2 f + b_2$$

$$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial b_2} \qquad A_2 = \sigma(Z_2)$$

$$\frac{\partial L}{\partial a_2} = \frac{\partial}{\partial a_2}\left[-y_i \log(a_2) - (1-y_i)\log(1-a_2)\right]$$

single image

$$= \frac{-y_i}{a_2} + \frac{(1-y_i)}{1-a_2} = \frac{-y_i(1-a_2) + a_2(1-y_i)}{a_2(1-a_2)}$$

$$\Rightarrow \frac{\partial L}{\partial a_2} = \frac{-y_i + \cancel{y_i a_2} + a_2 - \cancel{y_i a_2}}{a_2(1-a_2)} = \frac{a_2 - y_i}{a_2(1-a_2)} \quad \text{—— (1)}$$

$$\frac{\partial a_2}{\partial Z_2} = \sigma(Z_2)[1-\sigma(Z_2)] = a_2(1-a_2) \quad \text{—— (2)}$$

$$\frac{\partial Z_2}{\partial w_2} = \frac{\partial}{\partial w_2}[w_2 f + b_2] = f \quad \text{—— (3)}$$

$$\frac{\partial Z_2}{\partial b_2} = \frac{\partial}{\partial b_2}[w_2 f + b_2] = 1 \quad \text{—— (4)}$$

RAISONI GROUP
— a vision beyond —

$$\therefore \frac{\partial L}{\partial w_2} = ① \times ② \times ③ = \frac{a_2 - y_i}{a_2(1-a_2)} \times a_2(1-a_2) \times f = (a_2 - y_i) f$$

$$\frac{\partial L}{\partial b_2} = ① \times ② \times ④ = \frac{a_2 - y_i}{a_2(1-a_2)} \times a_2(1-a_2) \times 1 = a_2 - y_i$$

For multiple images,

$$\frac{\partial L}{\partial w_2} = \underset{(1\times1)}{(A} - \underset{(1\times1)}{Y)}\ \underset{(4\times1)}{F} = \underset{(1\times1)}{(A} - \underset{(1\times1)}{Y)}\ \underset{(1\times4)}{F^T}$$

$$\frac{\partial L}{\partial b_2} = (A - Y)$$

When passing one image at a time (SGD)

In case of suppose Mini-batch G.D. when passing m no. of images at a time -

$$\frac{\partial L}{\partial w_2} = \frac{1}{m} \times (A - Y) F^T$$

$$\frac{\partial L}{\partial b_2} = \frac{1}{m} \times (A - Y)$$

- CNN Portion Backpropagation

$$\frac{\partial L}{\partial w_1} = \underbrace{\boxed{\frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2}}}_{A_2 - Y} \times \frac{\partial Z_2}{\partial f} \times \frac{\partial f}{\partial P_1} \times \frac{\partial P_1}{\partial A_1} \times \frac{\partial A_1}{\partial Z_1} \times \frac{\partial Z_1}{\partial w_1}$$

$$\frac{\partial L}{\partial b_1} = \underbrace{\boxed{\frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2}}}_{A_2 - Y} \times \frac{\partial Z_2}{\partial f} \times \frac{\partial f}{\partial P_1} \times \frac{\partial P_1}{\partial A_1} \times \frac{\partial A_1}{\partial Z_1} \times \frac{\partial Z_1}{\partial b_1}$$

$$\frac{\partial Z_2}{\partial f} = \frac{\partial}{\partial f}\left[w_2 f + b_2 \right] = w_2$$

$$\Rightarrow \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial f} = \frac{\partial L}{\partial f} = (A_2 - Y) w_2$$

$\frac{\partial f}{\partial P_1} \rightarrow$ f - Flatten layer has no trainable parameters.

$f \rightarrow (4\times1)$ $\quad P_1 \rightarrow (2\times2)$

So, transforming Flatten layer's shape to that of Max Pooling layers' shape

$$\therefore \frac{\partial f}{\partial P_1} = .\text{reshape}(P_1.\text{shape})$$

$$\Rightarrow \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial f} \times \frac{\partial f}{\partial P_1} = \frac{\partial L}{\partial P_1} = (A_2 - Y)\, W_2 .\text{reshape}(P_1.\text{shape})$$

(2×2)

$\frac{\partial P_1}{\partial A_1} \longrightarrow$ P_1 - Max Pooling layer has no trainable parameters.

$P_1 \rightarrow (2\times 2)$ $A_1 \longrightarrow (4\times 4)$

Intuition: $\partial P_1 / \partial A_1$ refers to A_1 to check the positions of each maximum value in a pooling region. Then in those positions elements of P_1 are substituted & in other positions 0 is substituted as they don't contribute to the loss function calculation.

Eg: $P_1 = \begin{bmatrix} 4 & 8 \\ 12 & 16 \end{bmatrix}_{(2\times 2)}$ $A_1 = \begin{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} & \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \\ \begin{bmatrix} 9 & 10 \\ 11 & 12 \end{bmatrix} & \begin{bmatrix} 13 & 14 \\ 15 & 16 \end{bmatrix} \end{bmatrix}$

$$\Rightarrow \frac{\partial P_1}{\partial A_1} = \begin{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0 & 8 \end{bmatrix} \\ \begin{bmatrix} 0 & 0 \\ 0 & 12 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0 & 16 \end{bmatrix} \end{bmatrix}_{(4\times 4)}$$

Mathematically: $\frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial f} \times \frac{\partial f}{\partial P_1} \times \frac{\partial P_1}{\partial A_1} = \frac{\partial L}{\partial A_1} = \begin{cases} \left.\frac{\partial L}{\partial P_1}\right|_{xy}, & \text{if } A_{mn} \text{ is the max element} \\ 0, & \text{otherwise} \end{cases}$

$$\frac{\partial A_1}{\partial Z_1} = \begin{cases} 1 & \text{if } Z_{1_{xy}} > 0 \\ 0 & \text{if } Z_{1_{xy}} \leq 0 \end{cases}$$

Inverse of ReLU operation

ReLU differentiation
$f(x) = \max(0, x)$
$\Rightarrow f'(x) = \begin{cases} 0, & \text{if } \leq 0 \\ 1, & \text{if } x > 0 \end{cases}$

$$\Rightarrow \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial f} \times \frac{\partial f}{\partial P_1} \times \frac{\partial P_1}{\partial A_1} \times \frac{\partial A_1}{\partial Z_1} = \frac{\partial L}{\partial Z_1} \ (4\times4)$$

<u>Backpropagation on Convolution Layer</u>

Conv. Layer $\rightarrow$ Consists trainable parameters

For calculation simplicity assume input shape = (3×3) (2×2)

$$\Rightarrow \underset{(3\times3)}{X} \longrightarrow \circledast \longrightarrow Z_1$$

$\uparrow$ w_1, b_1 (2×2)

$Z_1 \hookrightarrow \frac{\partial L}{\partial Z_1} \Rightarrow (2\times2)$

$$\frac{\partial L}{\partial b_1} = \underset{(2\times2)}{\frac{\partial L}{\partial Z_1}} \times \frac{\partial Z_1}{\partial b_1}$$

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \quad w_1 = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \quad Z_1 = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}$$

$$\frac{\partial L}{\partial Z_1} = \begin{bmatrix} \partial L/\partial Z_{11} & \partial L/\partial Z_{12} \\ \partial L/\partial Z_{21} & \partial L/\partial Z_{22} \end{bmatrix}$$

$$Z_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22} + b_1$$

$$Z_{12} = w_{11}x_{12} + w_{12}x_{13} + w_{21}x_{22} + w_{22}x_{23} + b_1$$

$$Z_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{32} + b_1$$

$$Z_{22} = w_{11}x_{22} + w_{12}x_{23} + w_{21}x_{32} + w_{22}x_{33} + b_1$$

RAISONI GROUP
— a vision beyond —

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial Z_1} \times \frac{\partial Z_1}{\partial b_1} = \frac{\partial L}{\partial Z_{11}} \times \frac{\partial Z_{11}}{\partial b_1} + \frac{\partial L}{Z_{12}} \times \frac{\partial Z_{12}}{\partial b_1} + \frac{\partial L}{\partial Z_{21}} \times \frac{\partial Z_{21}}{\partial b_1} + \frac{\partial L}{\partial Z_{22}} \times \frac{\partial Z_{22}}{\partial b_1}$$

$$= \frac{\partial L}{\partial Z_{11}} + \frac{\partial L}{\partial Z_{12}} + \frac{\partial L}{\partial Z_{21}} + \frac{\partial L}{\partial Z_{22}}$$

$$\Rightarrow \frac{\partial L}{\partial b_1} = \text{sum}\left(\frac{\partial L}{\partial Z_1}\right)$$

$$\frac{\partial L}{\partial w_1} = \begin{bmatrix} \partial L/\partial w_{11} & \partial L/\partial w_{12} \\ \partial L/\partial w_{21} & \partial L/\partial w_{22} \end{bmatrix} \qquad \frac{\partial L}{\partial Z_1} = \begin{bmatrix} \partial L/\partial Z_{11} & \partial L/\partial Z_{12} \\ \partial L/\partial Z_{21} & \partial L/\partial Z_{22} \end{bmatrix}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial Z_1} \times \frac{\partial Z_1}{\partial w_1} \quad \frac{\partial L}{\partial Z_{11}} \times \frac{\partial Z_{11}}{\partial w_{11}} + \frac{\partial L}{\partial Z_{12}} \times \frac{\partial Z_{12}}{\partial w_{11}} + \frac{\partial L}{\partial Z_{21}} \times \frac{\partial Z_{21}}{\partial w_{11}} + \frac{\partial L}{\partial Z_{22}} \times \frac{\partial Z_{22}}{\partial w_{11}} = \frac{\partial L}{\partial w_{11}}$$

$$\frac{\partial L}{\partial w_{12}} = \frac{\partial L}{\partial Z_{11}} \times \frac{\partial Z_{11}}{\partial w_{12}} + \frac{\partial L}{\partial Z_{12}} \times \frac{\partial Z_{12}}{\partial w_{12}} + \frac{\partial L}{\partial Z_{21}} \times \frac{\partial Z_{21}}{\partial w_{12}} + \frac{\partial L}{\partial Z_{22}} \times \frac{\partial Z_{22}}{\partial w_{12}}$$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial L}{\partial Z_{11}} \times \frac{\partial Z_{11}}{\partial w_{21}} + \frac{\partial L}{\partial Z_{12}} \times \frac{\partial Z_{12}}{\partial w_{21}} + \frac{\partial L}{\partial Z_{21}} \times \frac{\partial Z_{21}}{\partial w_{21}} + \frac{\partial L}{\partial Z_{22}} \times \frac{\partial Z_{22}}{\partial w_{21}}$$

$$\frac{\partial L}{\partial w_{22}} = \frac{\partial L}{\partial Z_{11}} \times \frac{\partial Z_{11}}{\partial w_{22}} + \frac{\partial L}{\partial Z_{12}} \times \frac{\partial Z_{12}}{\partial w_{22}} + \frac{\partial L}{\partial Z_{21}} \times \frac{\partial Z_{21}}{\partial w_{22}} + \frac{\partial L}{\partial Z_{22}} \times \frac{\partial Z_{22}}{\partial w_{22}}$$

$$\Rightarrow \frac{\partial L}{\partial w_{11}} = \frac{\partial L}{\partial Z_{11}} x_{11} + \frac{\partial L}{\partial Z_{12}} x_{12} + \frac{\partial L}{\partial Z_{21}} x_{21} + \frac{\partial L}{\partial Z_{22}} x_{22}$$

$$\frac{\partial L}{\partial w_{12}} = \frac{\partial L}{\partial Z_{11}} x_{12} + \frac{\partial L}{\partial Z_{12}} x_{13} + \frac{\partial L}{\partial Z_{21}} x_{22} + \frac{\partial L}{\partial Z_{22}} x_{23}$$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial L}{\partial Z_{11}} x_{21} + \frac{\partial L}{\partial Z_{12}} x_{22} + \frac{\partial L}{\partial Z_{21}} x_{31} + \frac{\partial L}{\partial Z_{22}} x_{32}$$

$$\frac{\partial L}{\partial w_{22}} = \frac{\partial L}{\partial Z_{11}} x_{22} + \frac{\partial L}{\partial Z_{12}} x_{23} + \frac{\partial L}{\partial Z_{21}} x_{32} + \frac{\partial L}{\partial Z_{22}} x_{33}$$

$$\Rightarrow \frac{\partial L}{\partial w_1} = \text{Conv}\left(X, \frac{\partial L}{\partial Z_1}\right)$$

$$\therefore \frac{\partial L}{\partial W_2} = (A - Y) F^T$$

$$\frac{\partial L}{\partial b_2} = A - Y$$

$$\frac{\partial L}{\partial b_1} = \text{sum}\left(\frac{\partial L}{\partial Z_1}\right) = \frac{\partial L}{\partial Z_{11}} + \frac{\partial L}{\partial Z_{12}} + \frac{\partial L}{\partial Z_{21}} + \frac{\partial L}{\partial Z_{22}}$$

$$\frac{\partial L}{\partial Z_1} = \begin{bmatrix} \partial L/\partial Z_{11} & \partial L/\partial Z_{12} \\ \partial L/Z_{21} & \partial L/\partial Z_{22} \end{bmatrix}$$

$$\frac{\partial L}{\partial W_1} = \text{Conv}\left(X, \frac{\partial L}{\partial Z_1}\right)$$

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

$$= \begin{bmatrix} \partial L/\partial W_{11} & \partial L/\partial W_{12} \\ \partial L/\partial W_{21} & \partial L/\partial W_{22} \end{bmatrix}$$

$$\frac{\partial L}{\partial W_{11}} = \frac{\partial L}{\partial Z_{11}} x_{11} + \frac{\partial L}{\partial Z_{12}} x_{12} + \frac{\partial L}{\partial Z_{21}} x_{21} + \frac{\partial L}{\partial Z_{22}} x_{22}$$

$$\frac{\partial L}{\partial W_{12}} = \frac{\partial L}{\partial Z_{11}} x_{12} + \frac{\partial L}{\partial Z_{12}} x_{13} + \frac{\partial L}{\partial Z_{21}} x_{22} + \frac{\partial L}{\partial Z_{22}} x_{23}$$

$$\frac{\partial L}{\partial W_{21}} = \frac{\partial L}{\partial Z_{11}} x_{21} + \frac{\partial L}{\partial Z_{12}} x_{22} + \frac{\partial L}{\partial Z_{21}} x_{31} + \frac{\partial L}{\partial Z_{22}} x_{32}$$

$$\frac{\partial L}{\partial W_{22}} = \frac{\partial L}{\partial Z_{11}} x_{22} + \frac{\partial L}{\partial Z_{12}} x_{23} + \frac{\partial L}{\partial Z_{21}} x_{32} + \frac{\partial L}{\partial Z_{22}} x_{33}$$