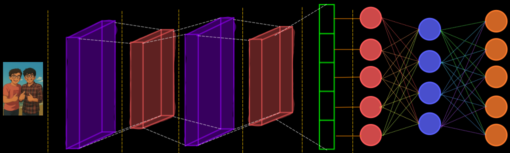
Basic CNN architecture



Most famous CNN Architectures.

1. LeNET
2. AlexNET
3. GoogleNET
4. VggNET
5. ResNET
6. Inception

1. LeNET (Yann lecun)

→ Structure

→ Input (32 × 32) → Structure/layer-1 → Structure/layer-2

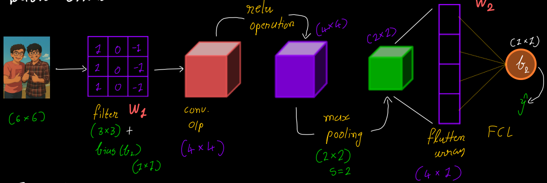
Structure/layer-1:
- Convolutional layer: kernel = (5×5), n_of_k = 6
- Pooling layer: filter = (2×2), n_of_f = 2

Structure/layer-2:
- Convolutional layer: kernel = (5×5), n_of_k = 16
- Pooling layer: filter = (2×2), n_of_f = 2

↳ FCL
- Layer-1 (120 neurons)
- Layer-2 (84 neurons)
- Layer-3 (o/p layer)

Backpropagation in CNN

Basic structure



Trainable parameters

↳ Total, we have mainly 2 locations, where we have some trainable parameters.

1. filter (3×3) = 9 parameters and 1 bias (1×1)

   $W_1 = (3\times3)$ $b_1 = (1\times1)$

   9   1

2. FCL

   4   1

   $W_2 = (1\times4)$, $b_2 = (1\times1)$

Total 15 trainable parameters.

Logical Structure of above CNN

$$x \longrightarrow \circledast \longrightarrow Z_1 \xrightarrow{\text{ReLU}} A_1 \xrightarrow[\text{pooling}]{\text{max}} P_1 \longrightarrow F \longrightarrow Z_2 \xrightarrow{\text{sigmoid}} A_2\,(\hat{y}) \longrightarrow L$$

- x: i/p image
- $\circledast$: convolution operation (W_1, b_1)
- Z_1: o/p of convolution operation
- A_1: o/p of ReLU
- P_1: o/p of max-pooling
- F: flatten array i/p of FCL
- Z_2: $(WX+b)$ (W_2, b_2)
- A_2: prediction
- L: loss calculation

Forward Propagation

$Z_1 = conv(x, W_1) + b_1$

$A_1 = ReLU(Z_1)$

$P_1 = MaxPool(A_1)$

$F = flatten(P_1)$

$Z_2 = \text{dot-product}(F, W_2) + b_2$

$A_2 / \hat{y} = sigmoid(Z_2)$

$$L = \frac{1}{m} \sum_{i=1}^{m} \left[-y_i \log(A_2) - (1-y_i) \log(1-A_2) \right]$$

Gradient descent

$W_1 = W_1 - \alpha \frac{\partial L}{\partial W_1}$ $\quad W_2 = W_2 - \alpha \frac{\partial L}{\partial W_2}$

$b_1 = b_1 - \alpha \frac{\partial L}{\partial b_1}$ $\quad b_2 = b_2 - \alpha \frac{\partial L}{\partial b_2}$

- here, since W_1 and W_2 are metrix, $\frac{\partial L}{\partial W_1}$ and $\frac{\partial L}{\partial W_2}$ will also be a matrix.
- b_1 and b_2 will be a single value.

$$x \longrightarrow \circledast \longrightarrow Z_1 \xrightarrow{ReLU} A_1 \xrightarrow[\text{pooling}]{\text{max}} P_1 \longrightarrow F \longrightarrow Z_2 \xrightarrow{sigmoid} \overset{(\hat{y})}{A_2} \longrightarrow L$$

x	$\circledast$	Z_1	A_1	P_1	F	Z_2	A_2	L
i/p image	convolution operation (W_1, b_1)	o/p of convolution operation	o/p of ReLU	o/p of max-pooling	flatten array i/p of FCL	$(WX + b)$ (W_2, b_2)	prediction	loss calculation

$$\frac{\partial L}{\partial W_2} = \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial W_2} \qquad \frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial b_2}$$

$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial F} \times \frac{\partial F}{\partial P_1} \times \frac{\partial P_1}{\partial A_1} \times \frac{\partial A_1}{\partial Z_1} \times \frac{\partial Z_1}{\partial W_1}$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial F} \times \frac{\partial F}{\partial P_1} \times \frac{\partial P_1}{\partial A_1} \times \frac{\partial A_1}{\partial Z_1} \times \frac{\partial Z_1}{\partial b_1}$$

So let's calculate derivative of ANN.

$Z_2 = W_2 F + b_2$

$A_2 = \sigma(Z_2)$

- let suppose we're currently working with single image, so for this time, consider a_2 instead of A_2, hence....

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

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

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

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

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

$$\frac{\partial A_2}{\partial Z_2} = \sigma(z_2)\left[1-\sigma(z_2)\right]$$

$$= a_2\left[1-a_2\right]$$

$$\frac{\partial Z_2}{\partial W_2} = F$$

$$\frac{\partial Z_2}{\partial b_2} = 1$$

$$\frac{\partial L}{\partial W_2} = \frac{(a_2 - y_i)}{a_2(1-a_2)} \times a_2(1-a_2) \times F$$

$$= (a_2 - y_i) \times F$$

$$= (A_2 - Y)F$$

$$(A_2 - Y)F$$

$$(1\times1)\quad(1\times1)\qquad(4\times1)^T$$

$$(1\times4)$$

$$(1\times1)\times(1\times4)$$

$$(1\times4)$$

hence...

$$\frac{\partial L}{\partial b_2} = \frac{(a_2 - y_i)}{a_2(1-a_2)}\, a_2(1-a_2)\, 1$$

$$= (A_2 - Y)$$

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

$$\boxed{\frac{\partial L}{\partial b_2} = (A_2 - Y)}$$

$$x \longrightarrow \circledast \longrightarrow Z_1 \xrightarrow{\text{ReLU}} A_1 \xrightarrow[\text{pooling}]{\text{max}} P_1 \longrightarrow F \longrightarrow Z_2 \xrightarrow{\text{sigmoid}} A_2\,(\hat{y}) \longrightarrow L$$

- x: I/P image
- $\circledast$: convolution operation (W_1, b_1)
- Z_1: o/p of convolution operation
- A_1: o/p of ReLU
- P_1: o/p of max-pooling
- F: flatten array i/p of FCL
- Z_2: $(WX+b)$ (W_2, b_2)
- A_2: prediction
- L: loss calculation

$$W_1 = W_1 - \alpha \frac{\partial L}{\partial W_1}, \quad b_1 = b_1 - \alpha \frac{\partial L}{\partial b_1}$$

$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial F} \times \frac{\partial F}{\partial P_1} \times \frac{\partial P_1}{\partial A_1} \times \frac{\partial A_1}{\partial Z_1} \times \frac{\partial Z_1}{\partial W_1}$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial F} \times \frac{\partial F}{\partial P_1} \times \frac{\partial P_1}{\partial A_1} \times \frac{\partial A_1}{\partial Z_1} \times \frac{\partial Z_1}{\partial b_1}$$

$Z_1 = \text{conv}(x, W_1) + b_1$

$A_1 = \text{ReLU}(Z_1)$

$P_1 = \text{MaxPool}(A_1)$

$F = \text{flatten}(P_1)$

$Z_2 = \text{dot-product}(F, W_2) + b_2$

$A_2 / \hat{y} = \text{sigmoid}(Z_2)$

$$\frac{\partial Z_2}{\partial F} = W_2$$

$$\frac{\partial F}{\partial P_1} = \text{No trainable parameters.}$$

$$\frac{\partial L}{\partial W_1} = \boxed{\frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial F} \times \frac{\partial F}{\partial P_1}} \times \frac{\partial P_1}{\partial A_1} \times \frac{\partial A_1}{\partial Z_1} \times \frac{\partial Z_1}{\partial W_1}$$

$\rightarrow (A_2 - Y) W_2 \cdot \text{reshape}(P_1.\text{shape})$

$$\frac{\partial L}{\partial b_1} = \boxed{\frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial F} \times \frac{\partial F}{\partial P_1}} \times \frac{\partial P_1}{\partial A_1} \times \frac{\partial A_1}{\partial Z_1} \times \frac{\partial Z_1}{\partial b_1}$$

- Now let's check, how we will do back-propagation in maxpooling layer.

→ We will do completely opposite of max pooling.

Exo.

$$A_1 \begin{bmatrix} 1 & 2 & 5 & 6 \\ 3 & 4 & 7 & 8 \\ 9 & 10 & 13 & 14 \\ 11 & 12 & 15 & 16 \end{bmatrix} \xrightarrow[\text{pooling}]{\text{max}} \begin{bmatrix} 4 & 8 \\ 12 & 16 \end{bmatrix}$$

→ We're taking only 1 value (max value) from each 2×2 window.

→ It means that remained 3 values of each frame are not contributing to calculate $\hat{y}$.

→ Now we need to regenerate a 4×4 matrix.

$$\frac{\partial L}{\partial P_1} = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \longrightarrow (4\times 4) \ldots ?$$

2×2

which will looks like this

→ During converting a 2×2 matrix back to a 4×4 we're ensuring the original positions of x_1, x_2, x_3 and x_4.

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & x_1 & 0 & x_2 \\ 0 & 0 & 0 & 0 \\ 0 & x_3 & 0 & x_4 \end{bmatrix}$$

4×4

$$\frac{\partial L}{\partial W_1} = \boxed{\frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial F} \times \frac{\partial F}{\partial P_2}} \times \frac{\partial P_2}{\partial A_1} \times \frac{\partial A_1}{\partial Z_1} \times \frac{\partial Z_1}{\partial W_1}$$

$$\frac{\partial L}{\partial b_1} = \overbrace{\underbrace{\boxed{\frac{\partial L}{\partial A_2} \times \frac{\partial A_2}{\partial Z_2} \times \frac{\partial Z_2}{\partial F} \times \frac{\partial F}{\partial P_2}}}_{\frac{\partial L}{\partial P_1}} \times \frac{\partial P_2}{\partial A_2}}^{\frac{\partial L}{\partial A_1}} \times \frac{\partial A_1}{\partial Z_1} \times \frac{\partial Z_1}{\partial b_1}$$

$$\frac{\partial L}{\partial A_1} = \begin{cases} \frac{\partial L}{\partial P_{1xy}} & \text{if } A_{mn} \text{ is maximum element.} \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial A_1}{\partial Z_1} = \begin{cases} 1 \rightarrow Z_{1xy} > 0 \\ 0 \rightarrow Z_{1xy} < 0 \end{cases}$$

- Now we need to calculate backpropagation of convolution operation.

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial Z_1} \times \frac{\partial Z_1}{\partial b_1}$$

Suppose our i/p shape is 3×3....

$$x = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \quad W_1 = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \quad Z_1 = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}$$

where....

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

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

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

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

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

Now we need to calculate $\frac{\partial Z_1}{\partial b_1}$. (differentiate Z_1 with respect to b_1)

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial Z_1} \times \frac{\partial Z_1}{\partial b_1} = \frac{\partial L}{\partial Z_{11}} \times \frac{\partial Z_{11}}{\partial b_1} + \frac{\partial L}{\partial Z_{12}} \times \frac{\partial Z_{12}}{\partial b_1} + \frac{\partial L}{\partial Z_{21}} \times \frac{\partial Z_{21}}{\partial b_1} + \frac{\partial L}{\partial Z_{22}} \times \frac{\partial Z_{2}}{\partial b_1}$$

Explanation: How individual Z_{ij} is changing with respect to b_1.

$$\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}$$

$$\frac{\partial Z_{11}}{\partial b_1} = \frac{\partial}{\partial b_1}\left[x_{11} w_{11} + x_{12} w_{12} + x_{21} w_{21} + x_{22} w_{22} + b_1\right]$$

$$= 1$$

$$\frac{\partial Z_{12}}{\partial b_1} = \frac{\partial}{\partial b_1}\left[x_{12} w_{11} + x_{13} + w_{12} + x_{22} w_{21} + x_{23} w_{22} + b_1\right]$$

$$= 1$$

$$\frac{\partial Z_{21}}{\partial b_1} = \frac{\partial}{\partial b_1}\left[x_{21} w_{11} + x_{22} + w_{12} + x_{31} w_{21} + x_{32} w_{22} + b_1\right]$$

$$= 1$$

$$\frac{\partial Z_{22}}{\partial b_1} = \frac{\partial}{\partial b_1}\left[x_{22} w_{11} + x_{23} + w_{12} + x_{32} w_{21} + x_{33} w_{22} + b_1\right]$$

$$= 1$$

hence

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

$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial Z_1} \times \frac{\partial Z_1}{\partial W_1}$$

$$\frac{\partial L}{\partial Z_1} = \begin{bmatrix} \frac{\partial L}{\partial Z_{11}} & \frac{\partial L}{\partial Z_{12}} \\ \frac{\partial L}{\partial Z_{21}} & \frac{\partial L}{\partial Z_{22}} \end{bmatrix} \qquad \frac{\partial L}{\partial W_1} = \begin{bmatrix} \frac{\partial L}{\partial W_{11}} & \frac{\partial L}{\partial W_{12}} \\ \frac{\partial L}{\partial W_{21}} & \frac{\partial L}{\partial W_{22}} \end{bmatrix}$$

Since....

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

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

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

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

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

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

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

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

Let's complete the equation with remained unknown ratives....

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

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

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

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

Hence....

$$\frac{\partial L}{\partial W_1} = \mathrm{conv}\left(x, \frac{\partial L}{\partial Z_1}\right)$$

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