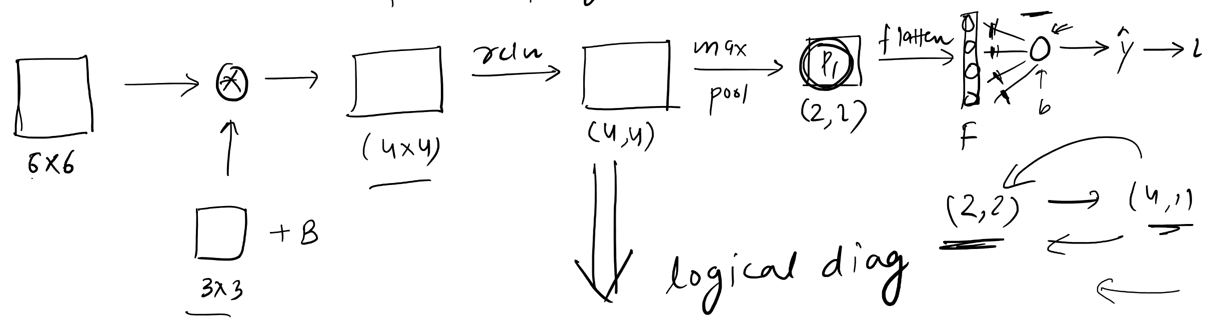
Lec-48 Backpropogation - II

Backpropogation in CNN

15 September 2022 09:37

1 conv + 1 pooling →



Forward Prop

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

$A_1 = relu(Z_1)$

$P_1 = maxpool(A_1)$

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

flatten

$Z_2 = W_2 F + b_2$

$A_2 = \sigma(Z_2)$

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

[_ _ _]

$(A_2 - y)$

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

5

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

1 2 3 4 6

reshape(P_1.shape)

6 derivatives

$\left[\frac{\partial Z_2}{\partial F}\right] = W_2 \rightarrow$

Shape ? → (F)

$conv\left(X, \frac{\partial L}{\partial Z_1}\right)$

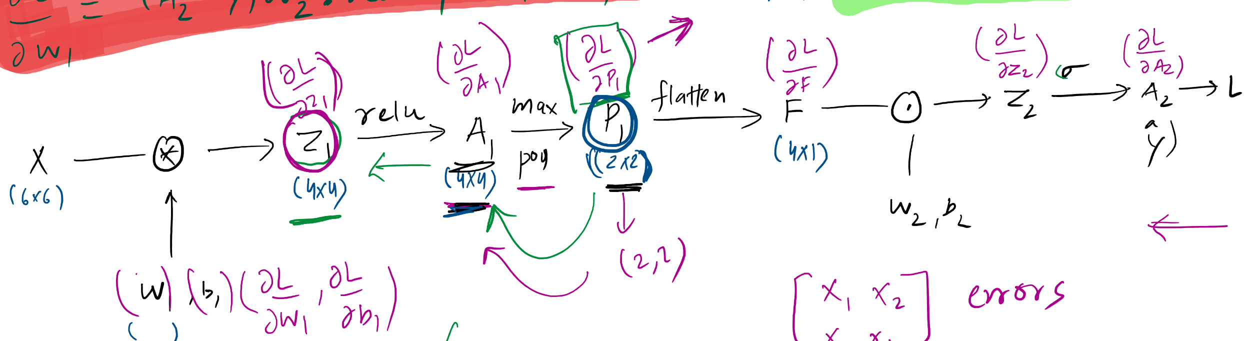
$\frac{\partial F}{\partial P_1}$ no trainable parameters

P_1 কে flatten করে F পেয়েছিলাম, back-propa. এর সময় P_1 এর সাপেক্ষে derivate করলে, F কে P_2 এ নিয়ে আসবো। (2,2) → (4,1)

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

এখন, অর্ধেক বের করেছি

$$\frac{\partial L}{\partial W_1} = (A_2 - y) W_2 \cdot reshape(P_1.shape)$$



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

$\frac{\partial L}{\partial A_1} = (4,4)$

$\frac{\partial L}{\partial A_1} = (4,4)$

In pooling like, Flatten → no trainable parameters

max pooling → A_1

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \longrightarrow \begin{bmatrix} 4 & 8 \\ 12 & 16 \end{bmatrix}$$

$$\begin{bmatrix} 9 & 10 \\ 11 & 12 \end{bmatrix} \begin{bmatrix} 13 & 14 \\ 15 & 16 \end{bmatrix}$$

(4×4)

$$\frac{\partial L}{\partial P_1} = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \quad 2\times 2$$

min value (1,2,3)

$\hat{y} \to 0$, ∂ / L

L ← **(simillarly)**

$\frac{\partial L}{\partial A_1}$

2×2 → (4×4) error

$$\frac{\partial L}{\partial A_1} \to \begin{bmatrix} 0 & x_1 & x_2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & x_3 & x_4 & 0 \end{bmatrix}$$

ŷ এর loss নির্ণয়ে min value গুলোর কোন contribution নেই। তাই max বাদে বাকী গুলো শূন্য দিয়ে দিব।

A, P

$$\frac{\partial L}{\partial W_2} = \frac{\partial L}{\partial A_2}\frac{\partial A_2}{\partial Z_2}\frac{\partial Z_2}{\partial F}\frac{\partial F}{\partial P_1}\frac{\partial P_1}{\partial A_1}\frac{\partial A_1}{\partial Z_1}\frac{\partial Z_1}{\partial W_1}$$

$(A_2 - Y)W_2.reshape(P_1.shape)$ → $\frac{\partial L}{\partial A_1}$

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

$\frac{\partial L}{\partial P_1}$ $\qquad$ $\frac{\partial L}{\partial A_1}$

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

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

relu $\max(0,x)$ $\quad \frac{dy}{dx}(x) = 0$ $\quad \frac{dy}{dx}(0) = 0$

Last step

$\left(\frac{\partial L}{\partial Z_1}\right)$

Convolution
↳ Backprop

max pooling
Flatten

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

Backprop on Convolution

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

assume input image (3×3)

$X \longrightarrow \circledast \longrightarrow Z_1$ $(2,2)$

$(3,3)$ $\qquad W_1, b_1$

$\left(\frac{\partial L}{\partial Z_1}\right) \to (2,2)$

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

W_1, b_1

$(2,2)$

Z_{21}

feature map এর সাথে bias থাকে। ২ টা filter থাকলে, feature map এর সাথে ২ টা constant bias add হবে।

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \circledast W_1 = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}$$

img $\qquad b_1 \qquad$ filter

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

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

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

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

$\frac{\partial A}{\partial z_1}$ → বের করা আছে।

$$\boxed{\frac{\partial L}{\partial b_1}} = \frac{\partial L}{\partial z_1} \times \frac{\partial z_1}{\partial b_1} = \boxed{\frac{\partial L}{\partial z_{11}} \times \frac{\partial z_{11}}{\partial b_1} + \frac{\partial L}{\partial z_{12}} \times \frac{\partial z_{12}}{\partial b_1} + \frac{\partial L}{\partial z_{21}} \times \frac{\partial z_{21}}{\partial b_1} + \frac{\partial L}{\partial z_{22}} \times \frac{\partial z_{22}}{\partial b_1}}$$

$$= \left(\frac{\partial L}{\partial z_{11}} + \frac{\partial L}{\partial z_{12}} + \frac{\partial L}{\partial z_{21}} + \frac{\partial L}{\partial z_{22}} \right) = sum\left(\frac{\partial L}{\partial z_1} \right)$$

$$\boxed{\frac{\partial L}{\partial b_1}} = sum\left(\frac{\partial L}{\partial z_1} \right) \to \text{scalar}$$

bias

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \qquad W_1 = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} + b$$

$X \to \circledast \to Z_1$

(3×3)

W_1, b_1

(2×2)

$\left(\frac{\partial L}{\partial z_1}\right)$

$$\frac{\partial L}{\partial W_1} = \begin{bmatrix} \frac{\partial L}{\partial w_{11}} & \frac{\partial L}{\partial w_{12}} \\ \frac{\partial L}{\partial w_{21}} & \frac{\partial L}{\partial w_{22}} \end{bmatrix} \qquad \frac{\partial L}{\partial z_1} = \begin{bmatrix} \frac{\partial L}{\partial z_{11}} & \frac{\partial L}{\partial z_{12}} \\ \frac{\partial L}{\partial z_{21}} & \frac{\partial L}{\partial z_{22}} \end{bmatrix}$$

$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial z_1} \times \frac{\partial z_1}{\partial W_1}$$

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

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

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

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

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

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

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

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

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

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \qquad \frac{\partial L}{\partial Z_1} = \begin{bmatrix} \frac{\partial L}{\partial z_{11}} & \frac{\partial L}{\partial z_{12}} \\ \frac{\partial L}{\partial z_{21}} & \frac{\partial L}{\partial z_{22}} \end{bmatrix}$$

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

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

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

Why use Pretrained models?

03 October 2022 12:52

1) Data hungry → image data → labelled data → cat dog → google

pinterest

labelling

10000 photos

manually

2) Time → model → training

Pretrained model