Back Prop in CNN

We know that output, O is

$$O = \text{convolution}(\underline{X}, \underline{W}) \quad \text{or} \quad \underline{X} * \underline{W}.$$

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

stride = 1

or

$$O_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$

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

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

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

Notice
(filter is not rotated for sake of simplicity)

Let us calculate the gradient with respect to filter:

$$\frac{\partial E}{\partial w_{11}} = \frac{\partial E}{\partial O_{11}}\frac{\partial O_{11}}{\partial w_{11}} + \frac{\partial E}{\partial O_{12}}\frac{\partial O_{12}}{\partial w_{11}} + \frac{\partial E}{\partial O_{21}}\frac{\partial O_{21}}{\partial w_{11}} + \frac{\partial E}{\partial O_{22}}\frac{\partial O_{22}}{\partial w_{11}}$$

$$\frac{\partial E}{\partial w_{12}} = \frac{\partial E}{\partial O_{11}}\frac{\partial O_{11}}{\partial w_{12}} + \frac{\partial E}{\partial O_{12}}\frac{\partial O_{12}}{\partial w_{12}} + \frac{\partial E}{\partial O_{21}}\frac{\partial O_{21}}{\partial w_{12}} + \frac{\partial E}{\partial O_{22}}\frac{\partial O_{22}}{\partial w_{22}}$$

$$\frac{\partial E}{\partial w_{21}} = \frac{\partial E}{\partial O_{11}}\frac{\partial O_{11}}{\partial w_{21}} + \frac{\partial E}{\partial O_{12}}\frac{\partial O_{12}}{\partial w_{21}} + \frac{\partial E}{\partial O_{21}}\frac{\partial O_{21}}{\partial w_{21}} + \frac{\partial E}{\partial O_{22}}\frac{\partial O_{22}}{\partial w_{21}}$$

$$\frac{\partial E}{\partial w_{22}} = \frac{\partial E}{\partial O_{11}}\frac{\partial O_{11}}{\partial w_{22}} + \frac{\partial E}{\partial O_{12}}\frac{\partial O_{12}}{\partial w_{22}} + \frac{\partial E}{\partial O_{21}}\frac{\partial O_{21}}{\partial O_{22}} + \frac{\partial E}{\partial O_{22}}\frac{\partial O_{22}}{\partial w_{22}}$$

this evaluates to

$$\frac{\partial E}{\partial W_{11}} = \frac{\partial E}{\partial O_{11}} \cdot x_{11} + \frac{\partial E}{\partial O_{12}} x_{12} + \frac{\partial E}{\partial O_{21}} x_{21} + \frac{\partial E}{\partial O_{22}} x_{22}$$

$$\frac{\partial E}{\partial W_{12}} = \;'' \; x_{12} + \;'' \; x_{13} \qquad x_{22} \qquad x_{23}$$

$$\frac{\partial E}{\partial W_{21}} = \quad x_{21} \qquad x_{22} \qquad x_{31} \qquad x_{32}$$

$$\frac{\partial E}{\partial W_{22}} = \quad x_{22} \qquad x_{23} \qquad x_{32} \qquad x_{33}$$

This in matrix form:

$$\begin{bmatrix} \frac{\partial E}{\partial W_{11}} & \frac{\partial E}{\partial W_{12}} \\ \frac{\partial E}{\partial W_{21}} & \frac{\partial E}{\partial W_{22}} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} * \begin{bmatrix} \frac{\partial E}{\partial O_{11}} & \frac{\partial E}{\partial O_{12}} \\ \frac{\partial E}{\partial O_{21}} & \frac{\partial E}{\partial O_{22}} \end{bmatrix}$$

If the matrix X is some hidden layer activation map, then the gradient of the error E w.r.t X has to be also computed.

In MLP:

$$E \rightarrow \hat{y} \xrightarrow{W} v_1 \xrightarrow{W} v_2 \xrightarrow{W} \cdots \xrightarrow{W} X.$$

Error $\quad$ o/p $\quad$ hidden layer activations so now looks like gradient at hidden activation map

$$\frac{\partial E}{\partial x_{11}} = \frac{\partial E}{\partial o_{11}} w_{11} + \frac{\partial E}{\partial o_{12}} 0 + \frac{\partial E}{\partial o_{21}} 0 + \frac{\partial E}{\partial o_{22}} 0$$

$$\frac{\partial E}{\partial x_{12}} = \frac{\partial E}{\partial o_{11}} w_{12} + \frac{\partial E}{\partial o_{12}} w_{11} + \frac{\partial E}{\partial o_{21}} 0 + \frac{\partial E}{\partial o_{22}} 0$$

$$\frac{\partial E}{\partial x_{13}} = \frac{\partial E}{\partial o_{11}} 0 + \frac{\partial E}{\partial o_{12}} w_{12} + \frac{\partial E}{\partial o_{21}} 0 + \frac{\partial E}{\partial o_{22}} 0$$

$$\frac{\partial E}{\partial x_{21}} = \quad w_{21} + \quad 0 \quad w_{11} \quad 0$$

$$\frac{\partial E}{\partial x_{22}} = \quad w_{22} \quad w_{21} \quad w_{12} \quad w_{11}$$

$$\frac{\partial E}{\partial x_{23}} = \quad 0 \quad w_{22} \quad 0 \quad w_{11}$$

$$\frac{\partial E}{\partial x_{31}} = \quad 0 \quad 0 \quad w_{21} \quad 0$$

$$\frac{\partial E}{\partial x_{32}} = \quad 0 \quad 0 \quad w_{22} \quad w_{21}$$

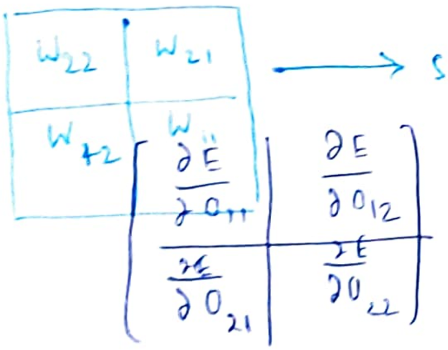
$$\frac{\partial E}{\partial x_{33}} = \quad 0 \quad 0 \quad 0 \quad w_{22}$$

we are calculating $\frac{\partial E}{\partial X}$ because we need to calculate the gradients of E w.r.t filter ~~w~~ in the previous layer of X.

$$\begin{bmatrix} \frac{\partial E}{\partial x_{11}} & \frac{\partial E}{\partial x_{12}} & \frac{\partial E}{\partial x_{13}} \\ \frac{\partial E}{\partial x_{21}} & \frac{\partial E}{\partial x_{22}} & \frac{\partial E}{\partial x_{23}} \\ \frac{\partial E}{\partial x_{31}} & \frac{\partial E}{\partial x_{32}} & \frac{\partial E}{\partial x_{33}} \end{bmatrix} = \text{full\_convolution}\left(\begin{bmatrix} \frac{\partial E}{\partial O_{11}} & \frac{\partial E}{\partial O_{12}} \\ \frac{\partial E}{\partial O_{21}} & \frac{\partial E}{\partial O_{22}} \end{bmatrix}, \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} \right)$$

$$\text{rotate}180^\circ\left(\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \right)$$

What is full_conv :



→ shift one cell at a time ①

↓

shift 1 cell at a time ②

$$\delta x_{11} = w_{11} \frac{\partial E}{\partial O_{11}}$$

Summary

$$\frac{\partial E}{\partial W} = \text{Conv}\left(\text{Input } X, \text{ loss Gradient } \frac{\partial E}{\partial O} \right)$$

$$\frac{\partial E}{\partial X} = \text{Full\_Conv}\left(\text{loss Gradient } \frac{\partial E}{\partial O}, \text{ Rotated } 180^\circ (W) \right)$$