Deep Learning

COMP 6721 Introduction of AI

Topics

- Gradient Descent in CNNs
- CNN Applications

Gradient Descent Recall

- There are multiple steps involved in performing (stochastic) gradient descent:

1. Compute the forward pass
2. Compute the derivative of the loss through a backward pass
3. Update the tunable parameters of the network

Problem Setup

- Consider a simple CNN, where our input, x, is 3x3 and kernel, w, is 2x2. No padding, stride = 1:

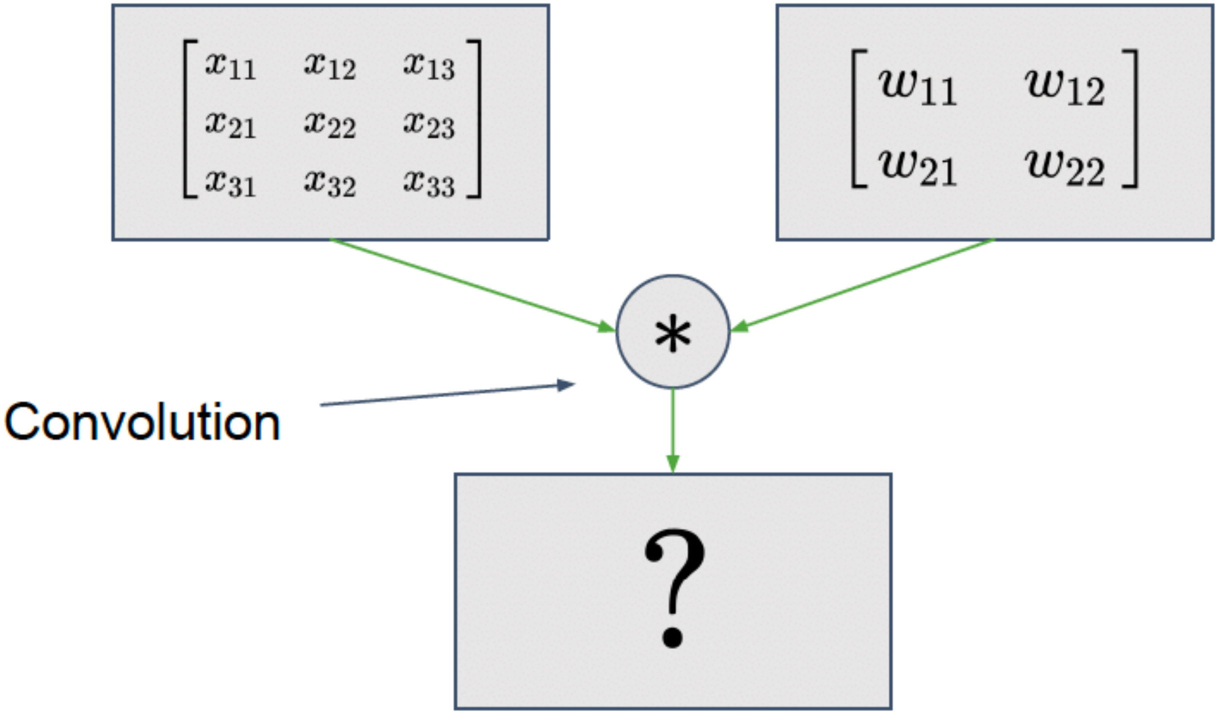
$$y = x * w$$

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

Convolution

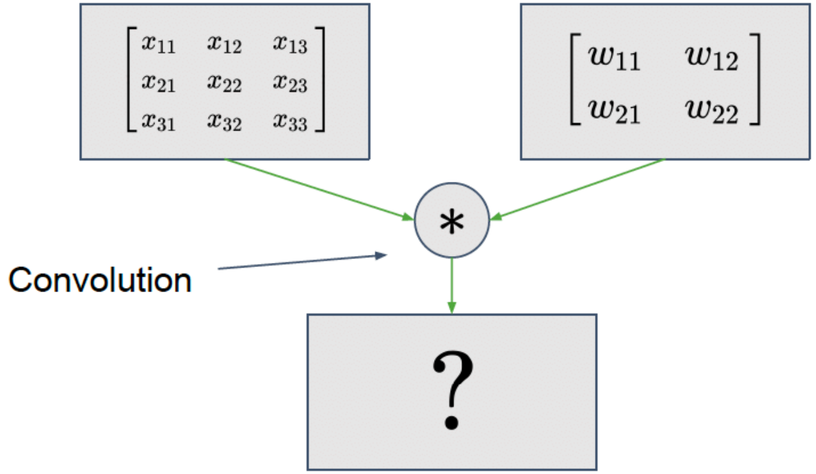
Forward Pass

- Graphically, this looks like this:



Forward Pass

- We can compute the forward pass, or the output, y:

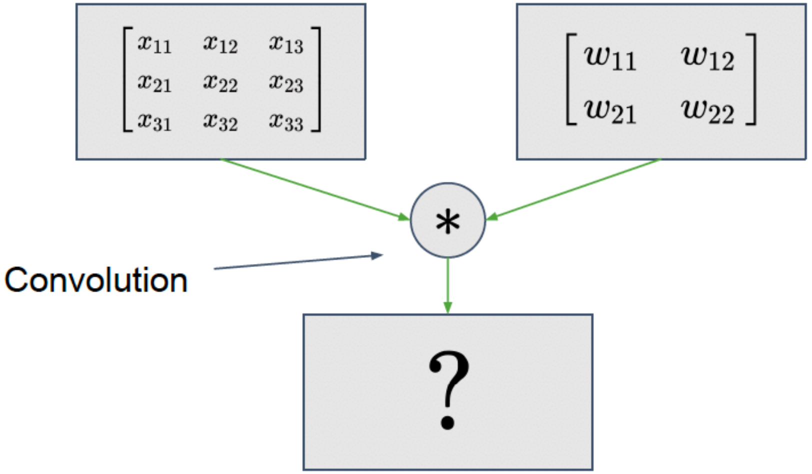


$$
\begin{aligned}
y &= x * w \\
&= \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} * \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}
\end{aligned}
$$

$$= \begin{bmatrix} \quad \end{bmatrix}$$

Forward Pass

- We can compute the forward pass, or the output, y:



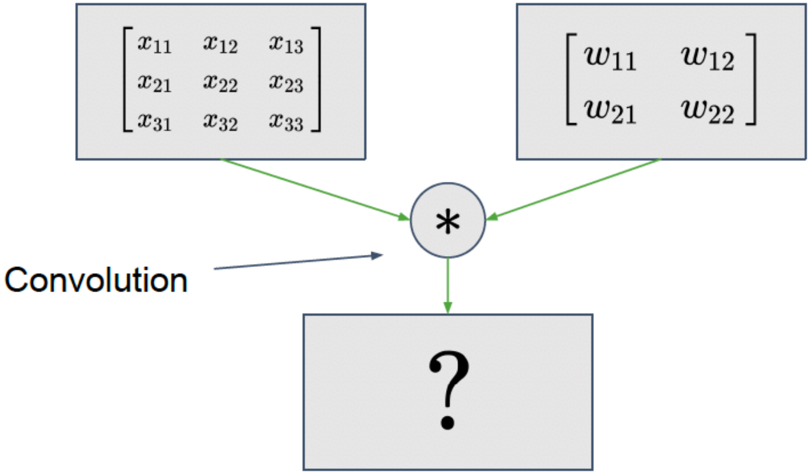
$$y = x * w$$

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

$$= \begin{bmatrix} \quad \end{bmatrix}$$

Forward Pass

- We can compute the forward pass, or the output, y:



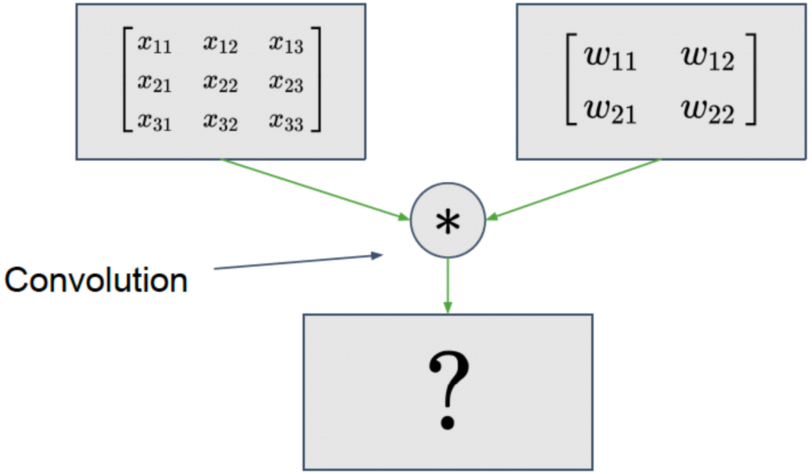
$$y = x * w$$

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

$$= \begin{bmatrix} \boxed{x_{11}w_{11}} + x_{12}w_{12} + x_{21}w_{21} + x_{22}w_{22}, \end{bmatrix}$$

Forward Pass

- We can compute the forward pass, or the output, y:



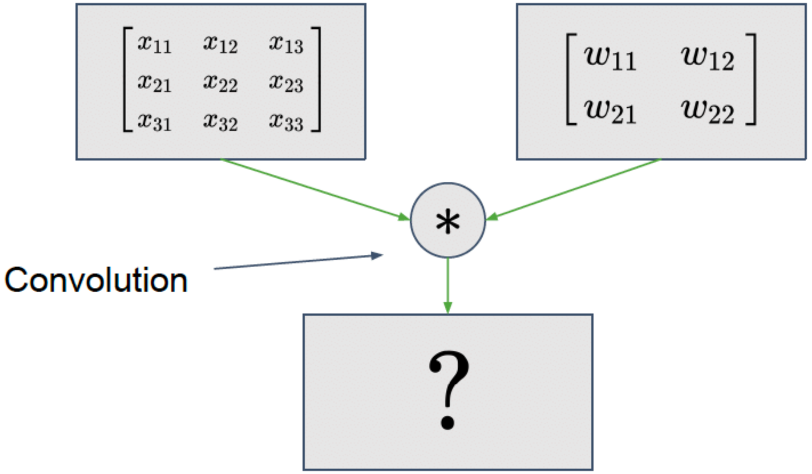
$$y = x * w$$

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

$$= \begin{bmatrix} x_{11}w_{11} + \boxed{x_{12}w_{12}} + x_{21}w_{21} + x_{22}w_{22}, & \\ & \end{bmatrix}$$

Forward Pass

- We can compute the forward pass, or the output, y:



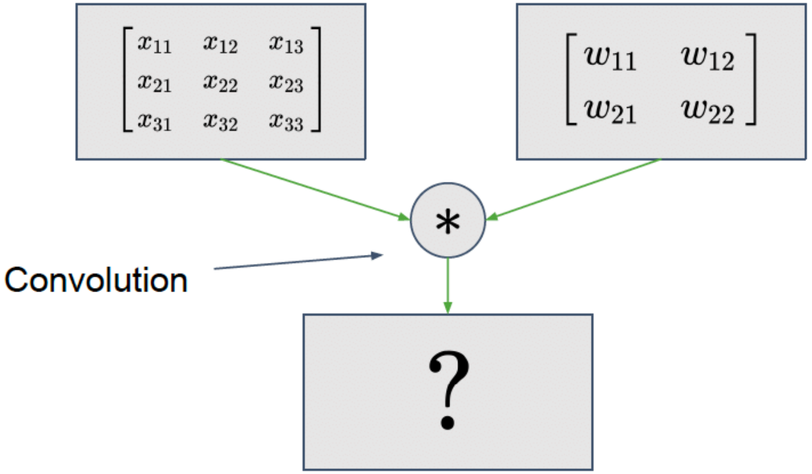
$$y = x * w$$

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

$$= \begin{bmatrix} x_{11}w_{11} + x_{12}w_{12} + \boxed{x_{21}w_{21}} + x_{22}w_{22}, & \end{bmatrix}$$

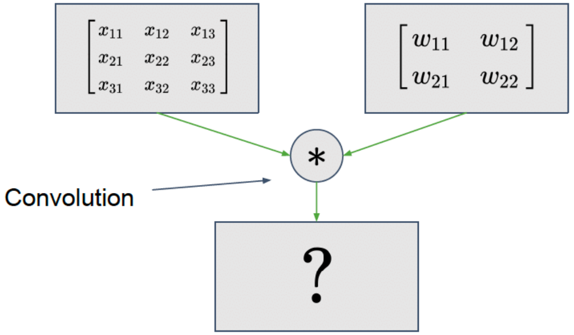
Forward Pass

- We can compute the forward pass, or the output, y:



$$
\begin{aligned}
y &= x * w \\
&= \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} * \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \\
&= \begin{bmatrix} x_{11}w_{11} + x_{12}w_{12} + x_{21}w_{21} + \boxed{x_{22}w_{22},} & \end{bmatrix}
\end{aligned}
$$

Forward Pass

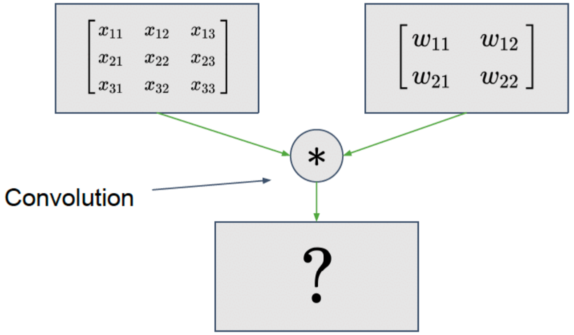


- We can compute the forward pass, or the output, y:

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

$$
= \begin{bmatrix} x_{11}w_{11} + x_{12}w_{12} + x_{21}w_{21} + x_{22}w_{22}, & x_{12}w_{11} + x_{13}w_{12} + x_{22}w_{21} + x_{23}w_{22} \\ & \end{bmatrix}
$$

Forward Pass

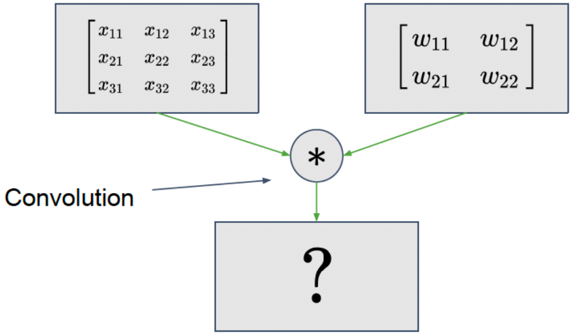


- We can compute the forward pass, or the output, y:

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

$$
= \begin{bmatrix} x_{11}w_{11} + x_{12}w_{12} + x_{21}w_{21} + x_{22}w_{22}, & x_{12}w_{11} + x_{13}w_{12} + x_{22}w_{21} + x_{23}w_{22} \\ x_{21}w_{11} + x_{22}w_{12} + x_{31}w_{21} + x_{32}w_{22}, & \end{bmatrix}
$$

Forward Pass

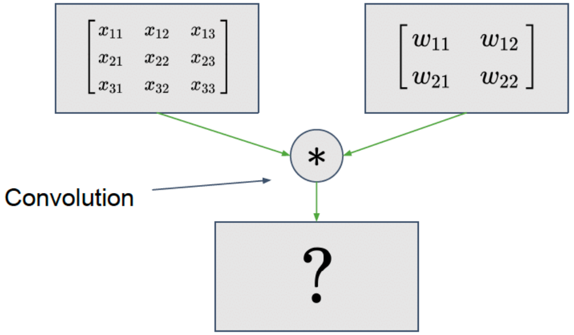


- We can compute the forward pass, or the output, y:

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

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

Forward Pass



- We can compute the forward pass, or the output, y:

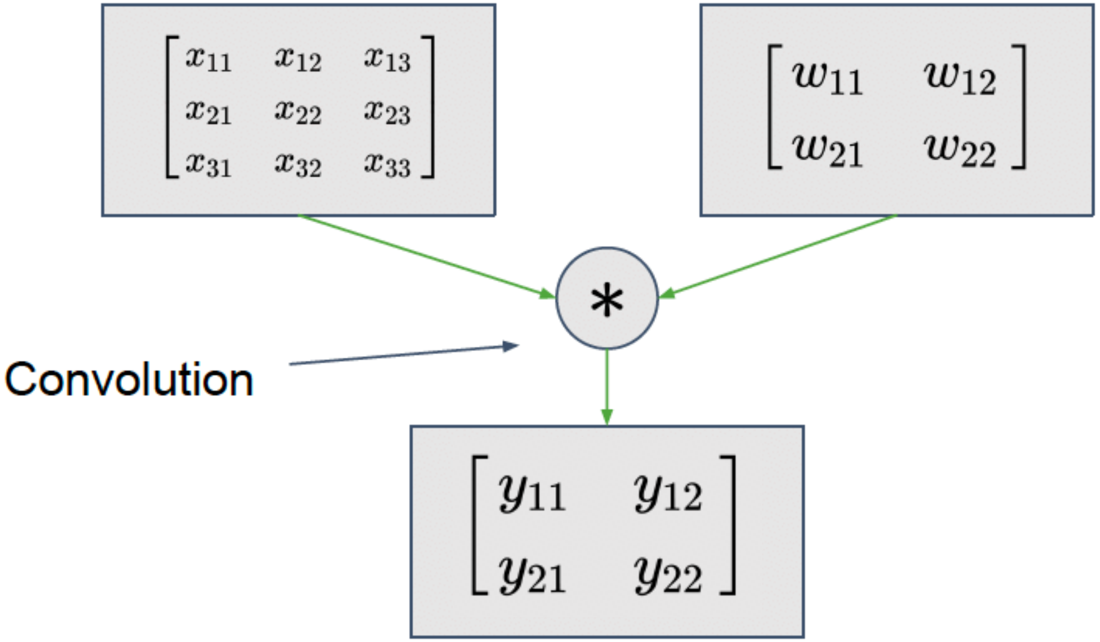
$$
\begin{aligned}
y &= x * w \\
&= \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} * \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}
\end{aligned}
$$

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

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

Forward Pass

- We can compute the forward pass, or the output, y:



Forward Pass

- Notice that this is equivalent to computing the following matrix multiplication: unroll the vector x_{ij} from left to right, top to bottom, construct W accordingly and unroll the resulting vector back to a 2x2 matrix. We will come back to this.

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

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

$$W = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}$$

$$W = \begin{bmatrix} w_{11} & 0 & 0 & 0 \\ w_{12} & w_{11} & 0 & 0 \\ 0 & w_{12} & 0 & 0 \\ w_{21} & 0 & w_{11} & 0 \\ w_{22} & w_{21} & w_{12} & w_{11} \\ 0 & w_{22} & 0 & w_{12} \\ 0 & 0 & w_{21} & 0 \\ 0 & 0 & w_{22} & w_{21} \\ 0 & 0 & 0 & w_{22} \end{bmatrix}$$

$$xW = \begin{bmatrix} y_{11} & y_{12} & y_{21} & y_{22} \end{bmatrix}$$

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

Loss

- Now let's define a loss function i.e. how we evaluate the error between the output we expect, ŷ, and the output we computed y. In this example, we use $L^{(2)}$ norm:

$$E = \frac{1}{2}\sum_{i,j}(y_{ij} - \hat{y}_{ij})^2 = \frac{1}{2}L^{(2)}$$

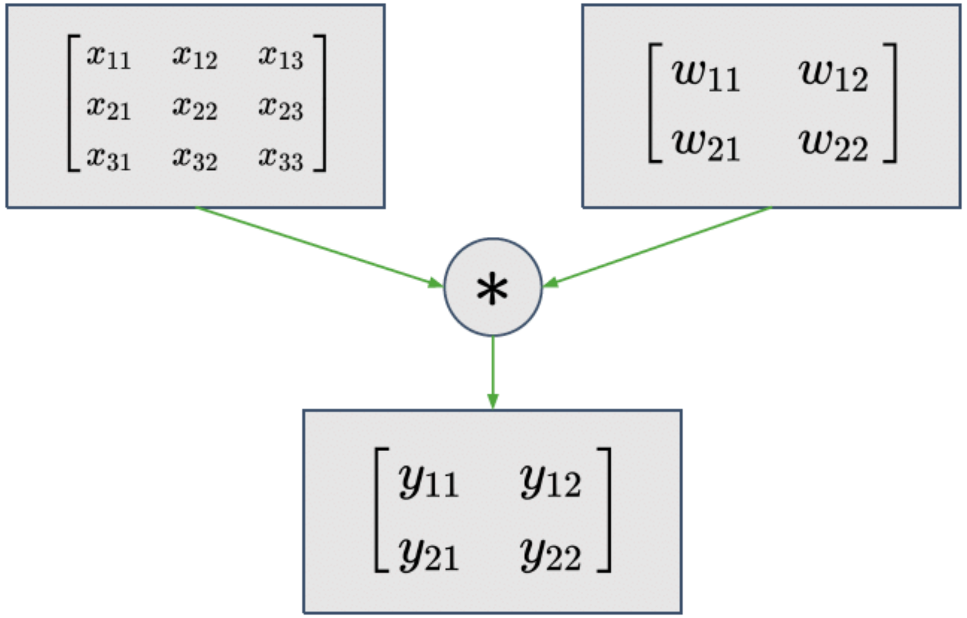
- Notice that the derivative w.r.t. each output can be expressed as:

$$\frac{\partial E}{\partial y_{ij}} = (y_{ij} - \hat{y}_{ij})$$

- There are many other types of loss functions that we could use, such as Binary Cross Entropy, etc.

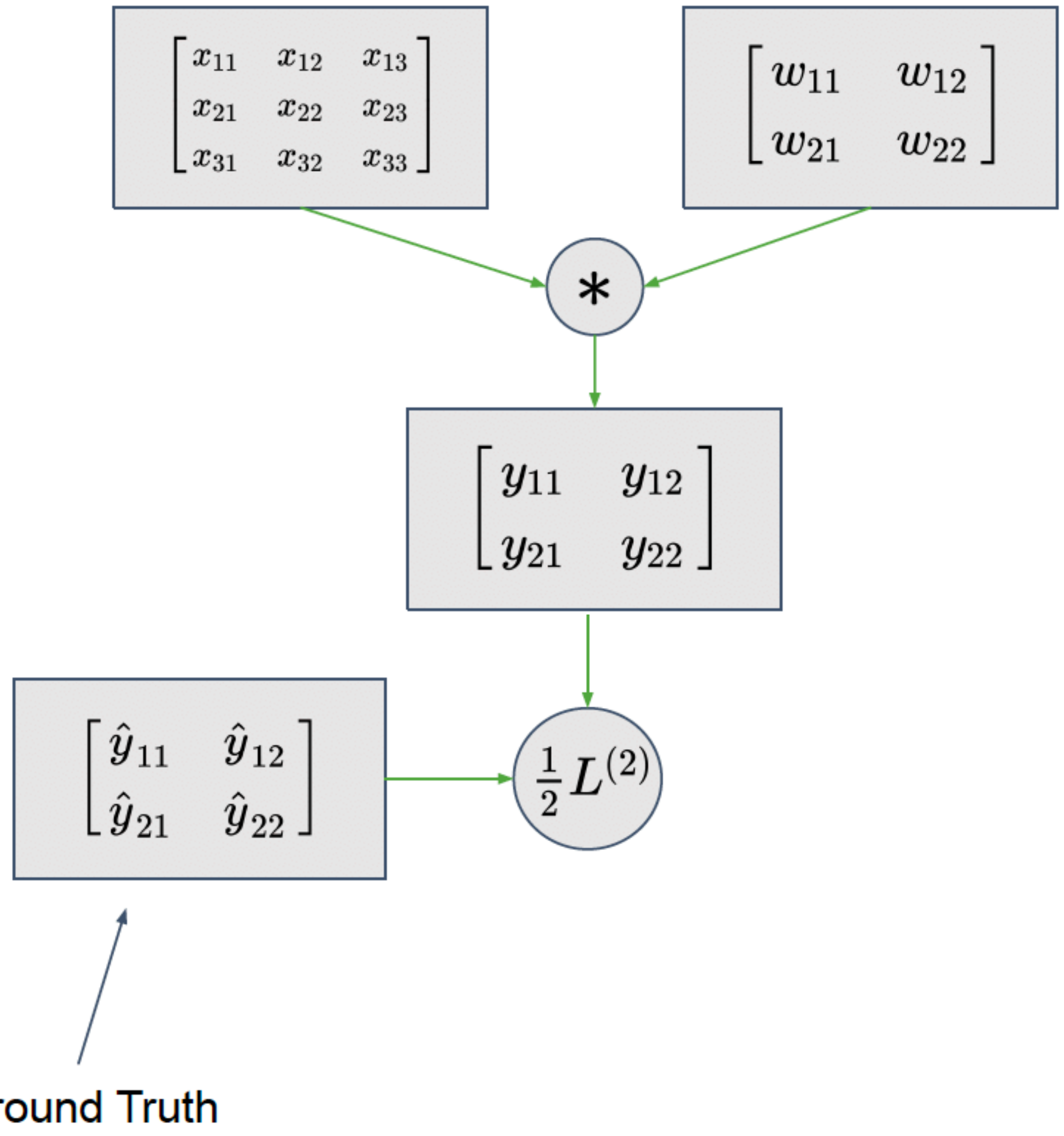
Loss

- We can continue building our computational graph:
- We compute the error between the ground truth and the output predicted by the network.



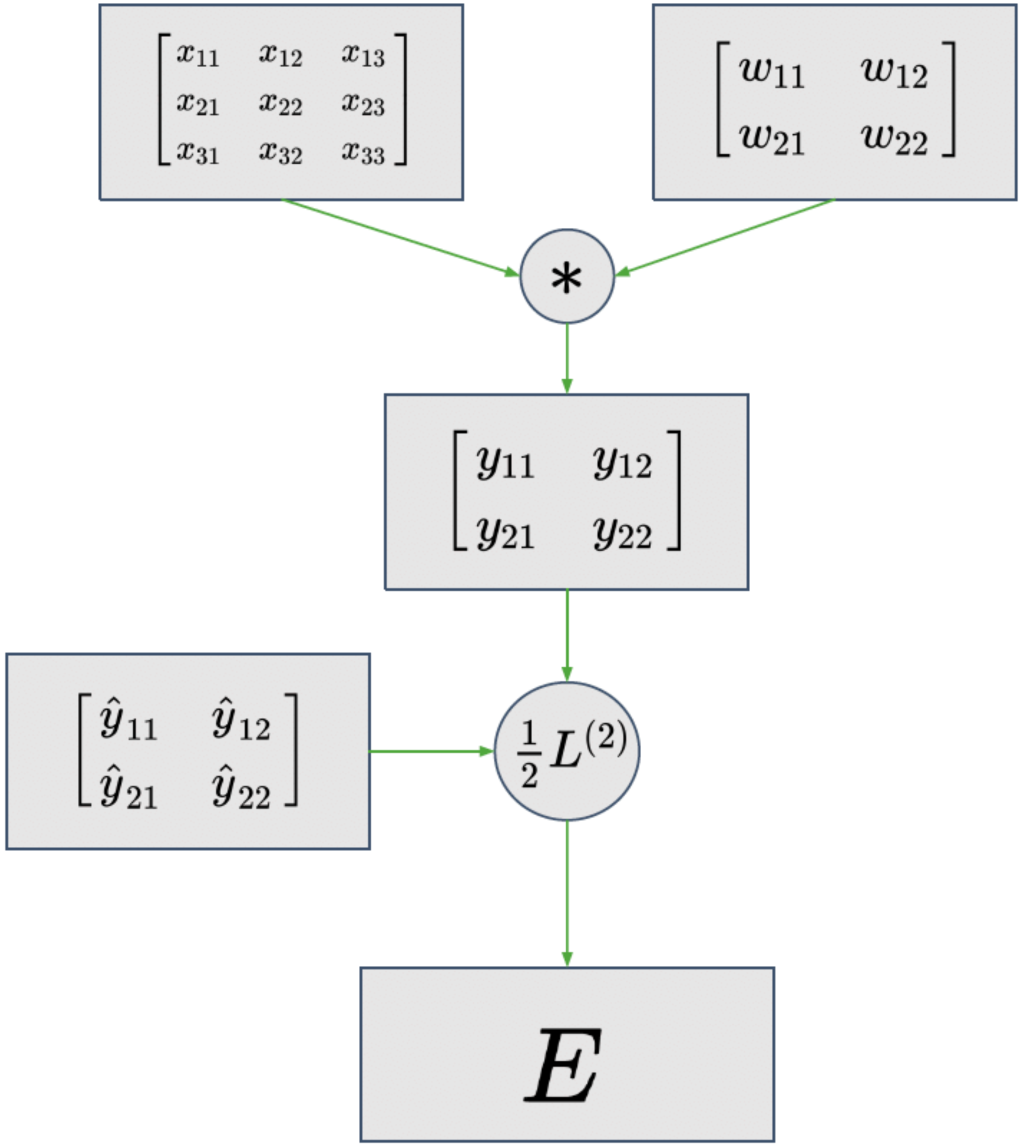
Loss

- We can continue building our computational graph:
- We compute the error between the ground truth and the output predicted by the network.



Gradient Descent Recall

- Recall that we are interested in calculating the update rule for the tunable parameters.

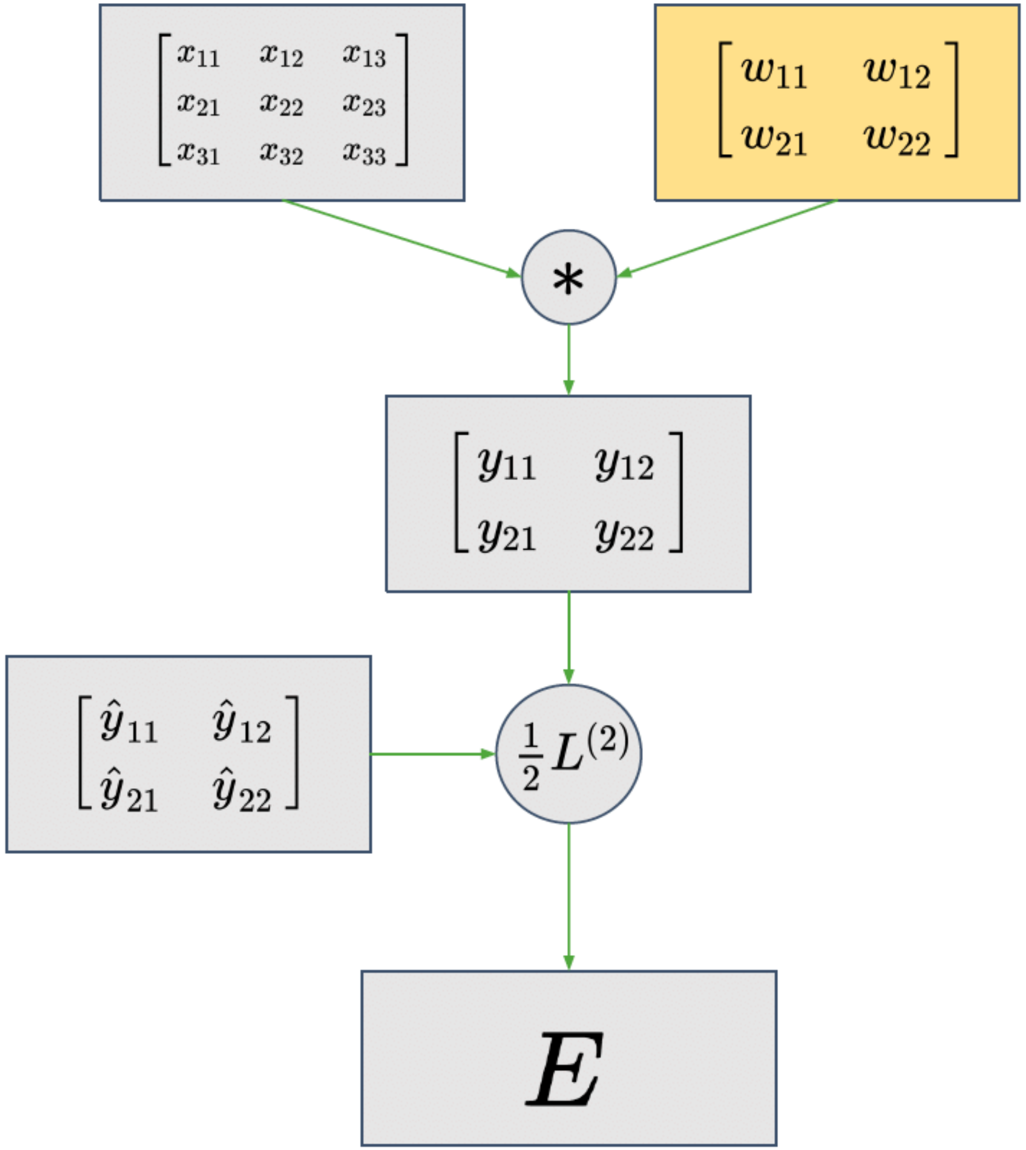


Gradient Descent Recall

- Recall that we are interested in calculating the **update rule** for the tunable parameters. The update rule for those parameters are:

$$\boxed{w_{ij}} \leftarrow \boxed{w_{ij}} - \eta \frac{\partial E}{\partial w_{ij}}$$

- In this case, the tunable parameters are in yellow.

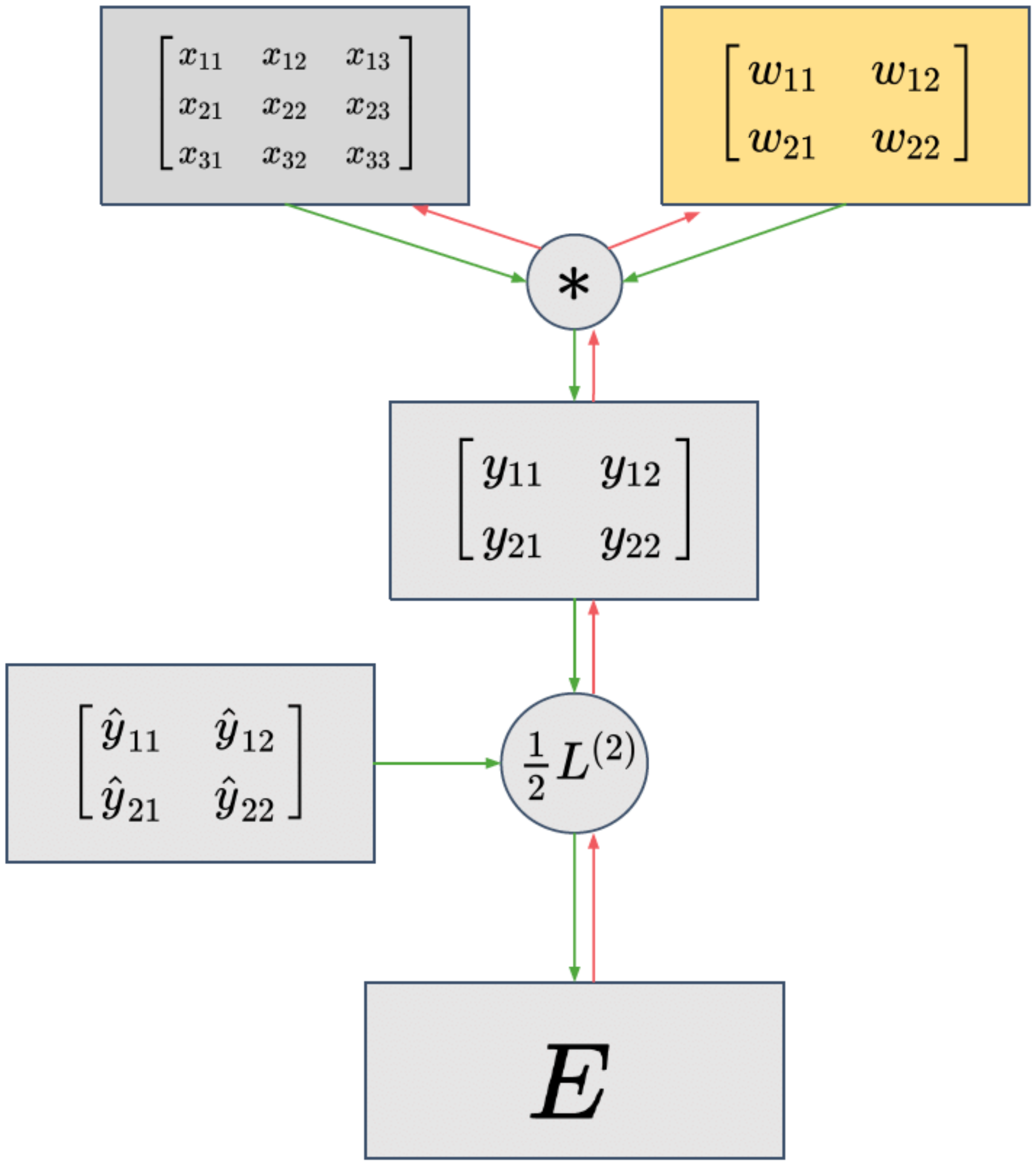


Gradient Descent Recall

- Recall that we are interested in calculating the **update rule** for the tunable parameters. The update rule for those parameters are:

$$\boxed{w_{ij}} \leftarrow \boxed{w_{ij}} - \eta \boxed{\frac{\partial E}{\partial w_{ij}}}$$

- In this case, the tunable parameters are in yellow. We must therefore calculate the derivative of the error with respect to all the w_{ij}.



Gradient

- Let's begin by calculating all the $\frac{\partial E}{\partial w_{ij}}$:
- Recall that our output, y, is:

$$y = x * w = \begin{bmatrix} x_{11}w_{11} + x_{12}w_{12} + x_{21}w_{21} + x_{22}w_{22}, & x_{12}w_{11} + x_{13}w_{12} + x_{22}w_{21} + x_{23}w_{22} \\ x_{21}w_{11} + x_{22}w_{12} + x_{31}w_{21} + x_{32}w_{22}, & x_{22}w_{11} + x_{23}w_{12} + x_{32}w_{21} + x_{33}w_{22} \end{bmatrix}$$

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

- Since each weight of the kernel affects each output and contributes to the **loss E**, we need to compute the **total derivative** over the output for each kernel weight w_{ij}:

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

Gradient

- In our example, we can expand all cases for all i,j:

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

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

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

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

Gradient

- In our example, we can expand all cases for all i,j. Now let's focus on i,j = 1,1

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

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

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

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

Gradient

- In our example,

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

- Recall that our output, y, is:

$$y = x * w = \begin{bmatrix} x_{11}w_{11} + x_{12}w_{12} + x_{21}w_{21} + x_{22}w_{22}, & x_{12}w_{11} + x_{13}w_{12} + x_{22}w_{21} + x_{23}w_{22} \\ x_{21}w_{11} + x_{22}w_{12} + x_{31}w_{21} + x_{32}w_{22}, & x_{22}w_{11} + x_{23}w_{12} + x_{32}w_{21} + x_{33}w_{22} \end{bmatrix}$$

Gradient

- In our example,

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

- Let's first compute:

$$\frac{\partial y_{11}}{\partial w_{11}} =$$

- Recall that our output, y, is:

$$y = x * w = \begin{bmatrix} x_{11}w_{11} + x_{12}w_{12} + x_{21}w_{21} + x_{22}w_{22}, & x_{12}w_{11} + x_{13}w_{12} + x_{22}w_{21} + x_{23}w_{22} \\ x_{21}w_{11} + x_{22}w_{12} + x_{31}w_{21} + x_{32}w_{22}, & x_{22}w_{11} + x_{23}w_{12} + x_{32}w_{21} + x_{33}w_{22} \end{bmatrix}$$

Gradient

- In our example,

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

- Let’s first compute:

$$\frac{\partial y_{11}}{\partial w_{11}} =$$

- Recall that our output, y, is:

$$y = x * w = \begin{bmatrix} \overbrace{\boxed{x_{11}w_{11} + x_{12}w_{12} + x_{21}w_{21} + x_{22}w_{22},}}^{y_{11}} & x_{12}w_{11} + x_{13}w_{12} + x_{22}w_{21} + x_{23}w_{22} \\ x_{21}w_{11} + x_{22}w_{12} + x_{31}w_{21} + x_{32}w_{22}, & x_{22}w_{11} + x_{23}w_{12} + x_{32}w_{21} + x_{33}w_{22} \end{bmatrix}$$

Gradient

- In our example,

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

- Let's first compute:

$$\frac{\partial y_{11}}{\partial w_{11}} = x_{11}$$

- Recall that our output, y, is:

$$y = x * w = \begin{bmatrix} \underbrace{\boxed{x_{11}w_{11} + x_{12}w_{12} + x_{21}w_{21} + x_{22}w_{22},}}_{y_{11}} & x_{12}w_{11} + x_{13}w_{12} + x_{22}w_{21} + x_{23}w_{22} \\ x_{21}w_{11} + x_{22}w_{12} + x_{31}w_{21} + x_{32}w_{22}, & x_{22}w_{11} + x_{23}w_{12} + x_{32}w_{21} + x_{33}w_{22} \end{bmatrix}$$

Gradient

- Expanding:

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

$$= \frac{\partial E}{\partial y_{11}}x_{11} + \frac{\partial E}{\partial y_{12}}x_{12} + \frac{\partial E}{\partial y_{21}}x_{21} + \frac{\partial E}{\partial y_{22}}x_{22}$$

- Recall that our output, y, is:

$$y = x * w = \begin{bmatrix} x_{11}w_{11} + x_{12}w_{12} + x_{21}w_{21} + x_{22}w_{22}, & x_{12}w_{11} + x_{13}w_{12} + x_{22}w_{21} + x_{23}w_{22} \\ x_{21}w_{11} + x_{22}w_{12} + x_{31}w_{21} + x_{32}w_{22}, & x_{22}w_{11} + x_{23}w_{12} + x_{32}w_{21} + x_{33}w_{22} \end{bmatrix}$$

Gradient

- We now have the derivative of the error with respect to the kernel weights

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

$$= \frac{\partial E}{\partial y_{11}} x_{11} + \frac{\partial E}{\partial y_{12}} x_{12} + \frac{\partial E}{\partial y_{21}} x_{21} + \frac{\partial E}{\partial y_{22}} x_{22}$$

These terms represent the error which is computed numerically $\frac{\partial E}{\partial y_{ij}} = (y_{ij} - \hat{y}_{ij})$

- Recall that our output, y, is:

$$y = x * w = \begin{bmatrix} x_{11}w_{11} + x_{12}w_{12} + x_{21}w_{21} + x_{22}w_{22}, & x_{12}w_{11} + x_{13}w_{12} + x_{22}w_{21} + x_{23}w_{22} \\ x_{21}w_{11} + x_{22}w_{12} + x_{31}w_{21} + x_{32}w_{22}, & x_{22}w_{11} + x_{23}w_{12} + x_{32}w_{21} + x_{33}w_{22} \end{bmatrix}$$

Gradient

- Expanding for every w_{ij}:

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

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

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

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

Gradient

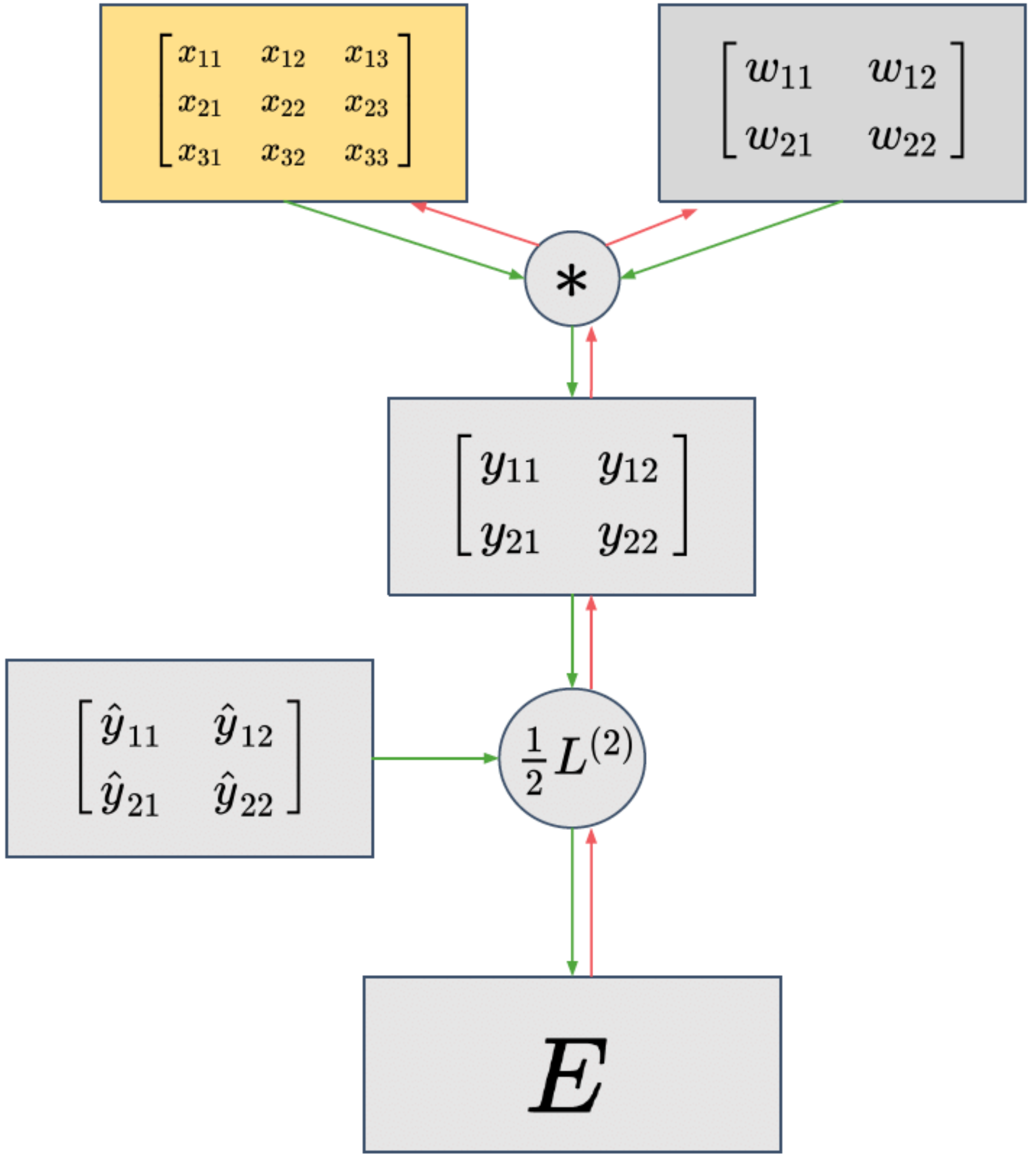
- … Which happens to be equivalent to:

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

Convolution

Gradient Descent Recall

- The inputs x_{ij} can also be the output from a previous layer. We therefore need to compute the derivative with respect to our inputs $\frac{\partial E}{\partial x_{ij}}$.



Gradient

- Now, we need to calculate all the $\frac{\partial E}{\partial x_{ij}}$
- We need to compute the derivative of the error with respect to the inputs x_{ij}. In general, for any pair (i,j) we have:

$$\frac{\partial E}{\partial x_{ij}} = \frac{\partial E}{\partial y_{11}} \frac{\partial y_{11}}{\partial x_{ij}} + \frac{\partial E}{\partial y_{12}} \frac{\partial y_{12}}{\partial x_{ij}} + \frac{\partial E}{\partial y_{21}} \frac{\partial y_{21}}{\partial x_{ij}} + \frac{\partial E}{\partial y_{22}} \frac{\partial y_{22}}{\partial x_{ij}}$$

Gradient

- Now, we need to calculate all the $\frac{\partial E}{\partial x_{ij}}$
- We need to compute the derivative of the error with respect to the inputs x_{ij}. In general, for any pair (i,j) we have:

$$\frac{\partial E}{\partial x_{ij}} = \frac{\partial E}{\partial y_{11}} \frac{\partial y_{11}}{\partial x_{ij}} + \frac{\partial E}{\partial y_{12}} \frac{\partial y_{12}}{\partial x_{ij}} + \frac{\partial E}{\partial y_{21}} \frac{\partial y_{21}}{\partial x_{ij}} + \frac{\partial E}{\partial y_{22}} \frac{\partial y_{22}}{\partial x_{ij}}$$

- Here expand for (i,j) = (1,1)

$$\frac{\partial E}{\partial x_{11}} = \frac{\partial E}{\partial y_{11}} \frac{\partial y_{11}}{\partial x_{11}} + \frac{\partial E}{\partial y_{12}} \frac{\partial y_{12}}{\partial x_{11}} + \frac{\partial E}{\partial y_{21}} \frac{\partial y_{21}}{\partial x_{11}} + \frac{\partial E}{\partial y_{22}} \frac{\partial y_{22}}{\partial x_{11}}$$

Gradient

- For (i,j) = (1,1)

$$\frac{\partial E}{\partial x_{11}} = \frac{\partial E}{\partial y_{11}}\frac{\partial y_{11}}{\partial x_{11}} + \frac{\partial E}{\partial y_{12}}\frac{\partial y_{12}}{\partial x_{11}} + \frac{\partial E}{\partial y_{21}}\frac{\partial y_{21}}{\partial x_{11}} + \frac{\partial E}{\partial y_{22}}\frac{\partial y_{22}}{\partial x_{11}}$$

Gradient

- For (i,j) = (1,1)

$$\frac{\partial E}{\partial x_{11}} = \frac{\partial E}{\partial y_{11}} \boxed{\frac{\partial y_{11}}{\partial x_{11}}} + \frac{\partial E}{\partial y_{12}} \frac{\partial y_{12}}{\partial x_{11}} + \frac{\partial E}{\partial y_{21}} \frac{\partial y_{21}}{\partial x_{11}} + \frac{\partial E}{\partial y_{22}} \frac{\partial y_{22}}{\partial x_{11}}$$

- Let's first compute:

$$\frac{\partial y_{11}}{\partial x_{11}} =$$

Gradient

- For (i,j) = (1,1)

$$\frac{\partial E}{\partial x_{11}} = \frac{\partial E}{\partial y_{11}} \boxed{\frac{\partial y_{11}}{\partial x_{11}}} + \frac{\partial E}{\partial y_{12}} \frac{\partial y_{12}}{\partial x_{11}} + \frac{\partial E}{\partial y_{21}} \frac{\partial y_{21}}{\partial x_{11}} + \frac{\partial E}{\partial y_{22}} \frac{\partial y_{22}}{\partial x_{11}}$$

- Let's first compute:

$$\frac{\partial y_{11}}{\partial x_{11}} =$$

Recall that our output, *y*, is:

$$y = x * w = \begin{bmatrix} \underbrace{\boxed{x_{11}w_{11} + x_{12}w_{12} + x_{21}w_{21} + x_{22}w_{22},}}_{y_{11}} & x_{12}w_{11} + x_{13}w_{12} + x_{22}w_{21} + x_{23}w_{22} \\ x_{21}w_{11} + x_{22}w_{12} + x_{31}w_{21} + x_{32}w_{22}, & x_{22}w_{11} + x_{23}w_{12} + x_{32}w_{21} + x_{33}w_{22} \end{bmatrix}$$

Gradient

- For (i,j) = (1,1)

$$\frac{\partial E}{\partial x_{11}} = \frac{\partial E}{\partial y_{11}} \boxed{\frac{\partial y_{11}}{\partial x_{11}}} + \frac{\partial E}{\partial y_{12}} \frac{\partial y_{12}}{\partial x_{11}} + \frac{\partial E}{\partial y_{21}} \frac{\partial y_{21}}{\partial x_{11}} + \frac{\partial E}{\partial y_{22}} \frac{\partial y_{22}}{\partial x_{11}}$$

- Let's first compute:

$$\frac{\partial y_{11}}{\partial x_{11}} = w_{11}$$

Recall that our output, *y*, is:

$$y = x * w = \begin{bmatrix} \overbrace{\boxed{x_{11}w_{11} + x_{12}w_{12} + x_{21}w_{21} + x_{22}w_{22},}}^{y_{11}} & x_{12}w_{11} + x_{13}w_{12} + x_{22}w_{21} + x_{23}w_{22} \\ x_{21}w_{11} + x_{22}w_{12} + x_{31}w_{21} + x_{32}w_{22}, & x_{22}w_{11} + x_{23}w_{12} + x_{32}w_{21} + x_{33}w_{22} \end{bmatrix}$$

Gradient

- Expanding for all (i,j), we get:
- Notice the **symmetry**? This is equivalent to calculating and reshaping the following matrix multiplication:

$$\partial Ey = \begin{bmatrix} \frac{\partial E}{\partial y_{11}} & \frac{\partial E}{\partial y_{12}} & \frac{\partial E}{\partial y_{21}} & \frac{\partial E}{\partial y_{22}} \end{bmatrix}$$

$$W^T = \begin{bmatrix} w_{11} & w_{12} & 0 & w_{21} & w_{22} & 0 & 0 & 0 & 0 \\ 0 & w_{11} & w_{12} & 0 & w_{21} & w_{22} & 0 & 0 & 0 \\ 0 & 0 & 0 & w_{11} & w_{12} & 0 & w_{21} & w_{22} & 0 \\ 0 & 0 & 0 & 0 & w_{11} & w_{12} & 0 & w_{21} & w_{22} \end{bmatrix}$$

$$\partial Ey \times W^T =$$

$$\frac{\partial E}{\partial x_{11}} = \frac{\partial E}{\partial y_{11}} w_{11}$$

$$\frac{\partial E}{\partial x_{12}} = \frac{\partial E}{\partial y_{11}} w_{12} + \frac{\partial E}{\partial y_{12}} w_{11}$$

$$\frac{\partial E}{\partial x_{13}} = \frac{\partial E}{\partial y_{12}} w_{12}$$

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

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

$$\frac{\partial E}{\partial x_{23}} = \frac{\partial E}{\partial y_{12}} w_{22} + \frac{\partial E}{\partial y_{22}} w_{12}$$

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

$$\frac{\partial E}{\partial x_{32}} = \frac{\partial E}{\partial y_{21}} w_{22} + \frac{\partial E}{\partial y_{22}} w_{21}$$

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

Gradient

- We are using the same **kernel matrix** we used to compute the forward pass. This is called "**Transposed Convolution**"

$$\partial Ey = \begin{bmatrix} \frac{\partial E}{\partial y_{11}} & \frac{\partial E}{\partial y_{12}} & \frac{\partial E}{\partial y_{21}} & \frac{\partial E}{\partial y_{22}} \end{bmatrix}$$

$$W^T = \begin{bmatrix} w_{11} & w_{12} & 0 & w_{21} & w_{22} & 0 & 0 & 0 & 0 \\ 0 & w_{11} & w_{12} & 0 & w_{21} & w_{22} & 0 & 0 & 0 \\ 0 & 0 & 0 & w_{11} & w_{12} & 0 & w_{21} & w_{22} & 0 \\ 0 & 0 & 0 & 0 & w_{11} & w_{12} & 0 & w_{21} & w_{22} \end{bmatrix}$$

$$\partial Ey \times W^T =$$

$$\frac{\partial E}{\partial x_{11}} = \frac{\partial E}{\partial y_{11}} w_{11}$$
$$\frac{\partial E}{\partial x_{12}} = \frac{\partial E}{\partial y_{11}} w_{12} + \frac{\partial E}{\partial y_{12}} w_{11}$$
$$\frac{\partial E}{\partial x_{13}} = \frac{\partial E}{\partial y_{12}} w_{12}$$
$$\frac{\partial E}{\partial x_{21}} = \frac{\partial E}{\partial y_{11}} w_{21} + \frac{\partial E}{\partial y_{21}} w_{11}$$
$$\frac{\partial E}{\partial x_{22}} = \frac{\partial E}{\partial y_{11}} w_{22} + \frac{\partial E}{\partial y_{12}} w_{21} + \frac{\partial E}{\partial y_{21}} w_{12} + \frac{\partial E}{\partial y_{22}} w_{11}$$
$$\frac{\partial E}{\partial x_{23}} = \frac{\partial E}{\partial y_{12}} w_{22} + \frac{\partial E}{\partial y_{22}} w_{12}$$
$$\frac{\partial E}{\partial x_{31}} = \frac{\partial E}{\partial y_{21}} w_{21}$$
$$\frac{\partial E}{\partial x_{32}} = \frac{\partial E}{\partial y_{21}} w_{22} + \frac{\partial E}{\partial y_{22}} w_{21}$$
$$\frac{\partial E}{\partial x_{33}} = \frac{\partial E}{\partial y_{22}} w_{22}$$

Gradient

- The **forward pass** can be computed via matrix multiplication:

$$x = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{21} & x_{22} & x_{23} & x_{31} & x_{32} & x_{33} \end{bmatrix} \quad W = \begin{bmatrix} w_{11} & 0 & 0 & 0 \\ w_{12} & w_{11} & 0 & 0 \\ 0 & w_{12} & 0 & 0 \\ w_{21} & 0 & w_{11} & 0 \\ w_{22} & w_{21} & w_{12} & w_{11} \\ 0 & w_{22} & 0 & w_{12} \\ 0 & 0 & w_{21} & 0 \\ 0 & 0 & w_{22} & w_{21} \\ 0 & 0 & 0 & w_{22} \end{bmatrix}$$

Gradient

- The **forward pass** can be computed via matrix multiplication:

$$x = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{21} & x_{22} & x_{23} & x_{31} & x_{32} & x_{33} \end{bmatrix} \quad W = \begin{bmatrix} w_{11} & 0 & 0 & 0 \\ w_{12} & w_{11} & 0 & 0 \\ 0 & w_{12} & 0 & 0 \\ w_{21} & 0 & w_{11} & 0 \\ w_{22} & w_{21} & w_{12} & w_{11} \\ 0 & w_{22} & 0 & w_{12} \\ 0 & 0 & w_{21} & 0 \\ 0 & 0 & w_{22} & w_{21} \\ 0 & 0 & 0 & w_{22} \end{bmatrix}$$

- The **gradient** w.r.t. the **kernel** weights can be calculated via **convolution**:

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

Gradient

- The **gradients** w.r.t. the **inputs** can be computed by "**Transposed convolution**":

$$\frac{\partial E}{\partial x_{ij}} = \partial Ey \times W^T$$

$$\partial Ey = \begin{bmatrix} \frac{\partial E}{\partial y_{11}} & \frac{\partial E}{\partial y_{12}} & \frac{\partial E}{\partial y_{21}} & \frac{\partial E}{\partial y_{22}} \end{bmatrix}$$

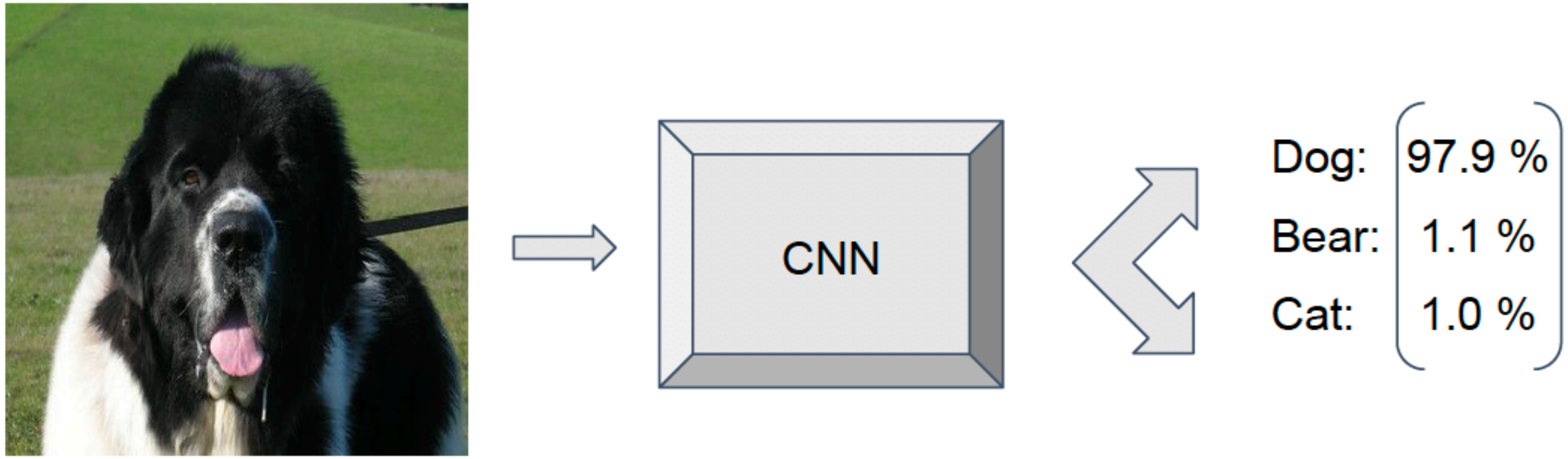
$$W^T = \begin{bmatrix} w_{11} & w_{12} & 0 & w_{21} & w_{22} & 0 & 0 & 0 & 0 \\ 0 & w_{11} & w_{12} & 0 & w_{21} & w_{22} & 0 & 0 & 0 \\ 0 & 0 & 0 & w_{11} & w_{12} & 0 & w_{21} & w_{22} & 0 \\ 0 & 0 & 0 & 0 & w_{11} & w_{12} & 0 & w_{21} & w_{22} \end{bmatrix}$$

Topics

- ***Gradient Descent in CNNs***
- CNN Applications

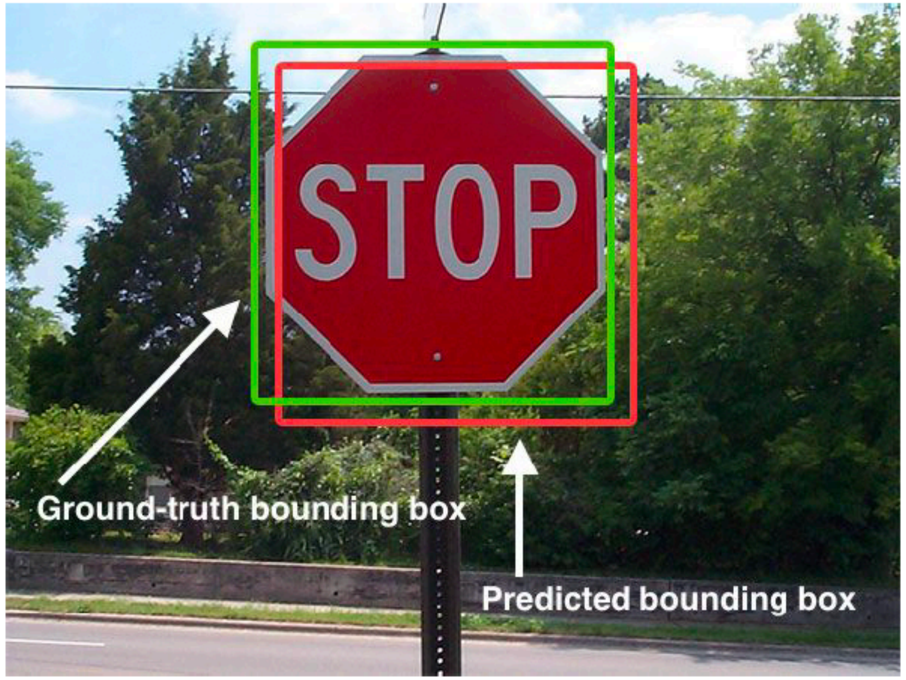
CNN Application – Image Classification

- Image classification can be defined as "given an image, what category does the image belong to?". CNNs are not the only way to do this but happen to be very good at it. This is the most commonly used task for CNNs.



CNN Application – Object Detection

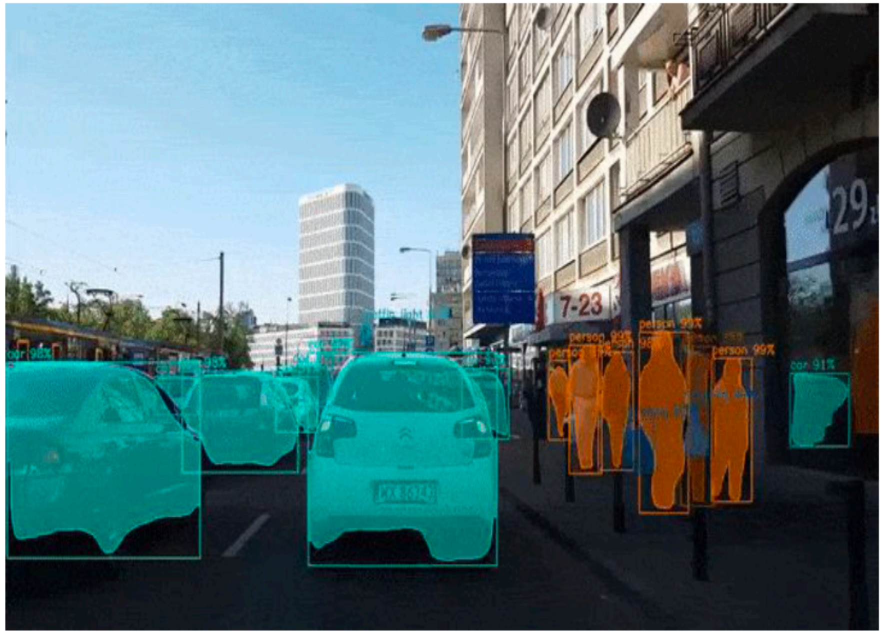
- In object detection, coordinates to a bounding box are predicted, as well as the category of object the box belongs to.



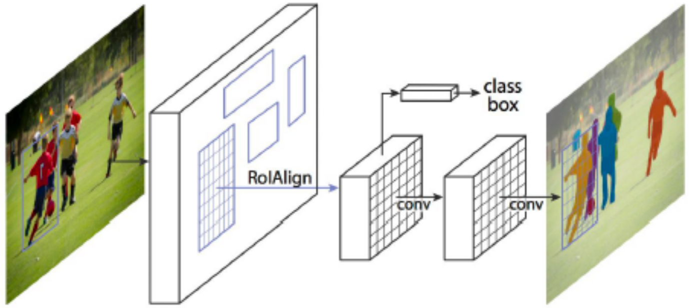
https://upload.wikimedia.org/wikipedia/commons/2/2d/Intersection_over_Union_-_object_detection_bounding_boxes.jpg

CNN Application – Image Segmentation

- Image segmentation is the next logical step in object detection; it determines which pixels correspond to the detected objects.



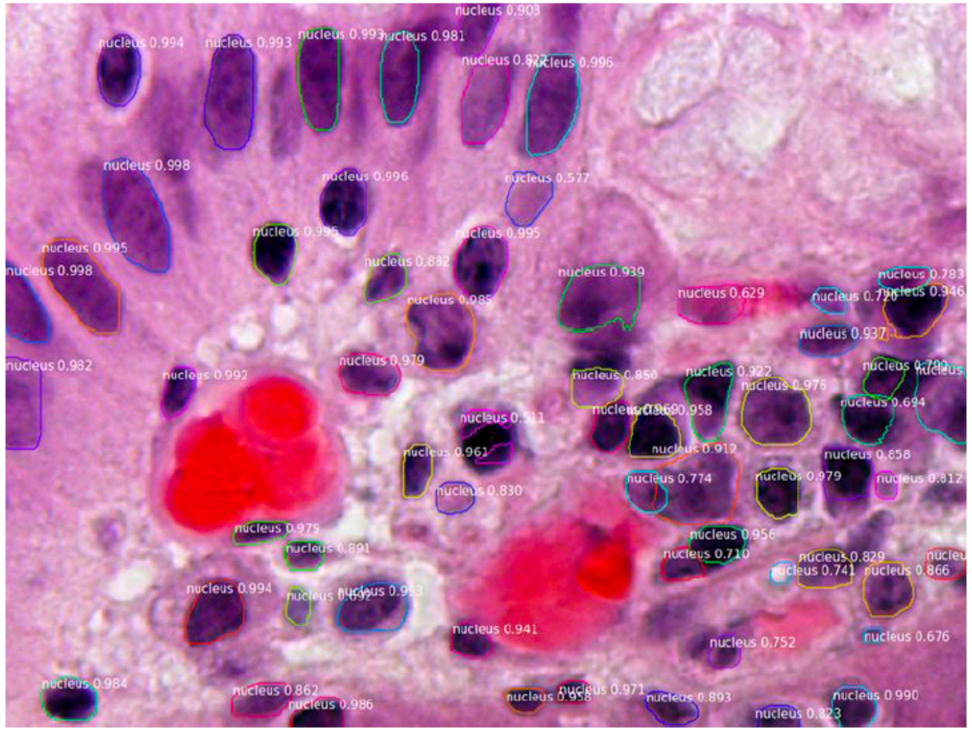
https://github.com/matterport/Mask_RCNN

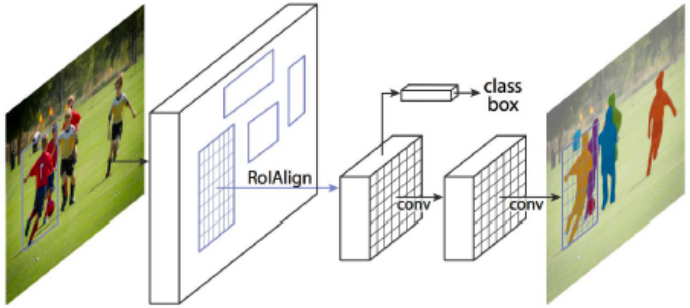


Mask RCNN Architecture

CNN Application – Image Segmentation

- It can be used in a wide variety of fields, such as medical imaging.



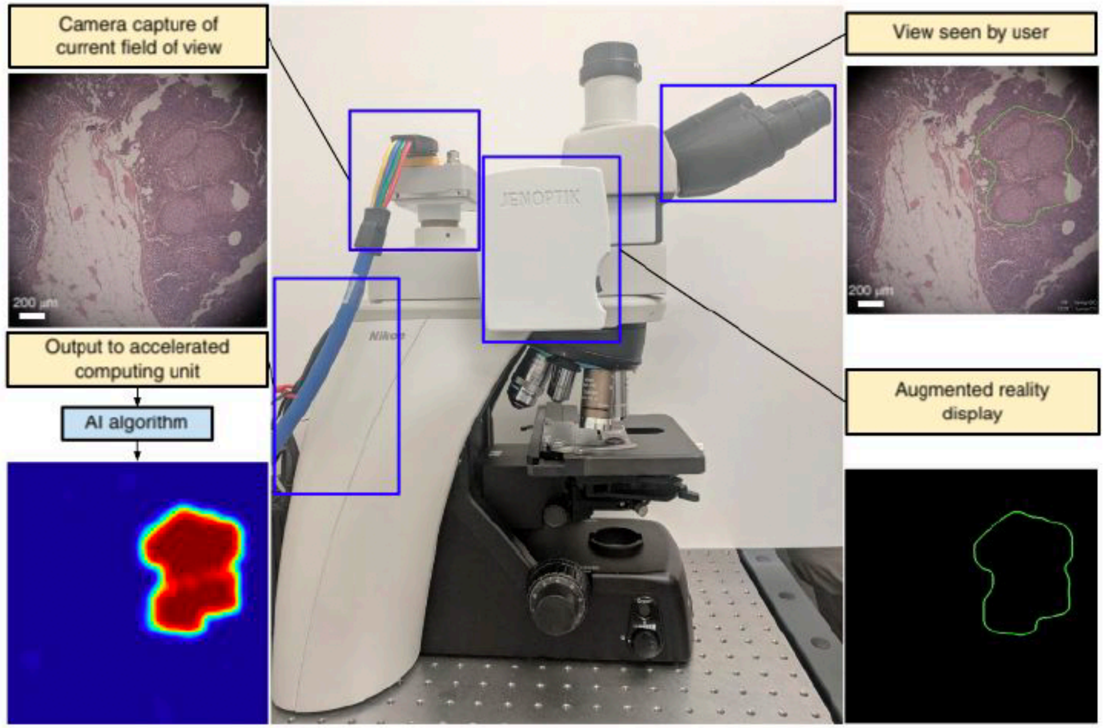


Mask RCNN Architecture

https://github.com/matterport/Mask_RCNN

Augmented (Medical) Reality

- Here is an example of a microscope rendering in real-time regions of interest in medical images.

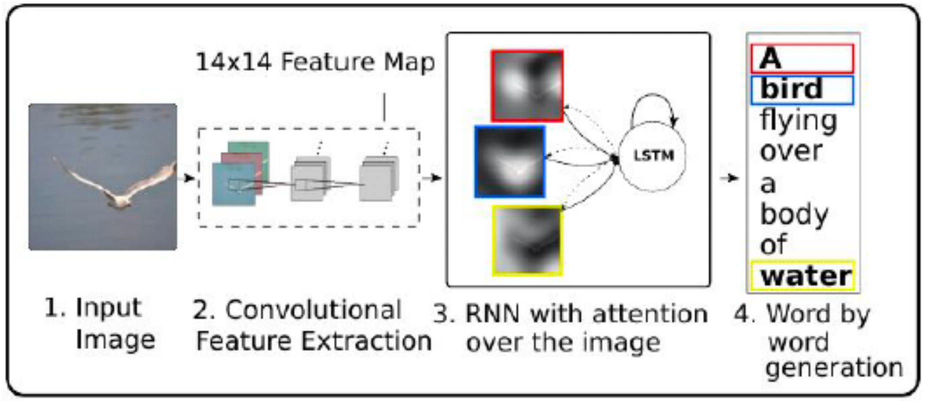


Caption Generation

- CNNs can be used in a pipeline (along with other NLP models) to help models learn to generate captions.

A woman is throwing a frisbee in a park.

Figure 1. Our model learns a words/image alignment. The visualized attentional maps (3) are explained in Sections 3.1 & 5.4



https://arxiv.org/pdf/1502.03044.pdf

Human Pose Estimation

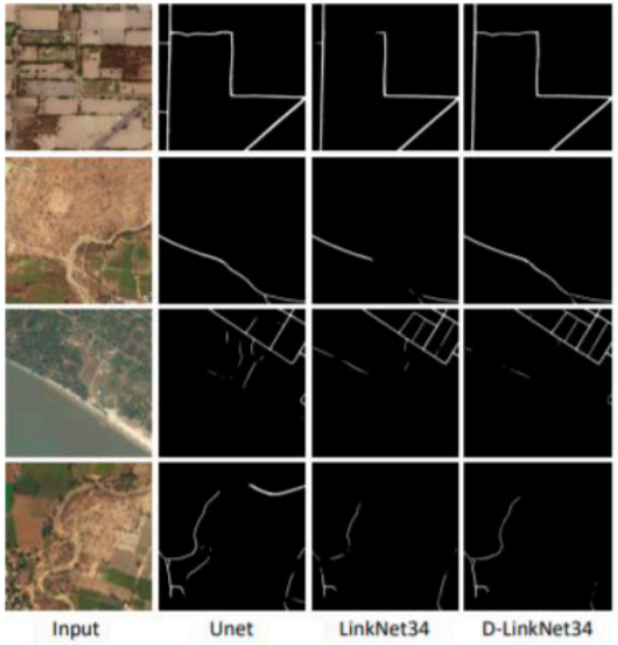
- It is possible to use CNNs to extract human poses from images

https://github.com/rachit2403/Open-Pose-Keras

https://github.com/facebookresearch/DensePose

Road Mapping

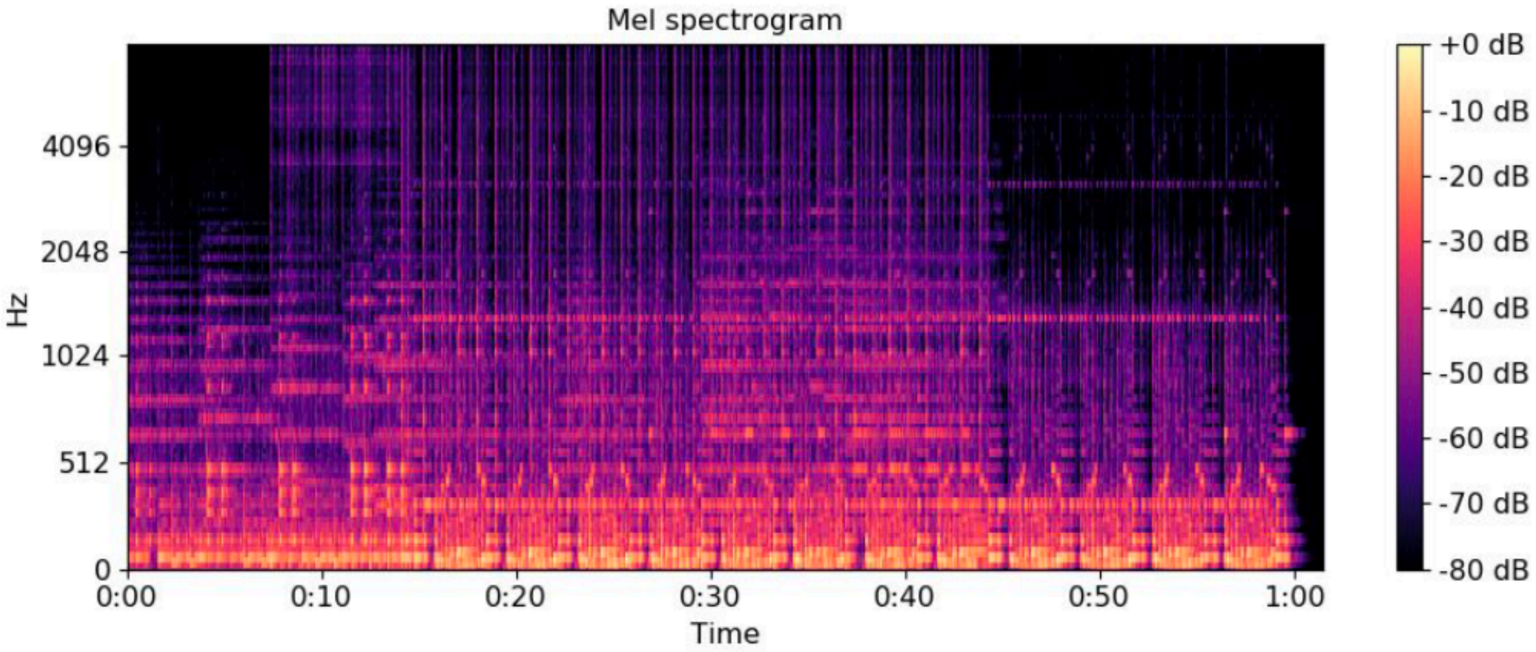
- Using satellite imagery, we can automate road + building mapping.



https://ai.facebook.com/blog/mapping-roads-through-deep-learning-and-weakly-supervised-training/
https://github.com/matterport/Mask_RCNN

Sound Classification

- Spectrograms of sounds can be computed using fourier transforms. The spectrograms can then be treated just like images in the context of sound classification.



https://librosa.github.io/librosa/generated/librosa.feature.melspectrogram.html

The End