Recitation 5

CNN: Basics and Backprop

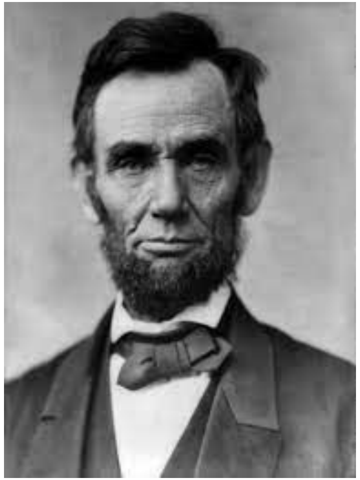
Instructors:

- Syed
- Miya
- Denis

17th Feb 2024

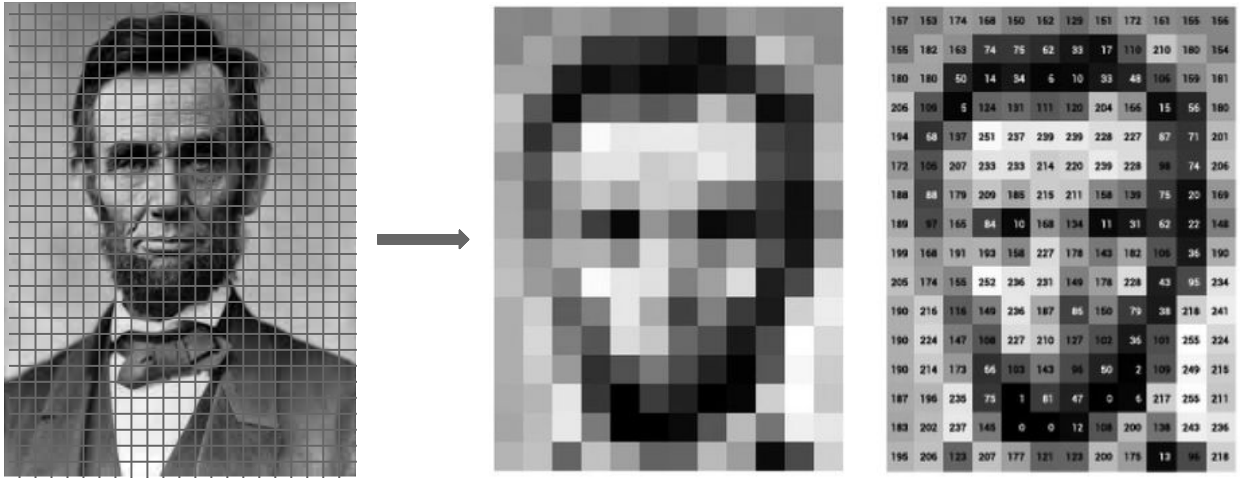
What is an image?

A visual representation



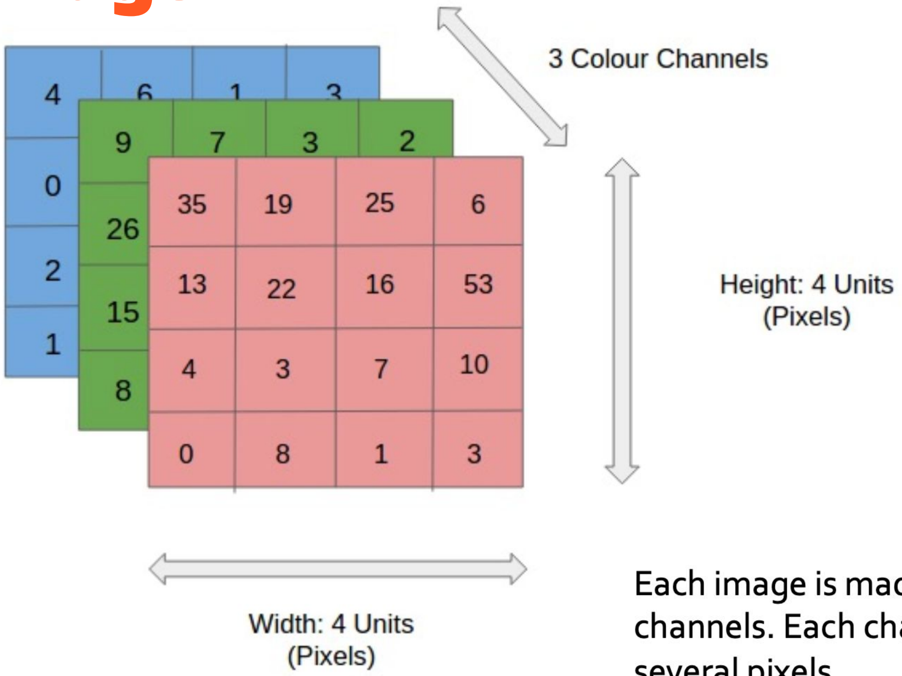
What is an image? : For a computer!

~~A visual representation~~. A Matrix **`I`** of dimensions **`(M,N)`** with **`I[i][j] = intensity(pixel(i,j))`**



157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

What is an image?



`I → (3,M,N)`

`I[c][i][j] =`

`Intensity at pixel(i,j) for channel c`

Each image is made up of a set of channels. Each channel comprises of several pixels

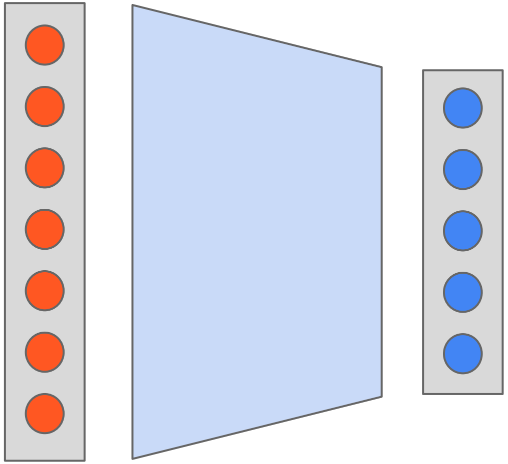
3 for a colored image, 1 for B&W.

The number of channels you encounter could even increase!

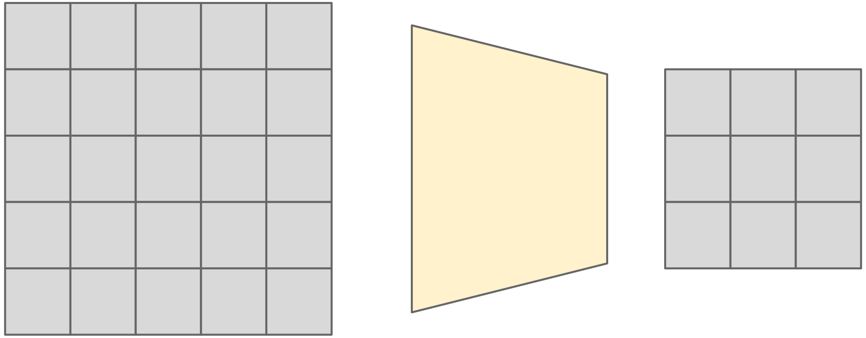
CNN

A CNN is a specialized neural network which employs convolutional and pooling layers to extract features and hierarchical patterns automatically from the input. It's widely used in tasks like image recognition and object detection due to its ability to learn and recognize complex visual patterns.

MLP Vs. CNN



Vector to Vector



Feature map to Feature map

Building Blocks of a CNN

Main Building blocks

- Convolution Layer
- Pooling Layer

Others(can also be found in MLP)

- Activation Layer
- Normalization Layer(LayerNorm, etc)
- Batch Normalization (BatchNorm)

Building Blocks of a CNN

Hyperparameters

Conv layer:

- Filter/kernel size
- Stride
- # of filters,
- Padding

Pooling layer:

- Pooling type & size(pool size)
- Stride

# of layers

Convolutional Layer(Conv layer)

Convolutional layers are the core components of CNNs. They apply convolution operations using learnable filters (kernels) to the input data. These filters slide across the input to detect patterns, edges, and features.

Kernel/Filter size

The size of the convolutional kernels (filters) determines the spatial extent over which the convolution operation is applied. Common kernel sizes are 3x3, 5x5, or 7x7.

Stride

The stride specifies the step size at which the convolutional kernel/filter is moved across the input data. A larger stride reduces the spatial dimensions of the output feature maps.

Taking bigger steps!

Padding

Padding in Convolutional Neural Networks (CNNs) is a technique used to control the spatial dimensions of the output feature maps produced by convolutional layers. It involves adding extra rows and columns of zeros (or other values) around the input data before applying the convolution operation

Convolution

Essentially element-wise (Hadamard) multiplications and summations**(Dot product)**

Input - **A**

0	0	0	0	1	1	1	1
0	0	0	0	1	1	1	1
0	0	0	0	1	1	1	1
0	0	0	0	1	1	1	1
0	0	0	0	1	1	1	1
0	0	0	0	1	1	1	1
0	0	0	0	1	1	1	1
0	0	0	0	1	1	1	1

*

Kernel - **W**

-1	0	1
-2	0	2
-1	0	1

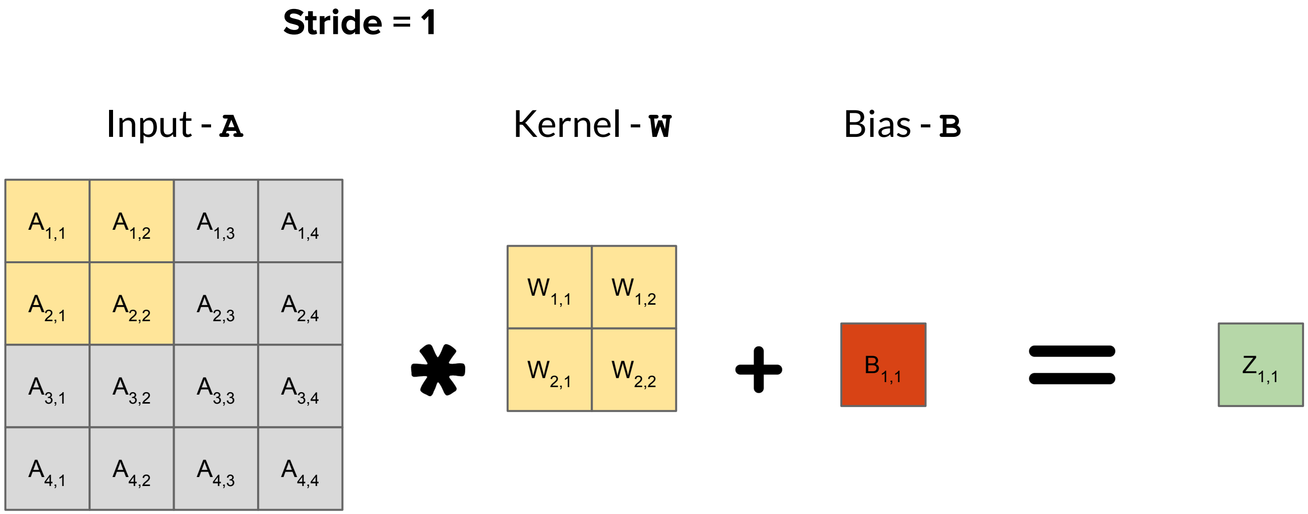
=

0	0	4	4	0	0
0	0	4	4	0	0
0	0	4	4	0	0
0	0	4	4	0	0
0	0	4	4	0	0
0	0	4	4	0	0

Here the stride is 1

Convolution

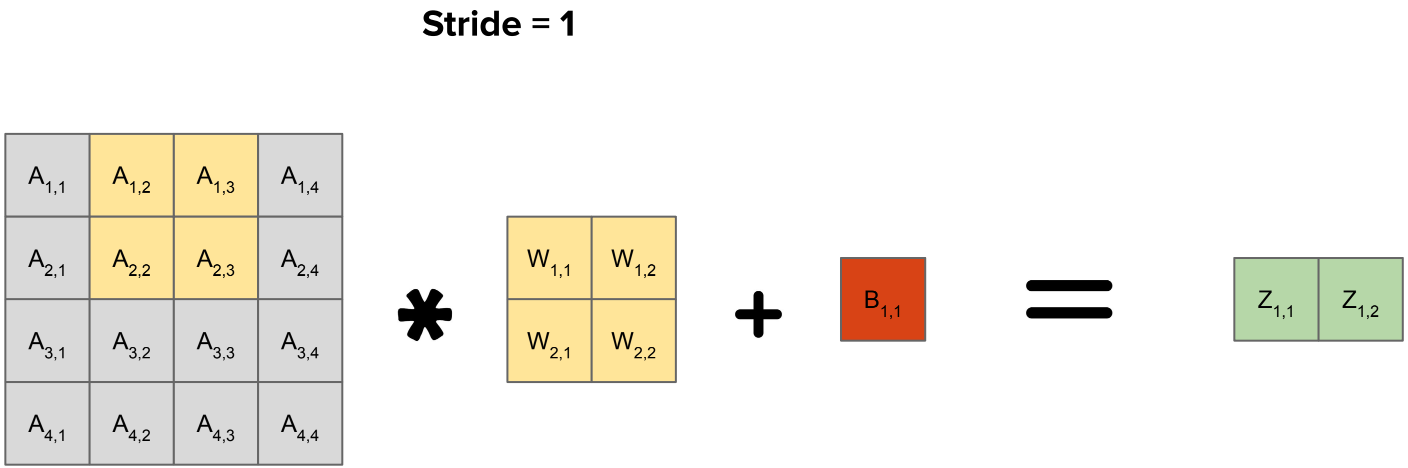
Essentially element-wise (Hadamard) multiplications and summations



$$Z_{1,1} = (A_{1,1} * W_{1,1}) + (A_{1,2} * W_{1,2}) + (A_{2,1} * W_{2,1}) + (A_{2,2} * W_{2,2}) + B$$

Convolution

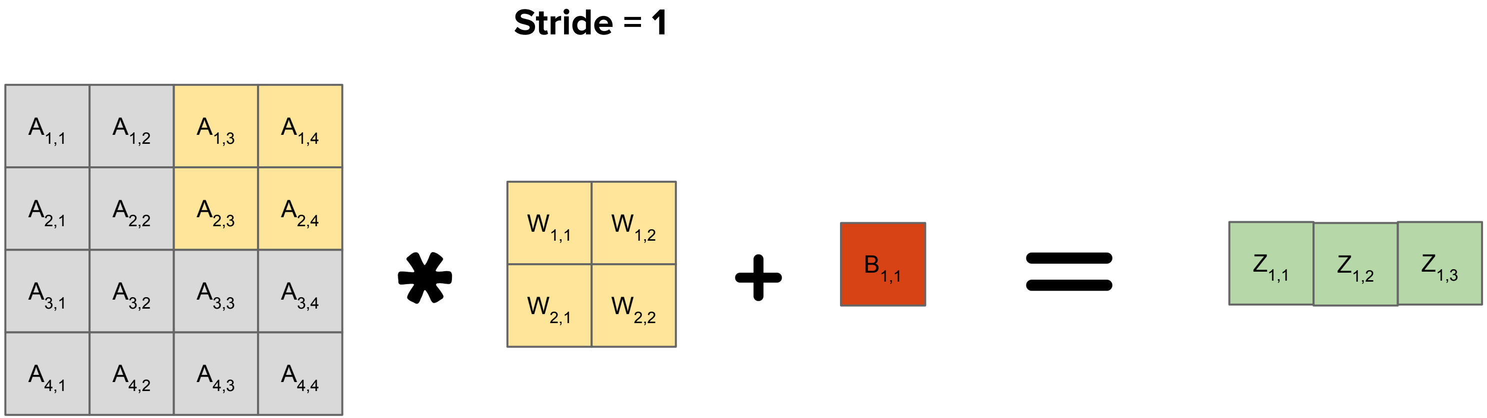
Essentially element-wise (Hadamard) multiplications and summations



$$Z_{1,2} = (A_{1,2} * W_{1,1}) + (A_{1,3} * W_{1,2}) + (A_{2,2} * W_{2,1}) + (A_{2,3} * W_{2,2}) + B$$

Convolution

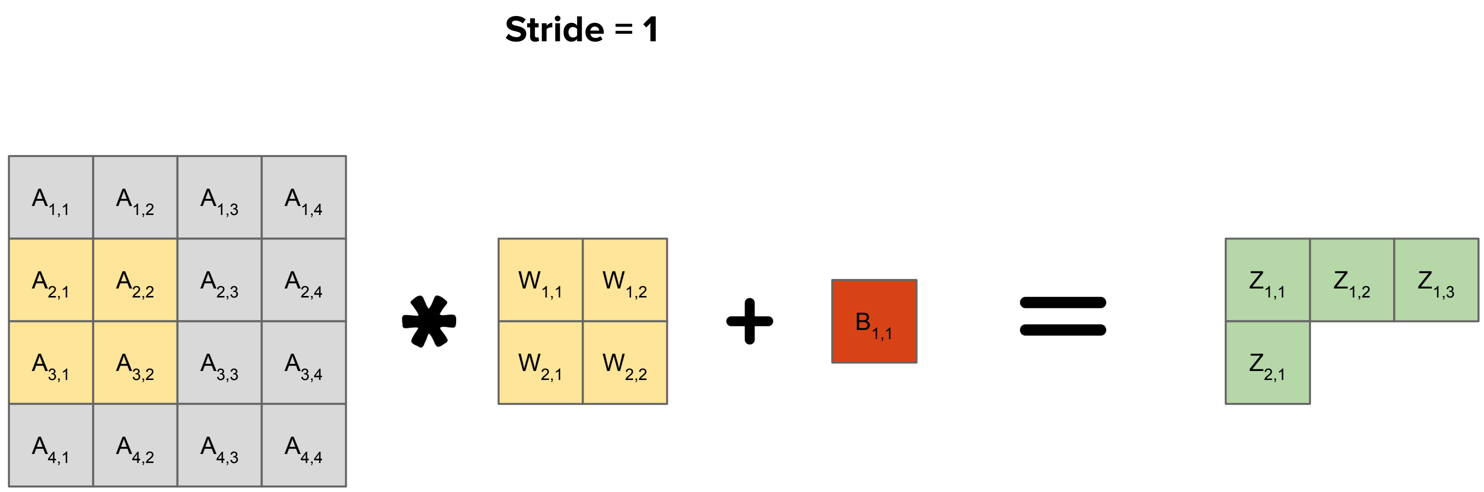
Essentially element-wise (Hadamard) multiplications and summations



$$Z_{1,3} = (A_{1,3}*W_{1,1}) + (A_{1,4}*W_{1,2}) + (A_{2,3}*W_{2,1}) + (A_{2,4}*W_{2,2}) + B$$

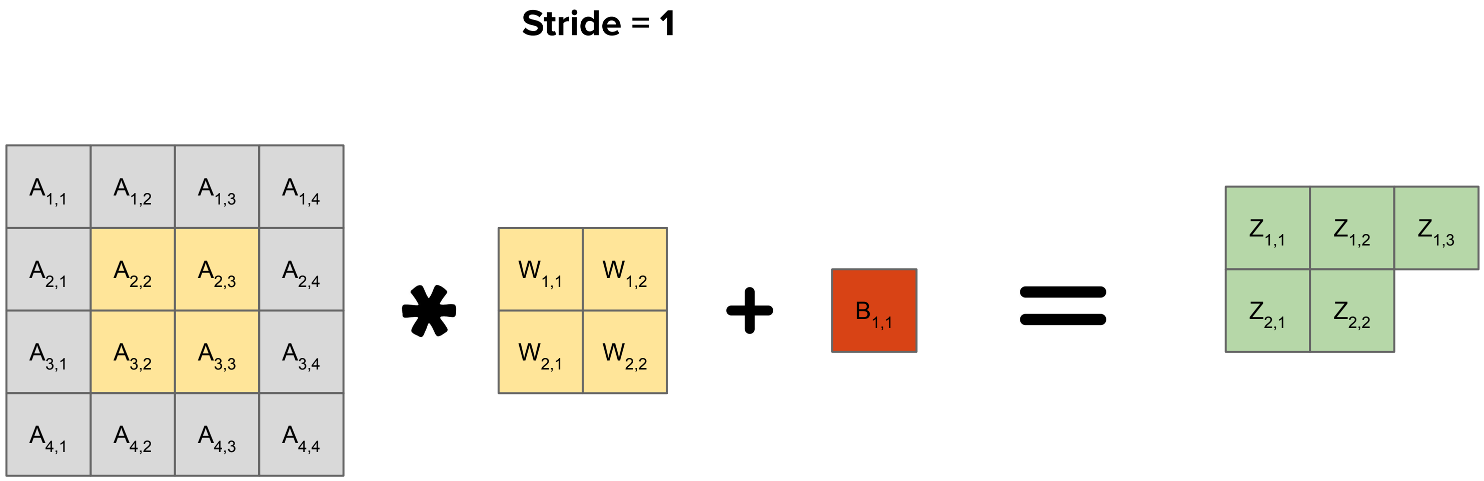
Convolution

Essentially element-wise (Hadamard) multiplications and summations



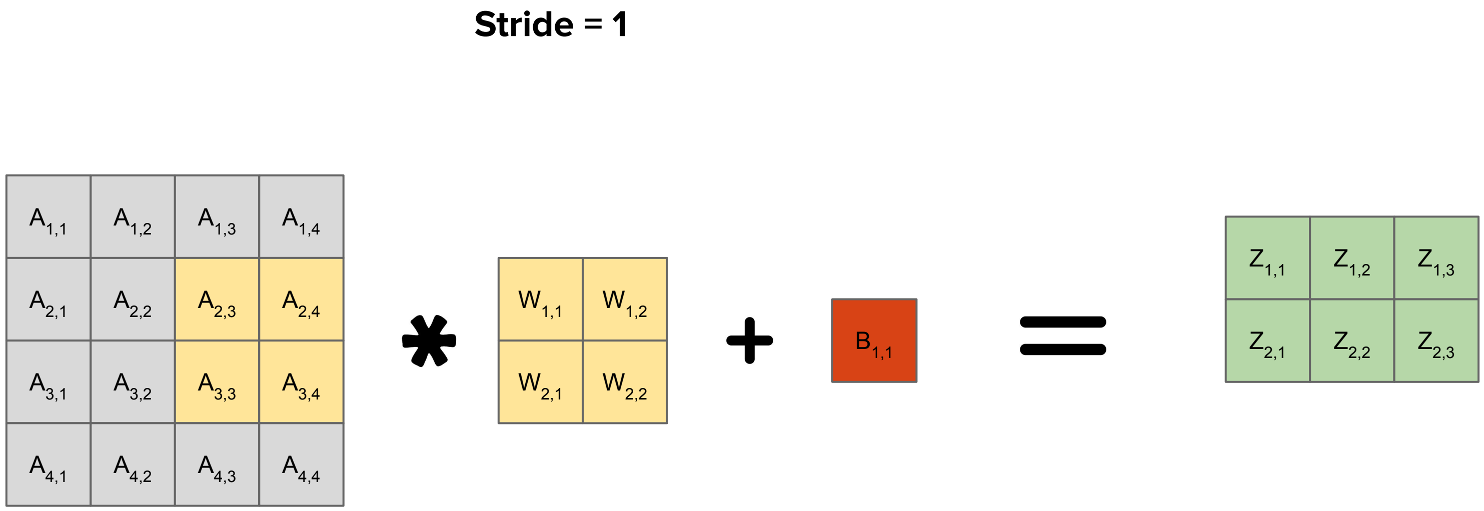
Convolution

Essentially element-wise (Hadamard) multiplications and summations



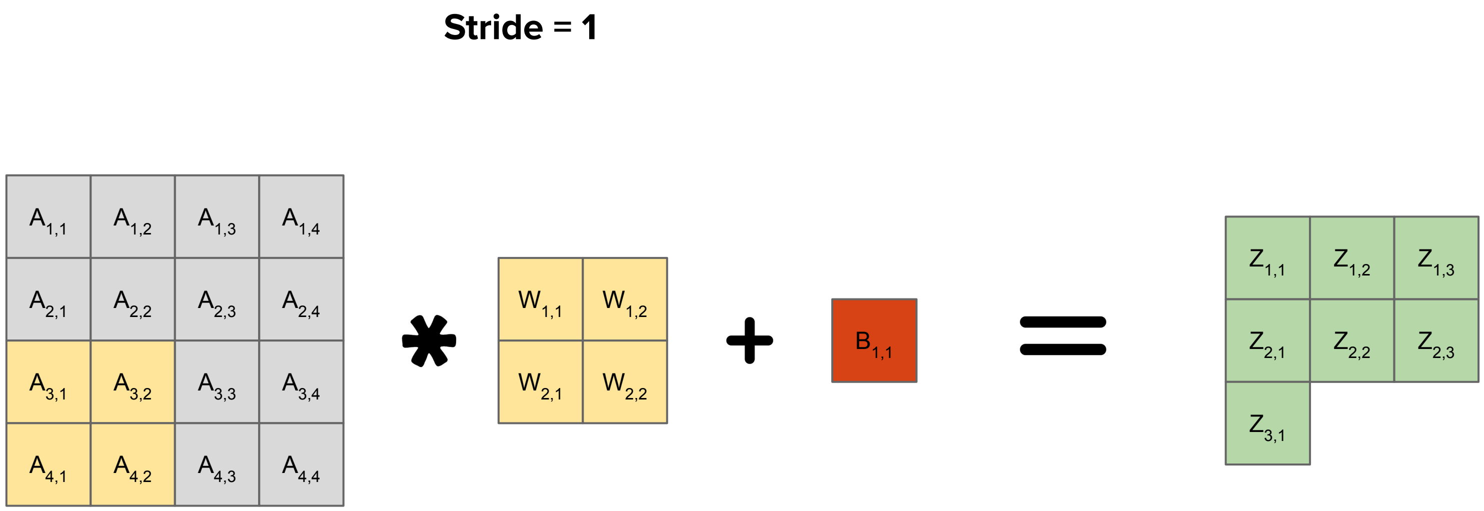
Convolution

Essentially element-wise (Hadamard) multiplications and summations



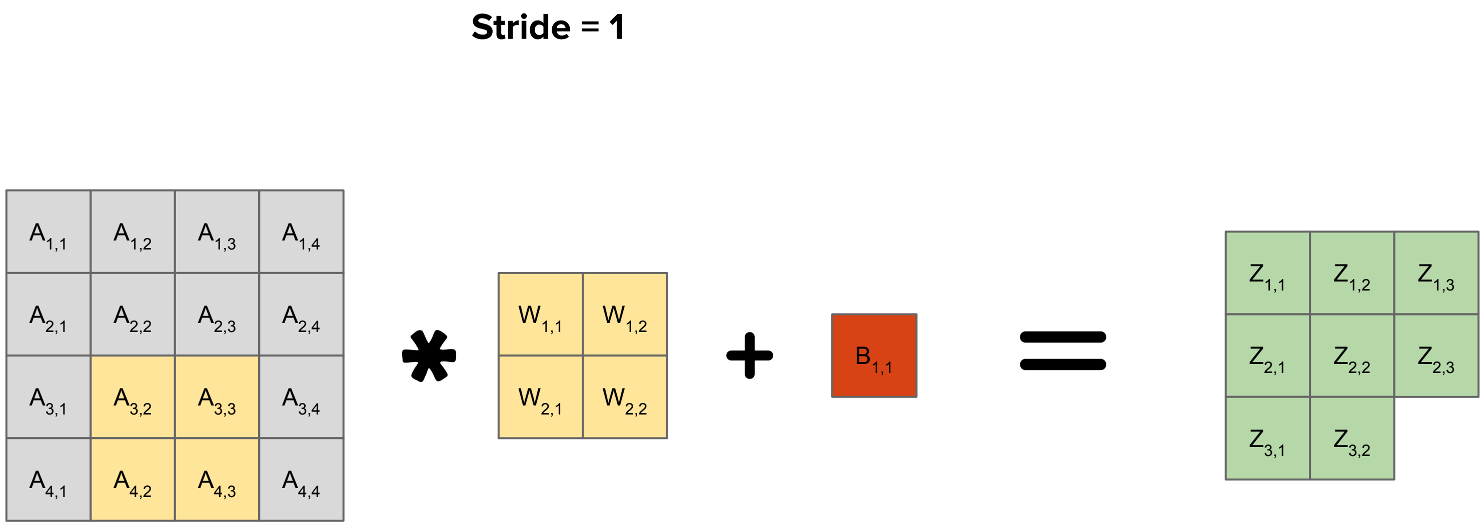
Convolution

Essentially element-wise (Hadamard) multiplications and summations



Convolution

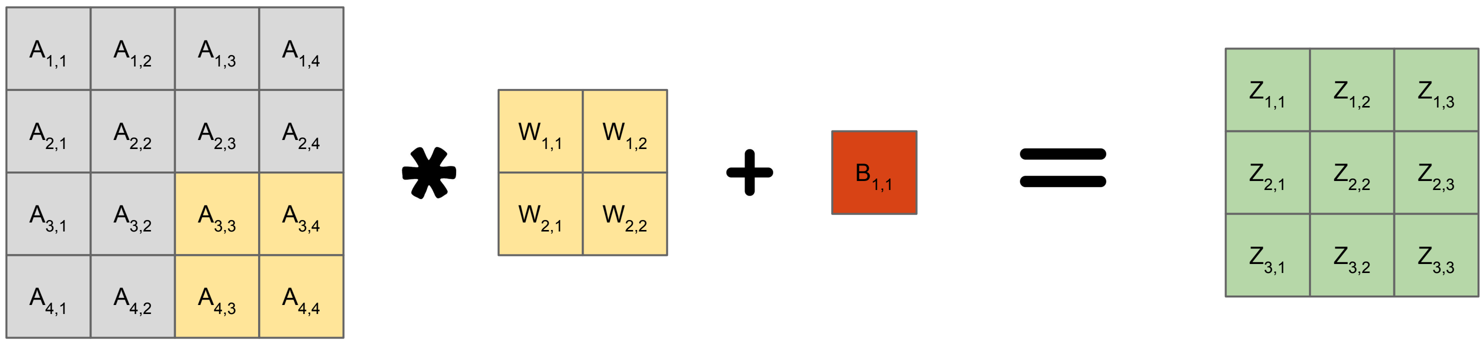
Essentially element-wise (Hadamard) multiplications and summations



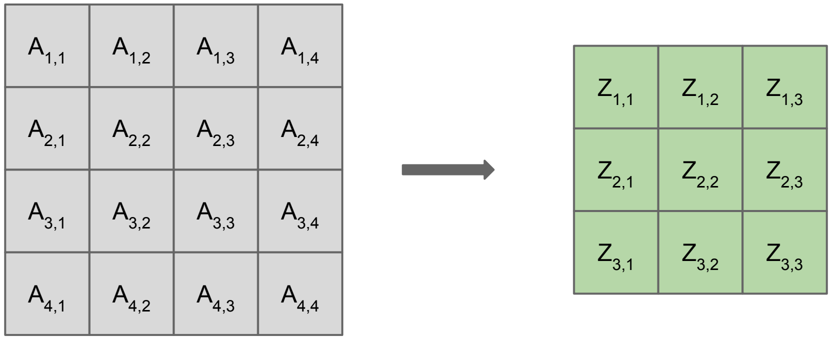
Convolution

Essentially element-wise (Hadamard) multiplications and summations

Stride = 1

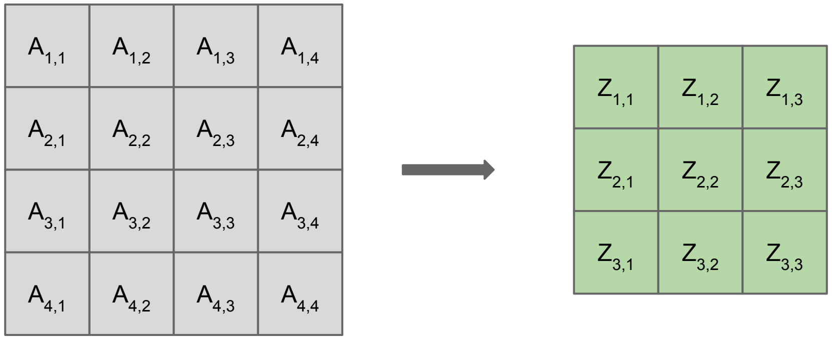


Output Size



$A_{1,1}$ $A_{1,2}$ $A_{1,3}$ $A_{1,4}$
$A_{2,1}$ $A_{2,2}$ $A_{2,3}$ $A_{2,4}$
$A_{3,1}$ $A_{3,2}$ $A_{3,3}$ $A_{3,4}$
$A_{4,1}$ $A_{4,2}$ $A_{4,3}$ $A_{4,4}$
$Z_{1,1}$ $Z_{1,2}$ $Z_{1,3}$
$Z_{2,1}$ $Z_{2,2}$ $Z_{2,3}$
$Z_{3,1}$ $Z_{3,2}$ $Z_{3,3}$

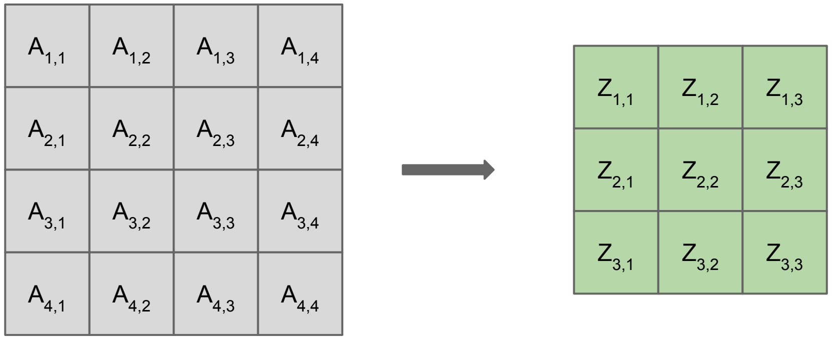
Output Size



Output Width =
$[(W_{in} - W_k + 2P) // (S)] + 1$

Same goes for Height.

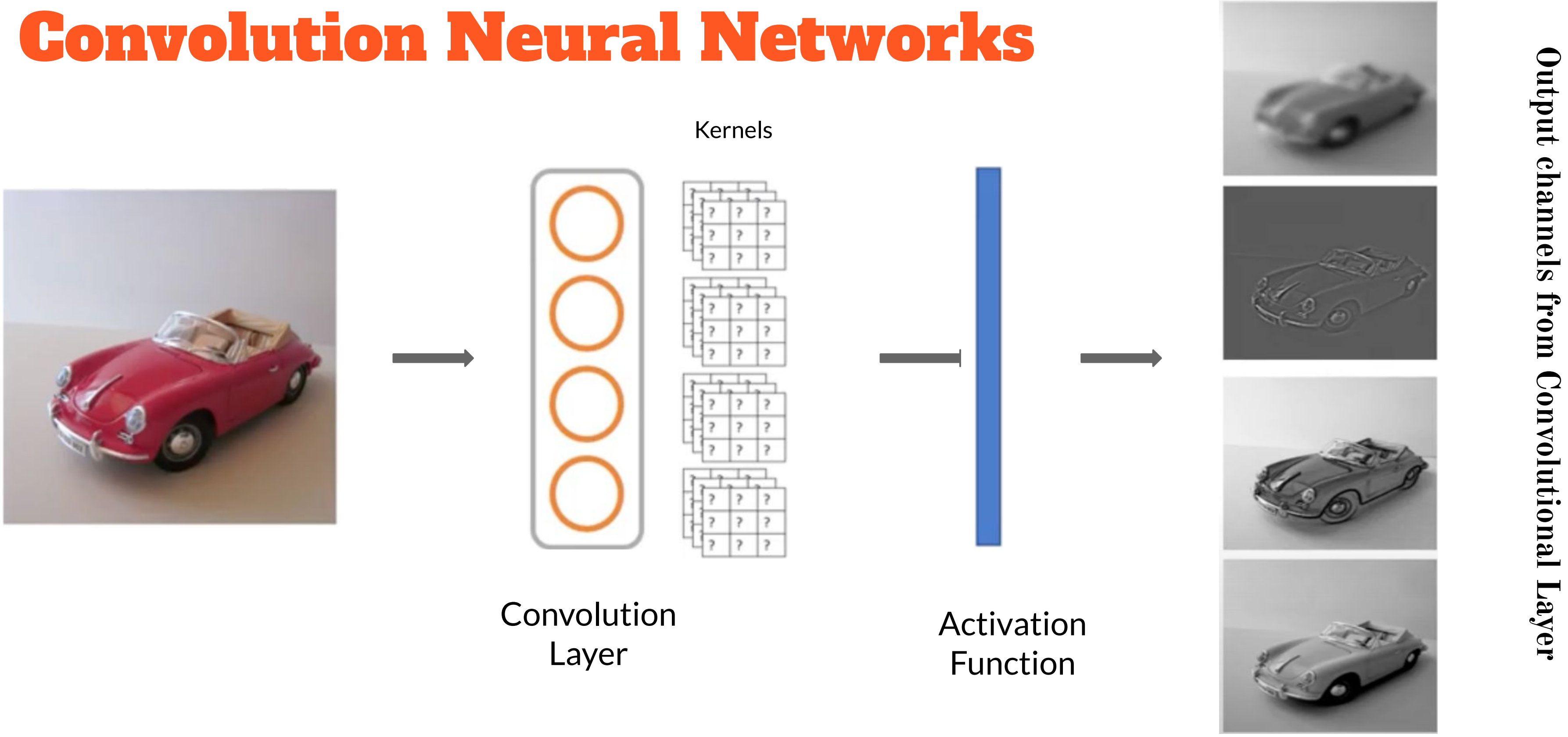
Output Size



Output Width =
$[(W_{in} - W_k + 2\mathbf{P}) \,//\, (\mathbf{S})] + 1$

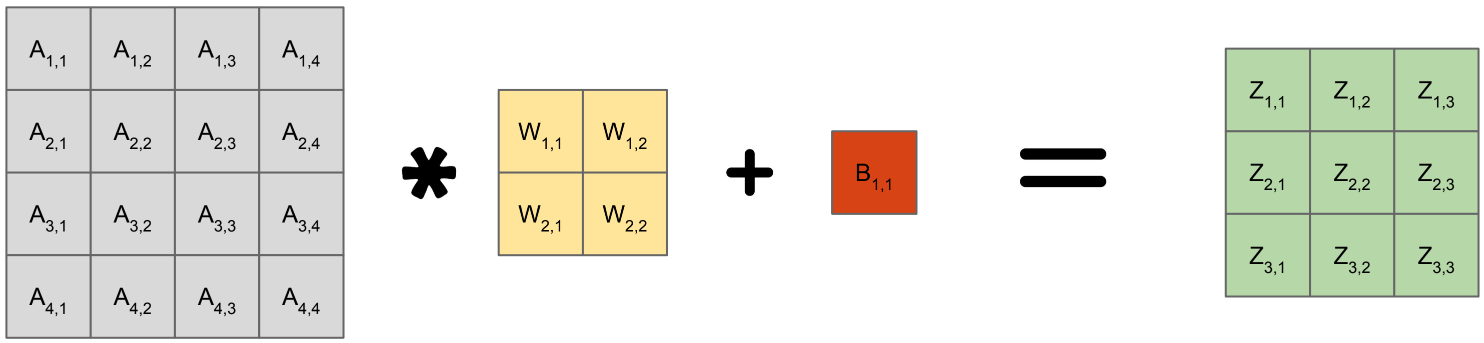
P: Padding (here - 0)
S: Stride (here - 1)

Convolution Neural Networks



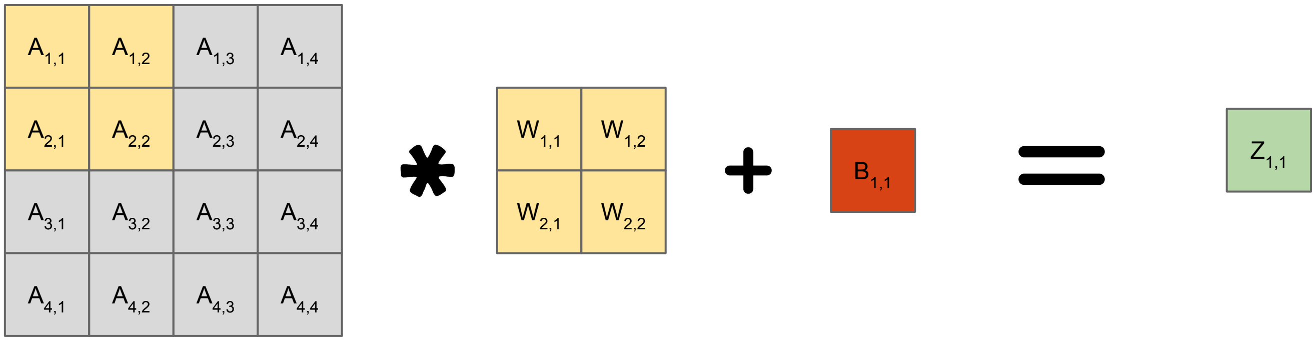
Stride = 1

What we did before - The kernel “moves” one pixel (or element) at a time.



Stride = 2

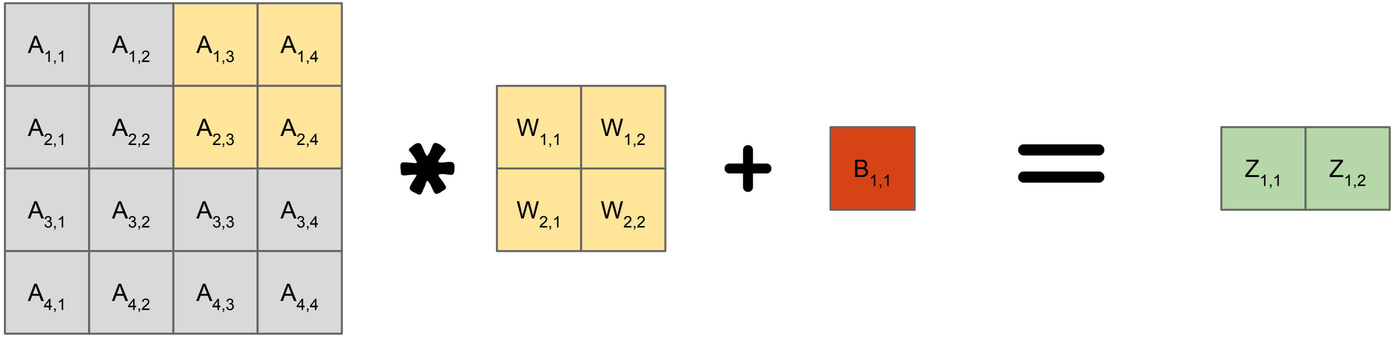
Start at the same place



$$Z_{1,1} = (A_{1,1} * W_{1,1}) + (A_{1,2} * W_{1,2}) + (A_{2,1} * W_{2,1}) + (A_{2,2} * W_{2,2}) + B$$

Stride = 2

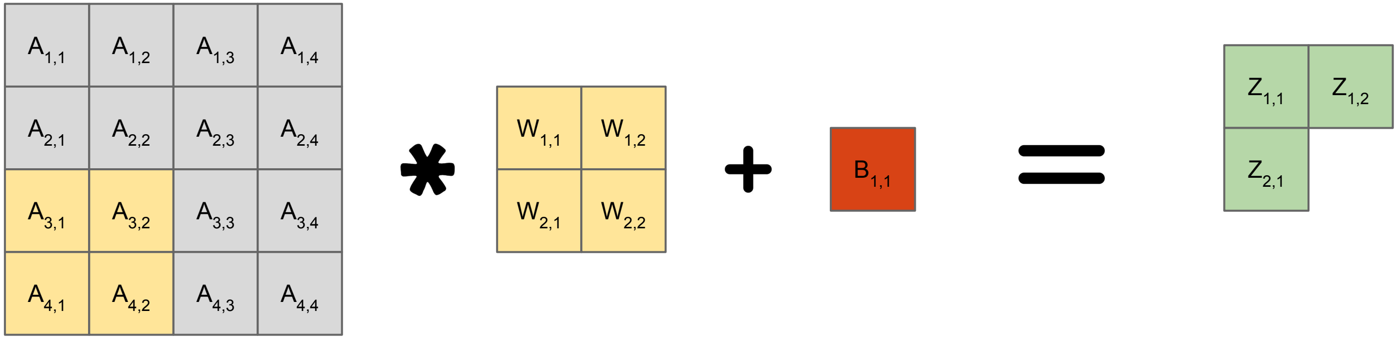
Move two elements to the right



$$Z_{1,2} = (A_{1,3}*W_{1,1}) + (A_{1,4}*W_{1,2}) + (A_{2,3}*W_{2,1}) + (A_{2,4}*W_{2,2}) + B$$

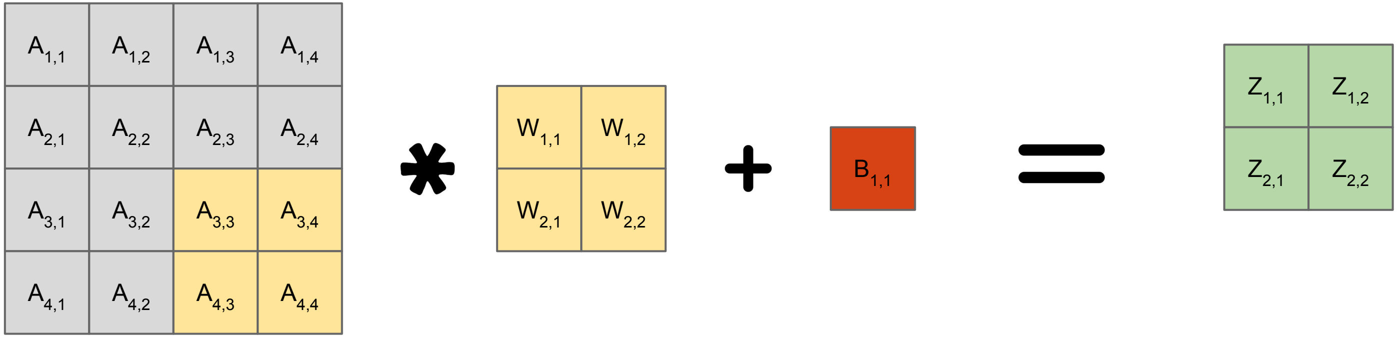
Stride = 2

Move two elements down.



Stride = 2

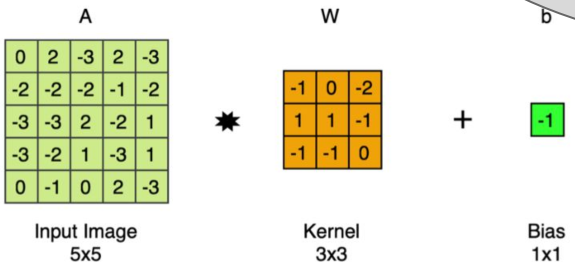
Move two elements to the right.

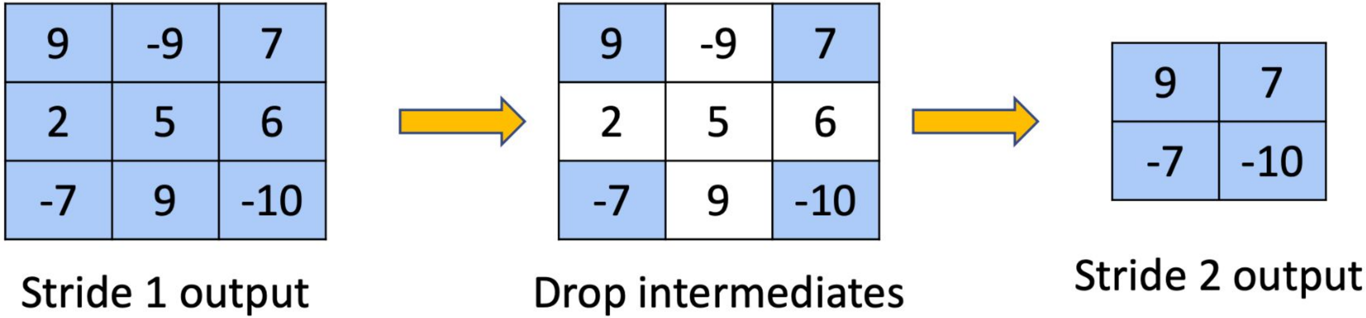


Interpreting Stride > 1

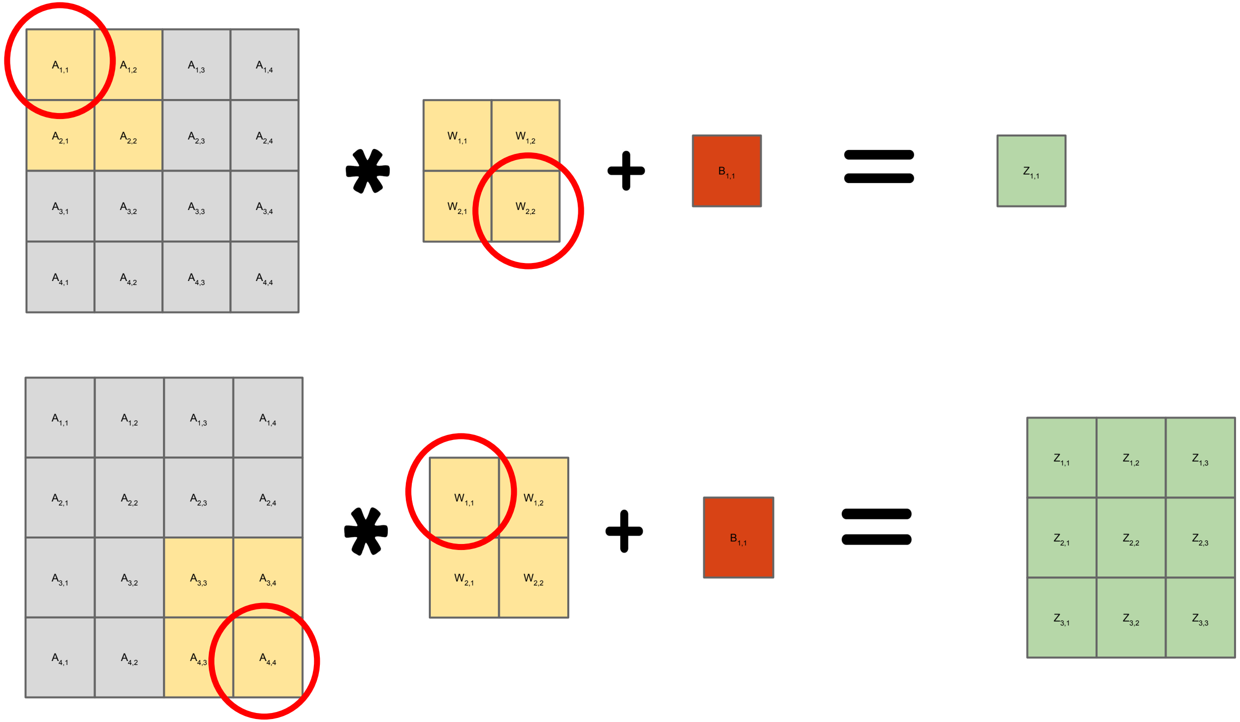
Think about how it is related to Upsampling(and Downsampling.

Will learn more in HW2





Padding



A1,1 A1,2 A1,3 A1,4
A2,1 A2,2 A2,3 A2,4
A3,1 A3,2 A3,3 A3,4
A4,1 A4,2 A4,3 A4,4
W1,1 W1,2
W2,1 W2,2
B1,1
Z1,1
A1,1 A1,2 A1,3 A1,4
A2,1 A2,2 A2,3 A2,4
A3,1 A3,2 A3,3 A3,4
A4,1 A4,2 A4,3 A4,4
W1,1 W1,2
W2,1 W2,2
B1,1
Z1,1 Z1,2 Z1,3
Z2,1 Z2,2 Z2,3
Z3,1 Z3,2 Z3,3

Padding

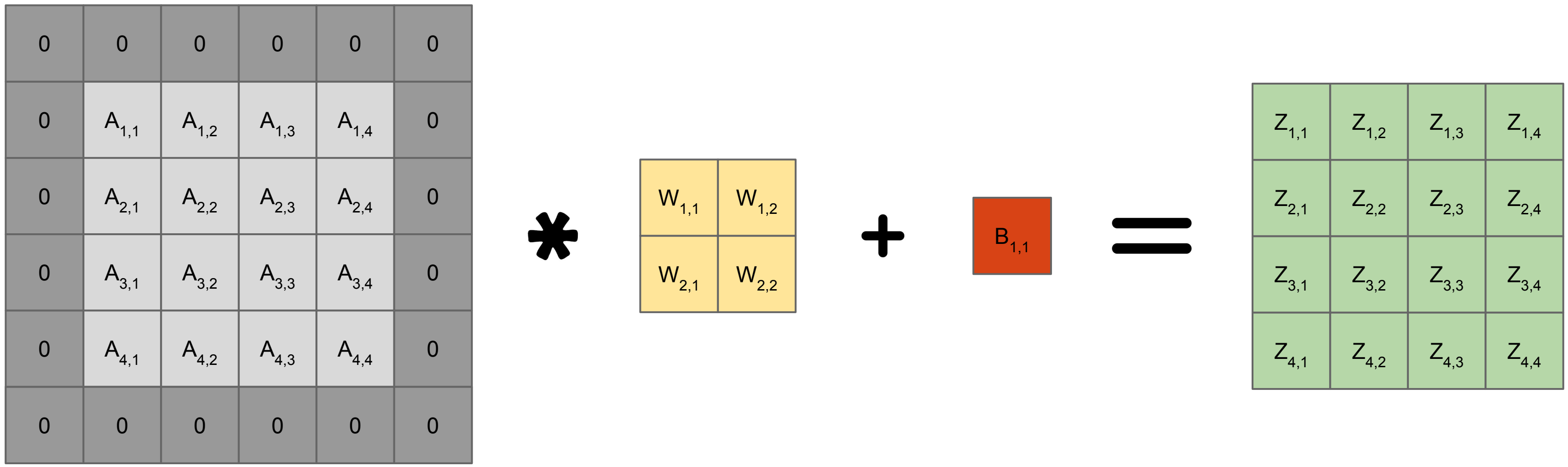
Increase output size

Preserve input size

More Kernel Interactions!

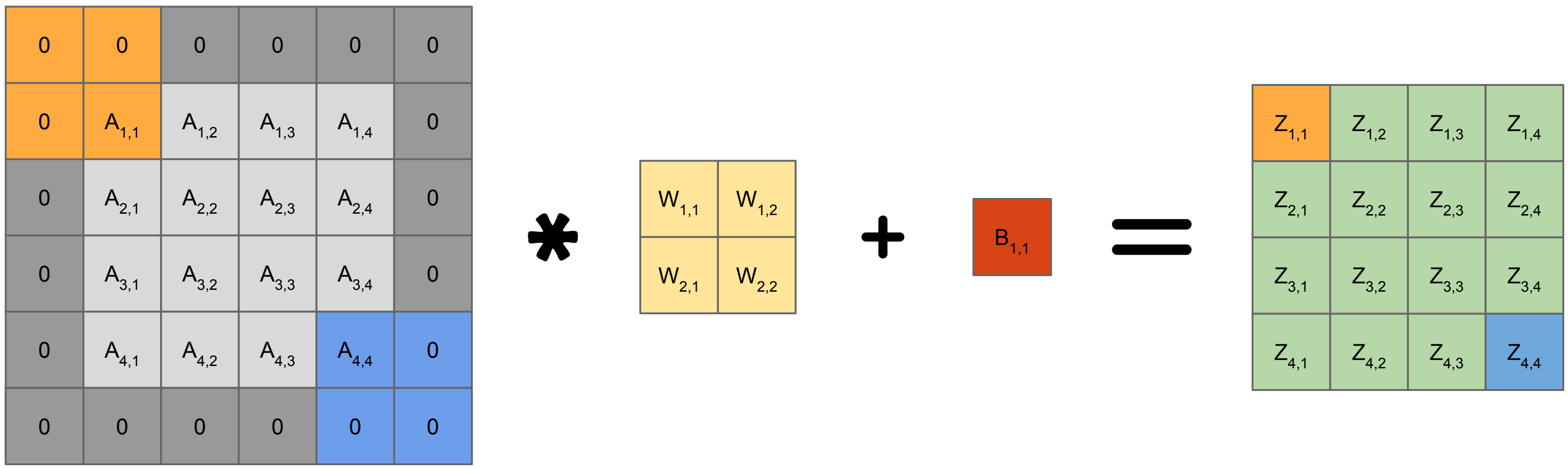
Padding

Padding = 1

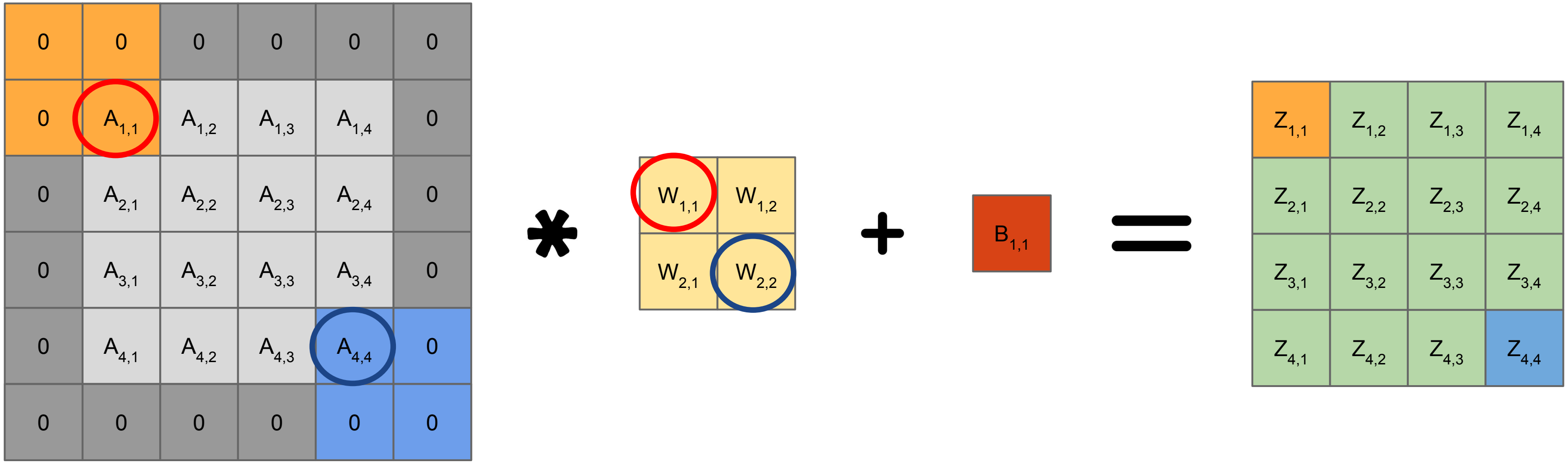


Padding

Padding = 1

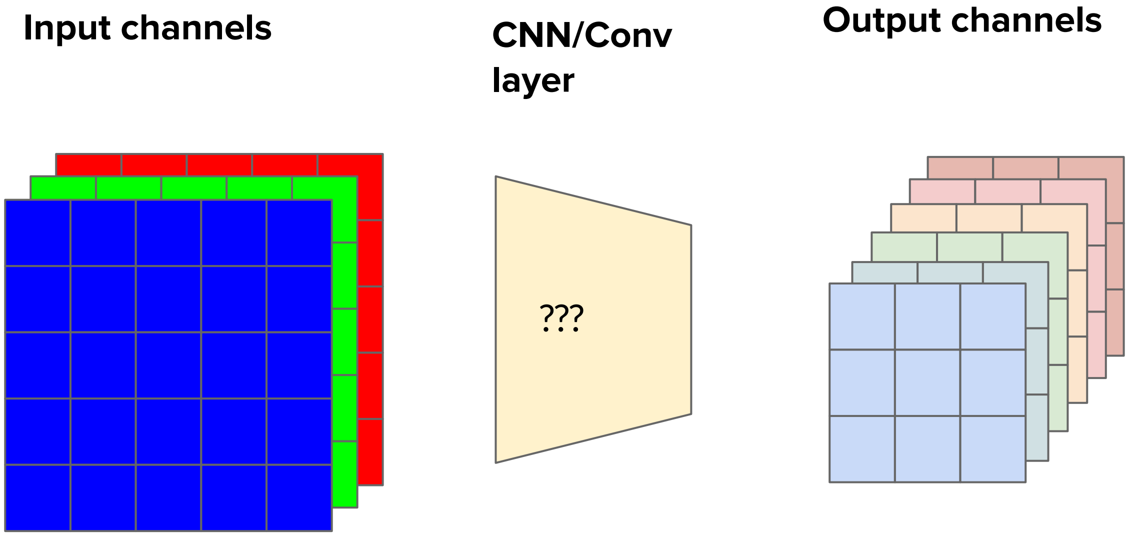


Padding



0 0 0 0 0 0
0 A1,1 A1,2 A1,3 A1,4 0
0 A2,1 A2,2 A2,3 A2,4 0
0 A3,1 A3,2 A3,3 A3,4 0
0 A4,1 A4,2 A4,3 A4,4 0
0 0 0 0 0 0
*
W1,1 W1,2
W2,1 W2,2
+
B1,1
=
Z1,1 Z1,2 Z1,3 Z1,4
Z2,1 Z2,2 Z2,3 Z2,4
Z3,1 Z3,2 Z3,3 Z3,4
Z4,1 Z4,2 Z4,3 Z4,4

Multi-channel CNN

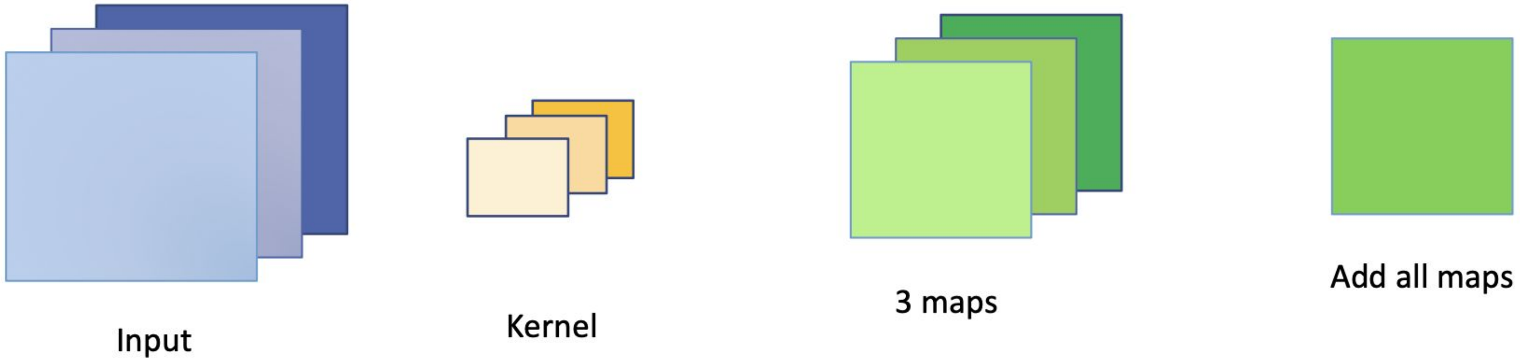


Multi-channel CNN

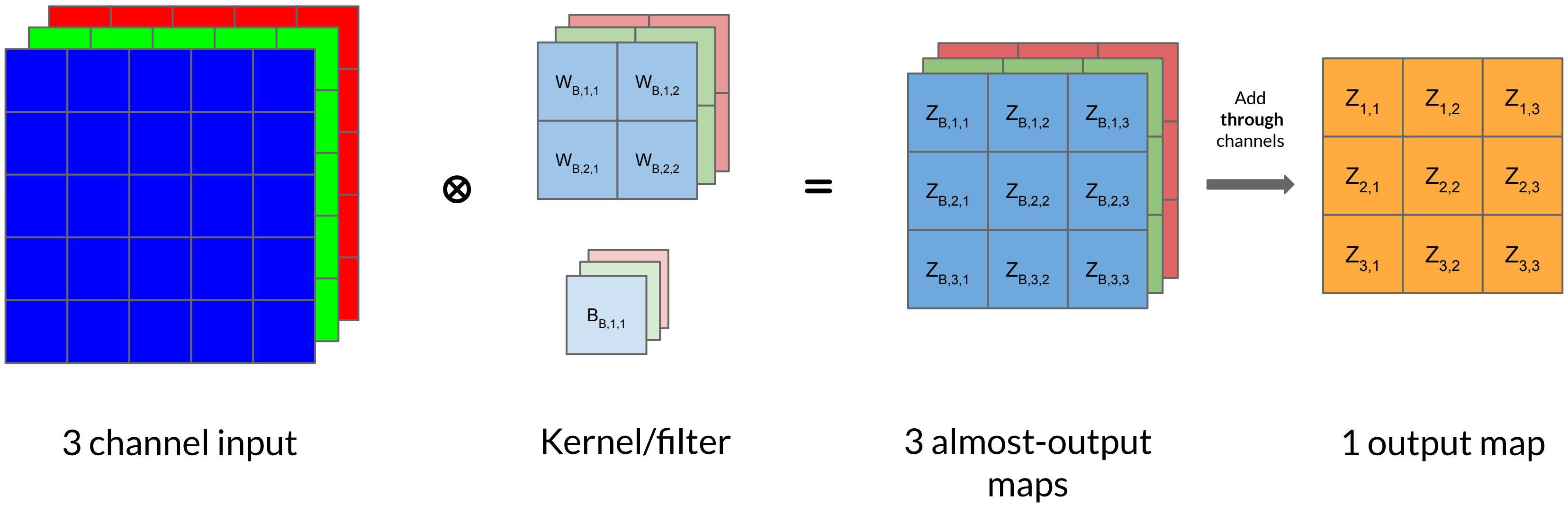
- Each kernel (or **filter**) has as many channels as the input does.

  `[kernel channels = Input channels]`

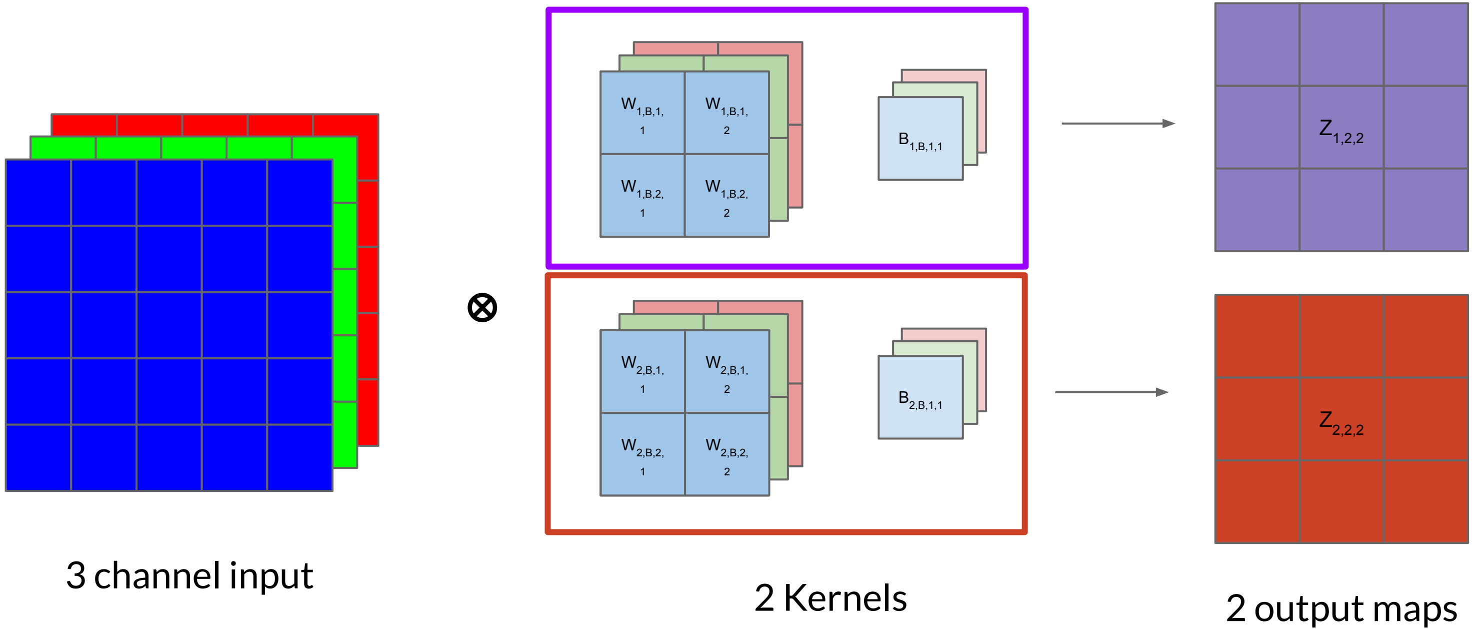
- Channel **c** of **the kernel** convolves with channel **c** (corresponding) of **the input**.
- The number of output channels from the convolution = number of **filters(kernels)** applied to the input.



1 Filter with 3-channel input



2 Filters with 3-channel input

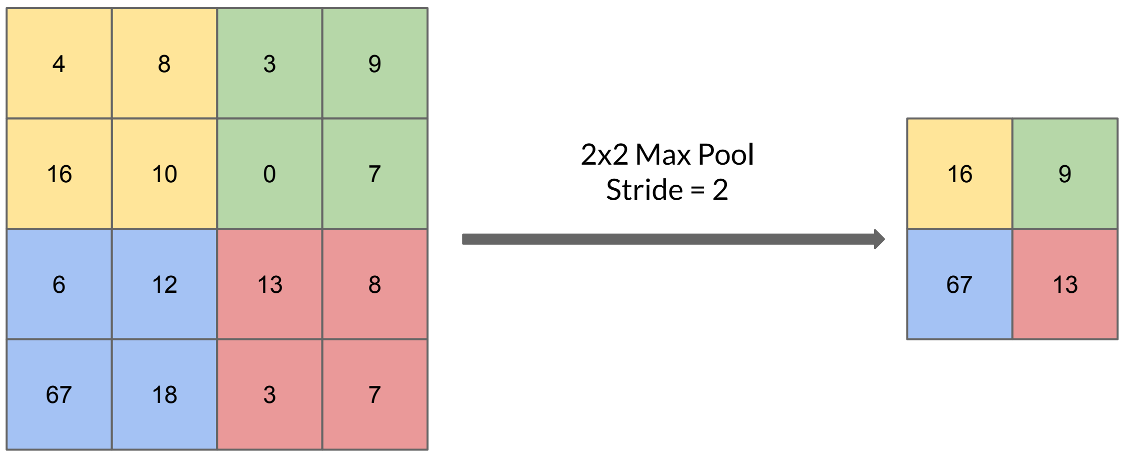


Pooling Layer

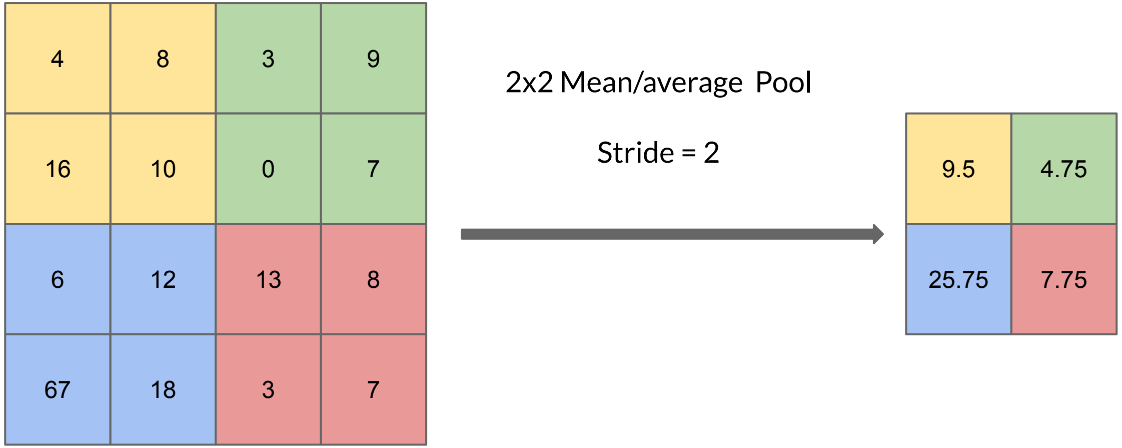
A pooling layer in a Convolutional Neural Network (CNN) is a fundamental component used to downsample the spatial dimensions of the feature maps produced by convolutional layers. Pooling layers are responsible for reducing the size of the feature maps while retaining the most important information.

- **Max-pooling** and **average-pooling** are common pooling operations.
- Introduces Jitter Invariance
- Reduces memory footprint by reducing the feature-map size

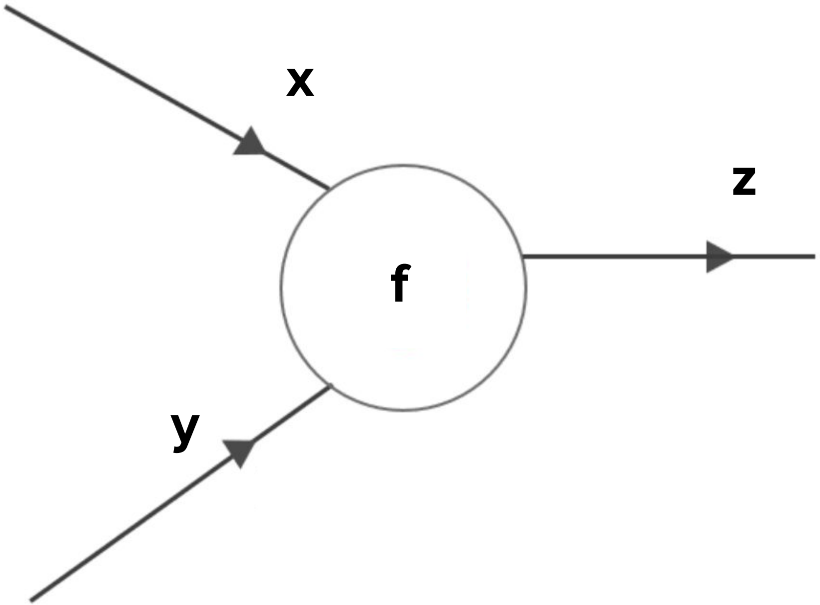
Pooling

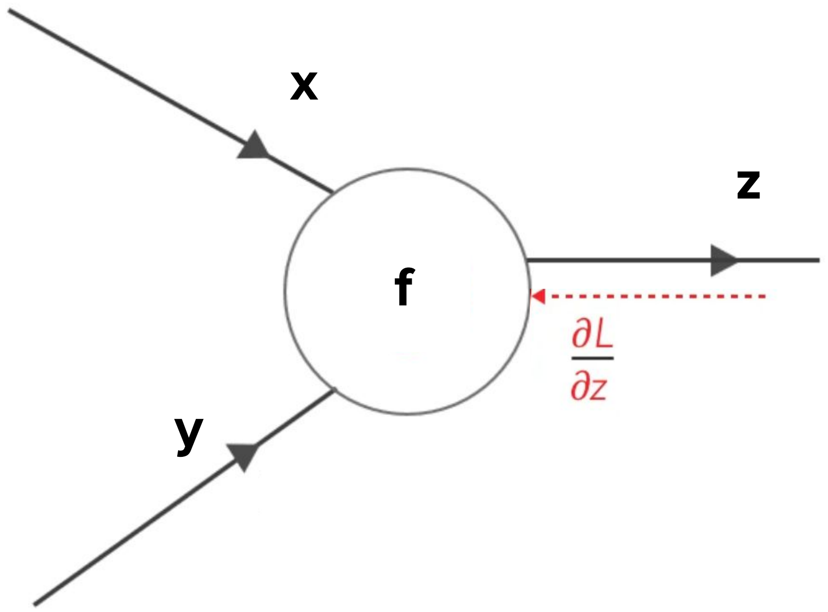


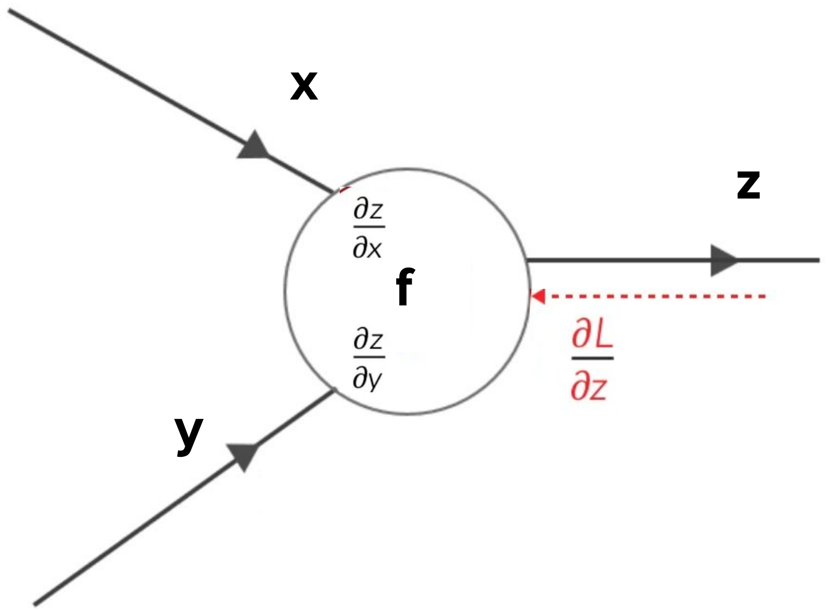
Pooling

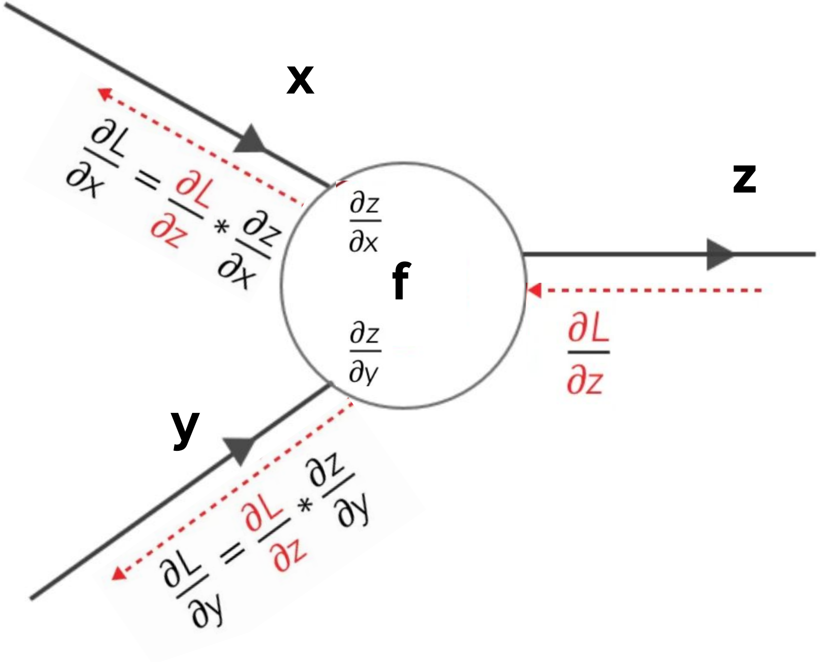

4
8
3
9
16
10
0
7
6
12
13
8
67
18
3
7
2x2 Mean/average Pool
Stride = 2
9.5
4.75
25.75
7.75

Backpropagation in CNN

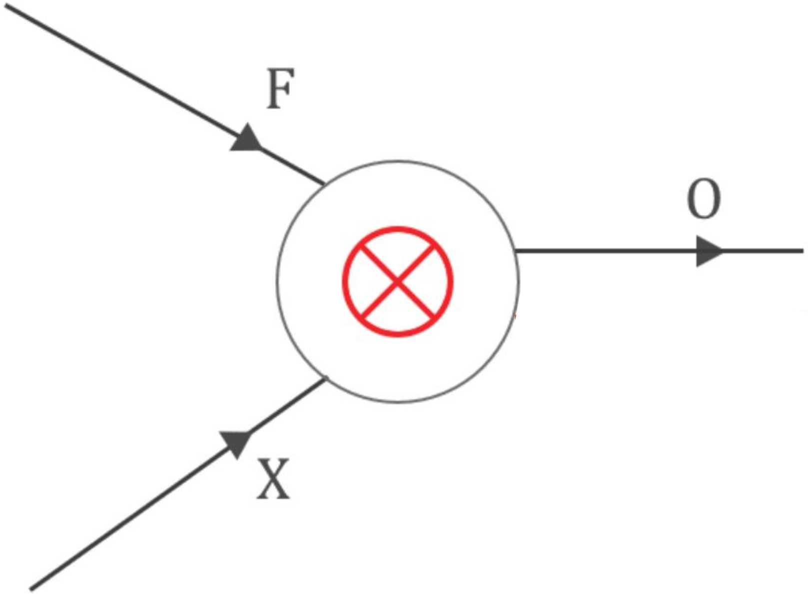

x
f
z
y

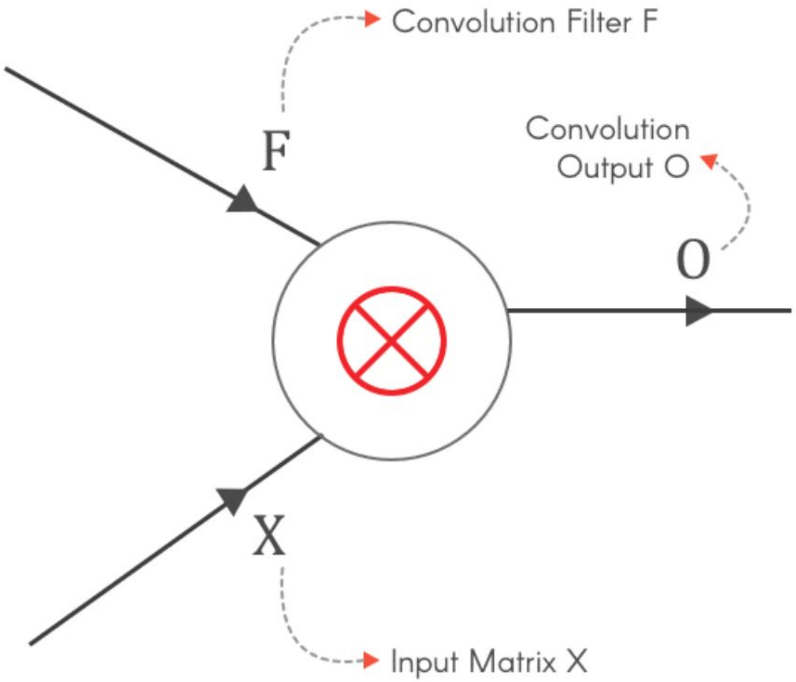

x
y
f
z
$\frac{\partial L}{\partial z}$


x
z
$\frac{\partial z}{\partial x}$
f
$\frac{\partial z}{\partial y}$
$\frac{\partial L}{\partial z}$
y

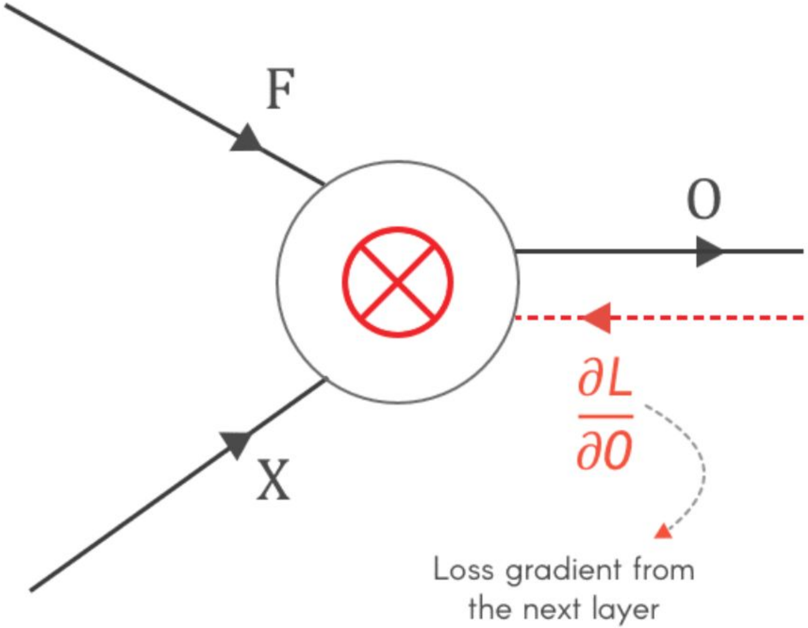


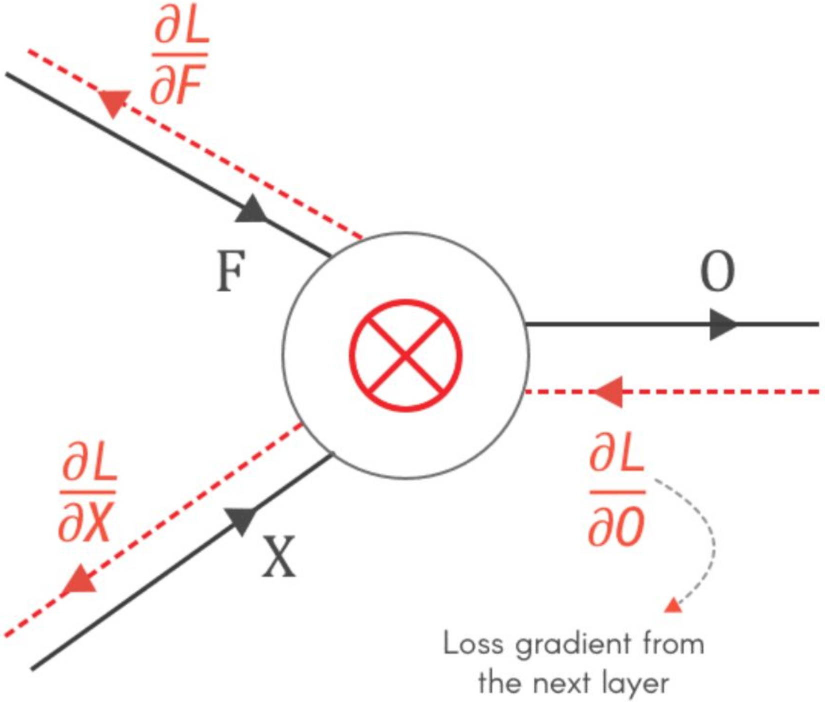
x
z
y
f
∂z/∂x
∂z/∂y
∂L/∂z
∂L/∂x = ∂L/∂z * ∂z/∂x
∂L/∂y = ∂L/∂z * ∂z/∂y

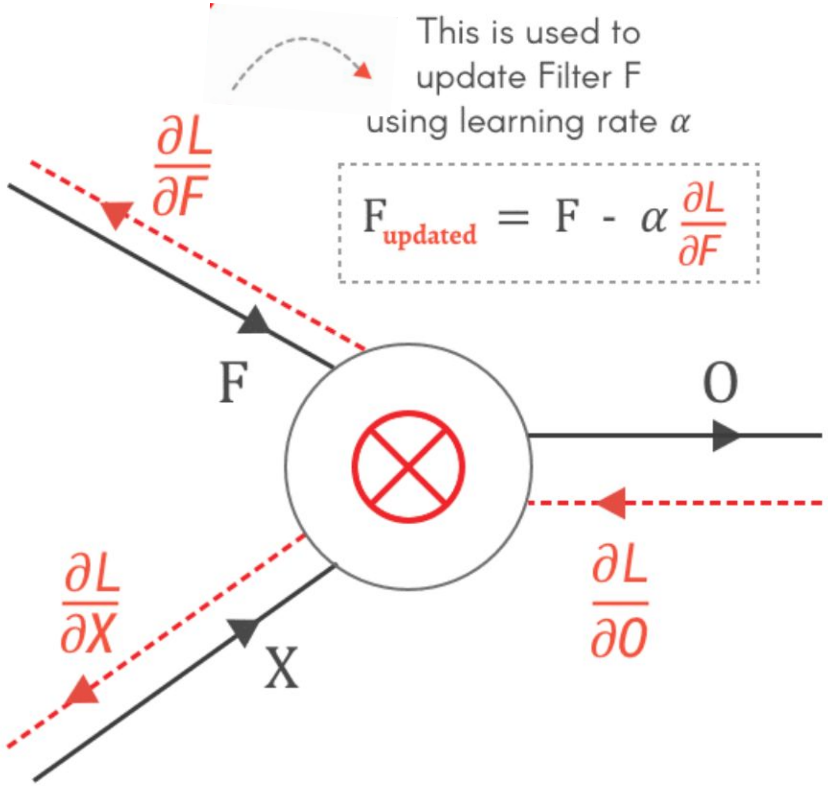

F
O
X

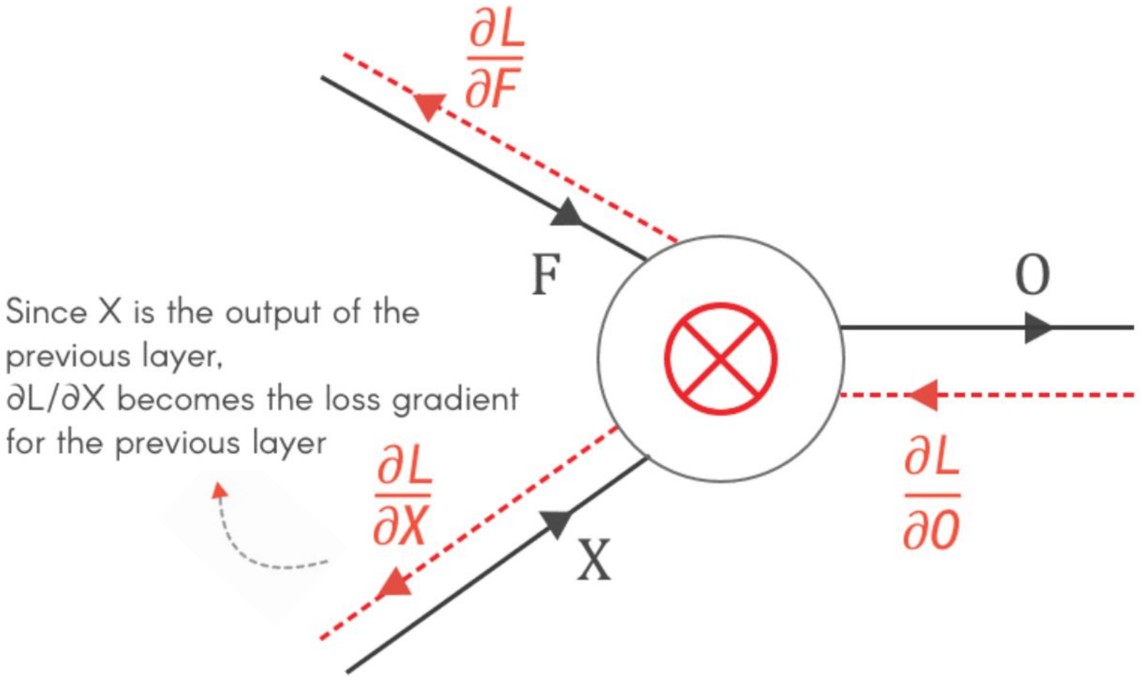


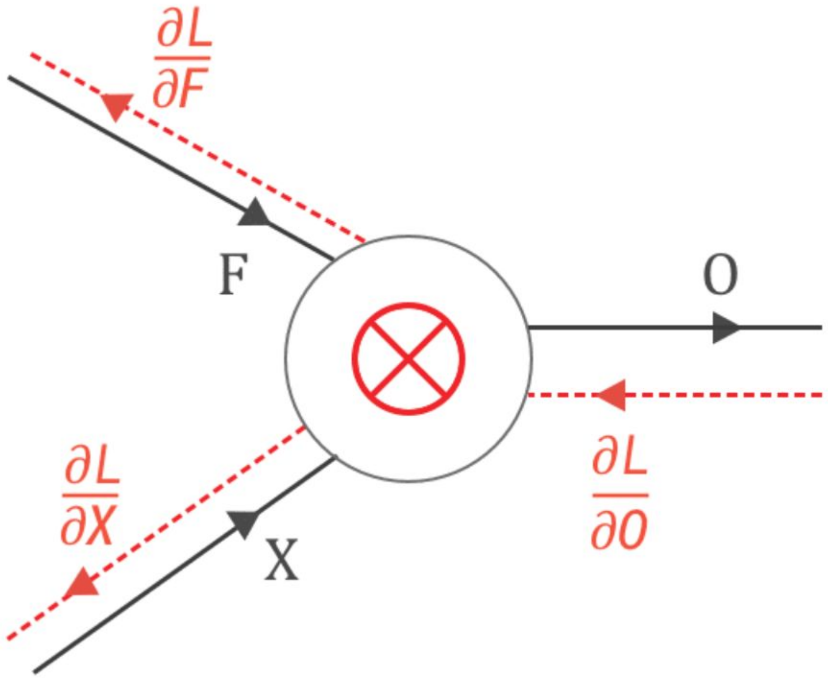
Convolution Filter F
F
Convolution
Output O
O
X
Input Matrix X


F
X
O
$\frac{\partial L}{\partial O}$
Loss gradient from the next layer

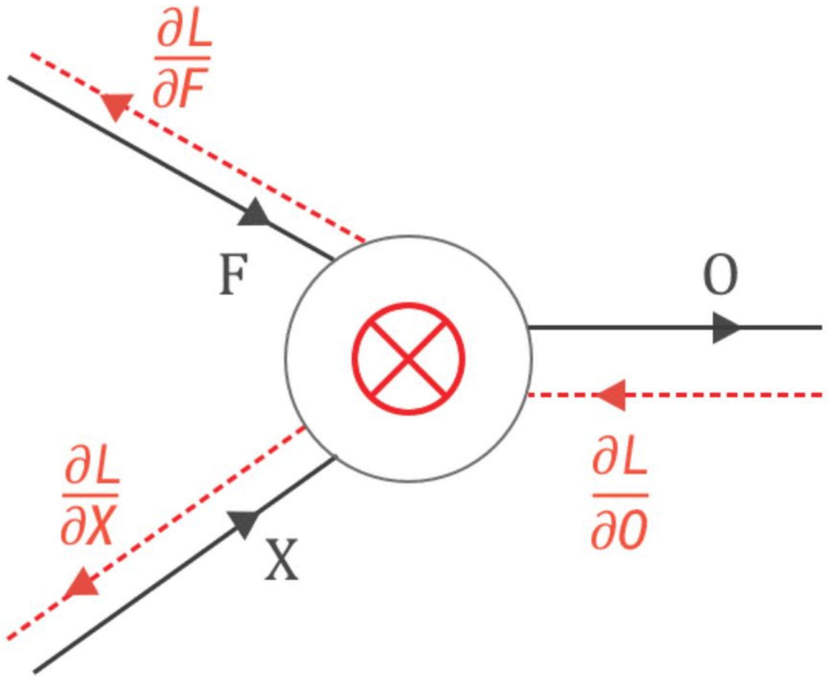

∂L/∂F
F
O
X
∂L/∂X
∂L/∂O
Loss gradient from the next layer


This is used to
update Filter F
using learning rate α
$F_{updated} = F - \alpha \frac{\partial L}{\partial F}$
$\frac{\partial L}{\partial F}$
F
O
X
$\frac{\partial L}{\partial X}$
$\frac{\partial L}{\partial O}$

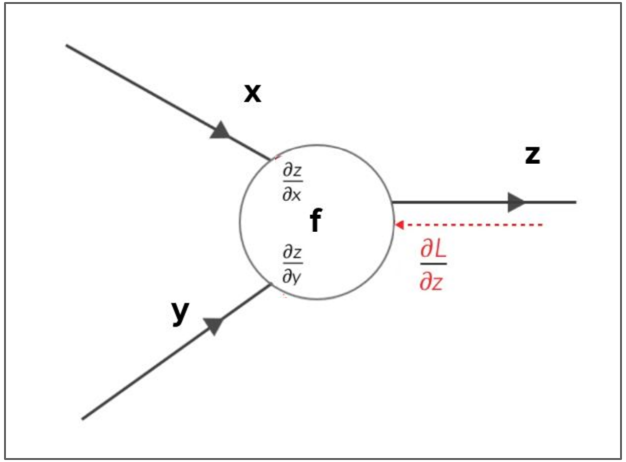

∂L/∂F
F
O
Since X is the output of the previous layer,
∂L/∂X becomes the loss gradient for the previous layer
∂L/∂X
X
∂L/∂O



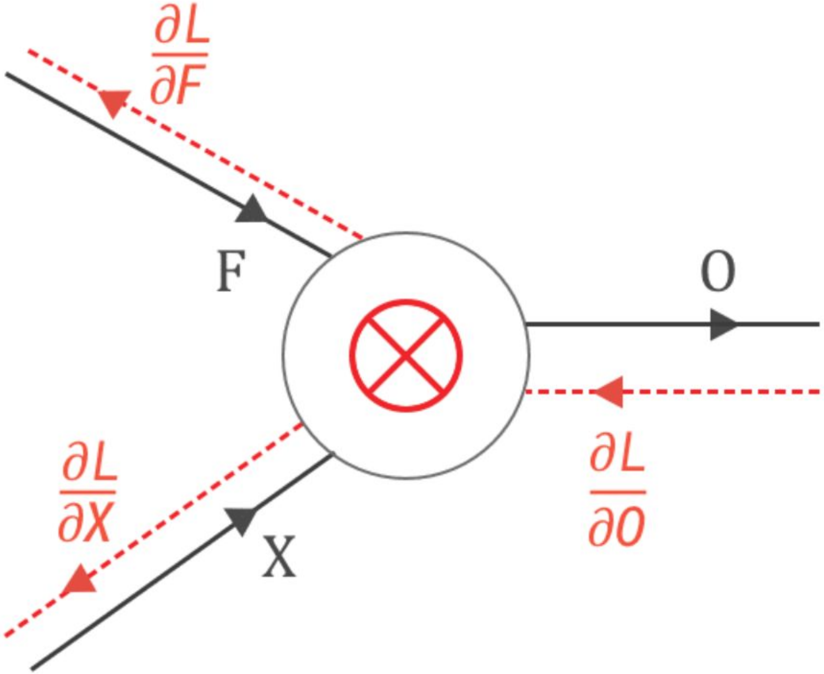
∂L/∂F
F
O
∂L/∂X
X
∂L/∂O



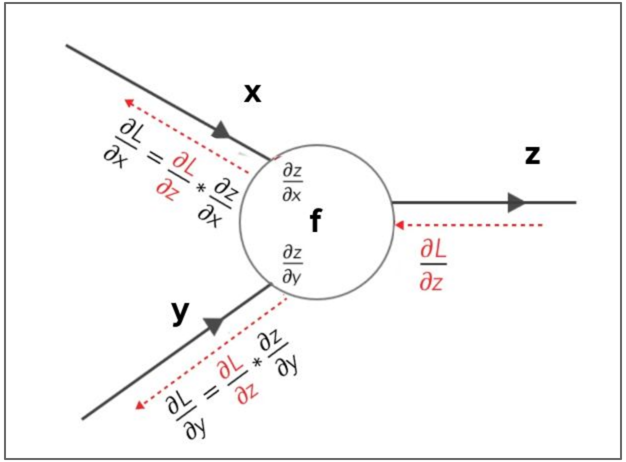
∂L/∂F
F
O
∂L/∂X
X
∂L/∂O



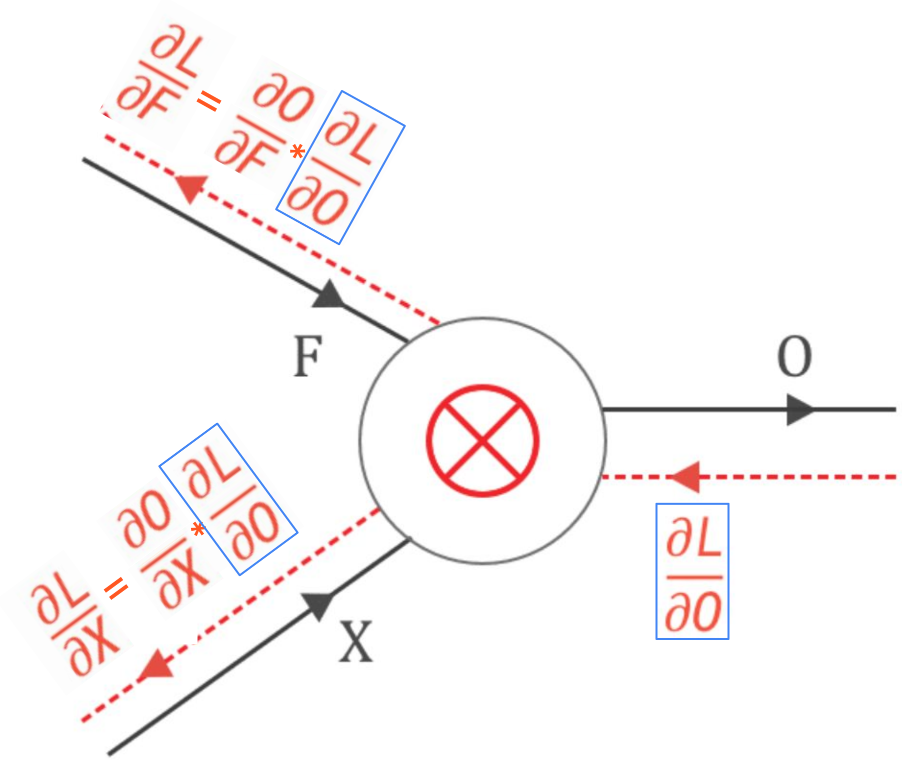
x
z
∂z/∂x
f
∂z/∂y
∂L/∂z
y



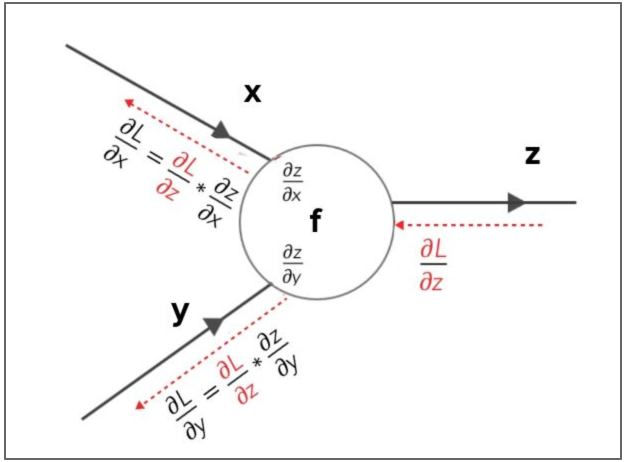
$\frac{\partial L}{\partial F}$
F
O
$\frac{\partial L}{\partial X}$
X
$\frac{\partial L}{\partial O}$



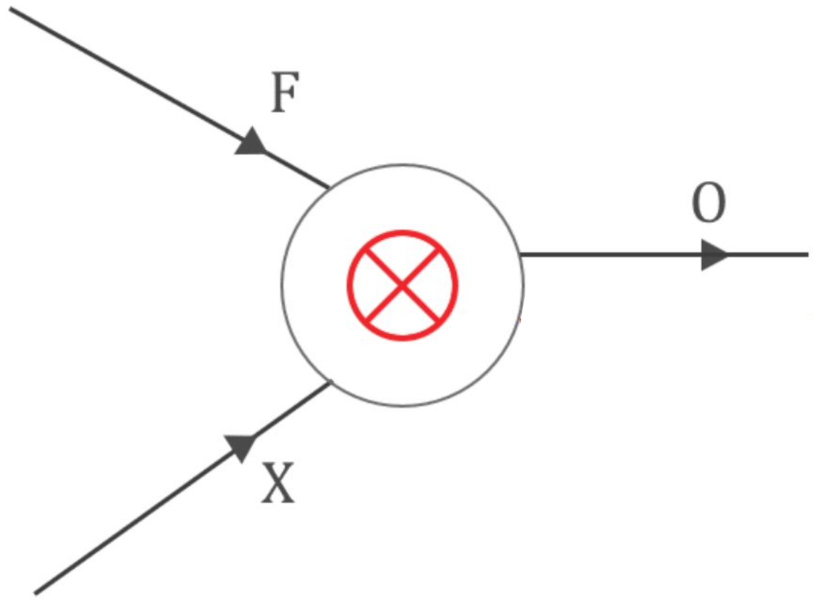
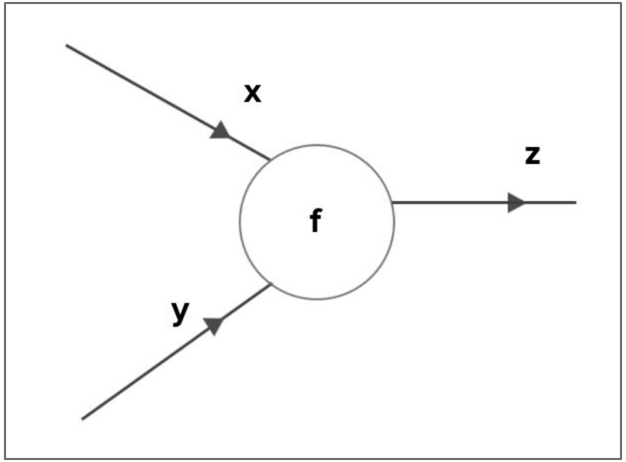
x
z
$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z} * \frac{\partial z}{\partial x}$
$\frac{\partial z}{\partial x}$
f
$\frac{\partial z}{\partial y}$
$\frac{\partial L}{\partial z}$
y
$\frac{\partial L}{\partial y} = \frac{\partial L}{\partial z} * \frac{\partial z}{\partial y}$

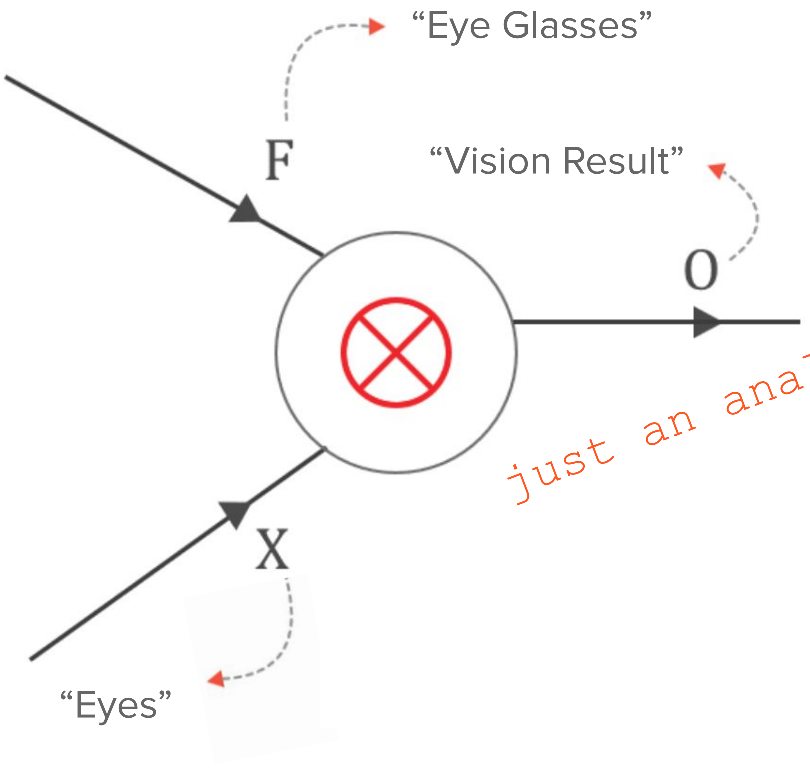


∂L/∂F = ∂O/∂F * ∂L/∂O
F
O
∂L/∂O
∂L/∂X = ∂O/∂X * ∂L/∂O
X

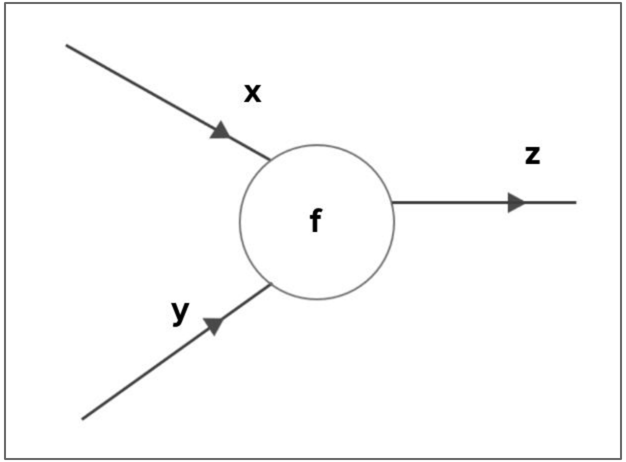


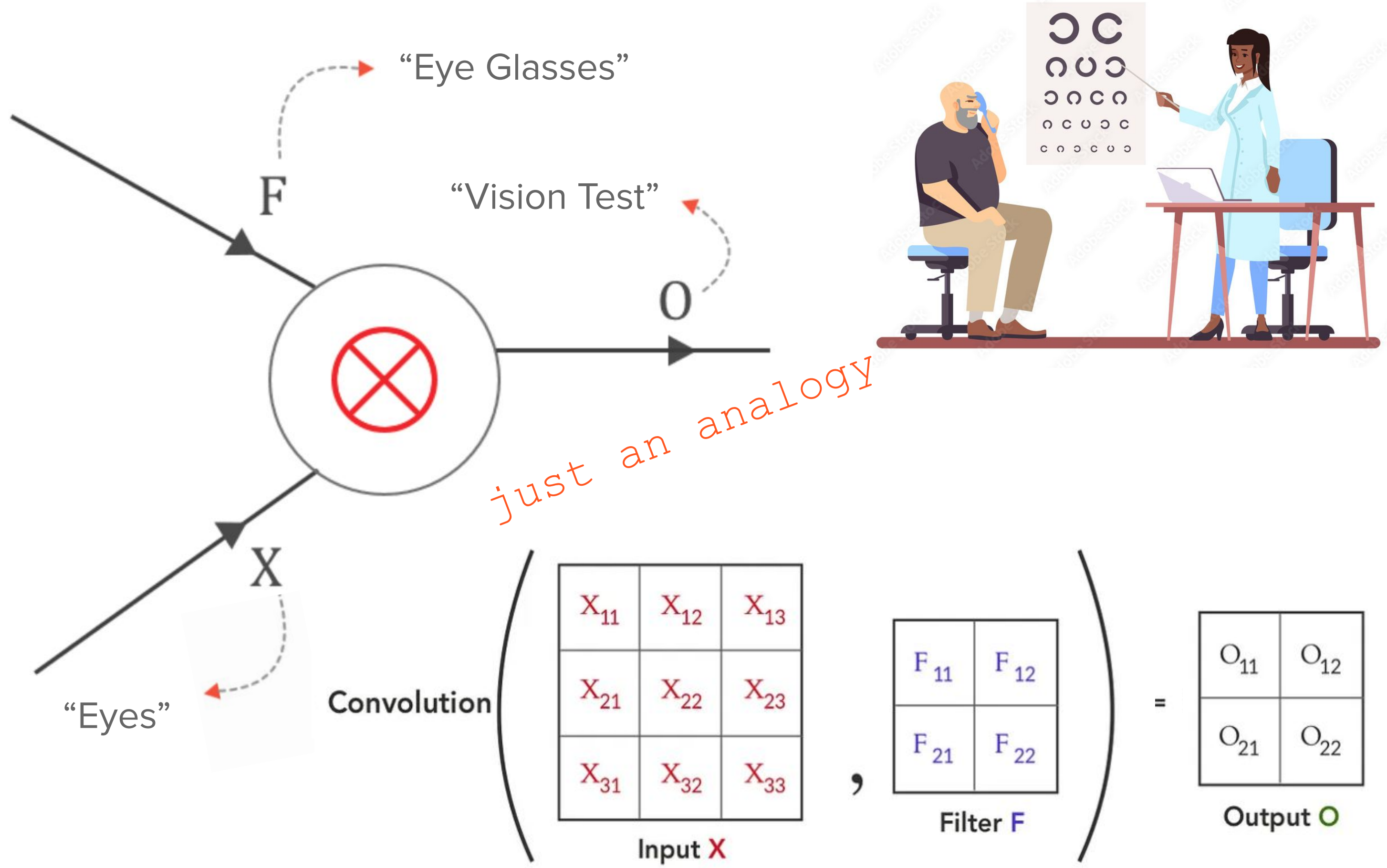
x
z
f
∂z/∂x
∂z/∂y
∂L/∂x = ∂L/∂z * ∂z/∂x
∂L/∂z
y
∂L/∂y = ∂L/∂z * ∂z/∂y


F
X
O

x
y
f
z

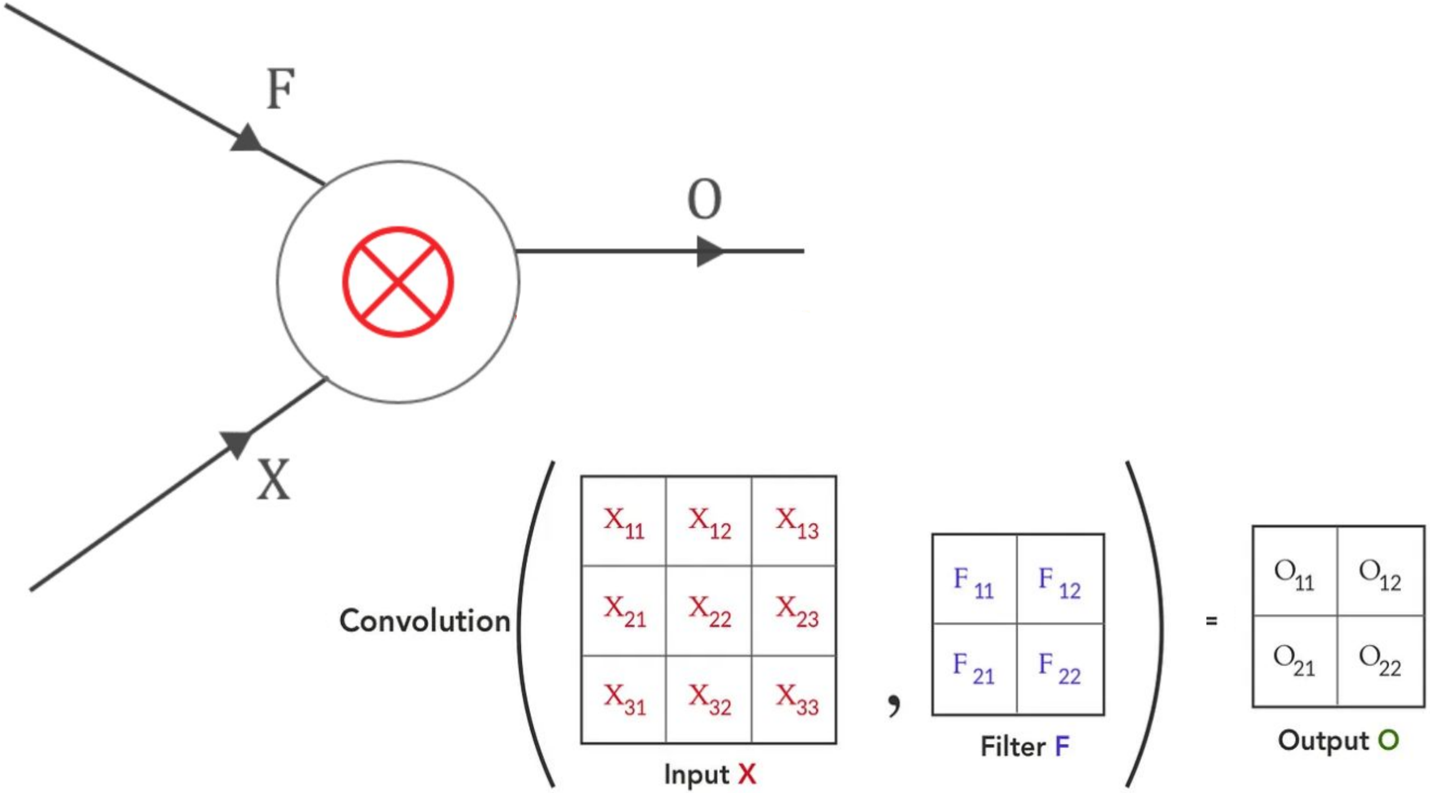


just an analogy

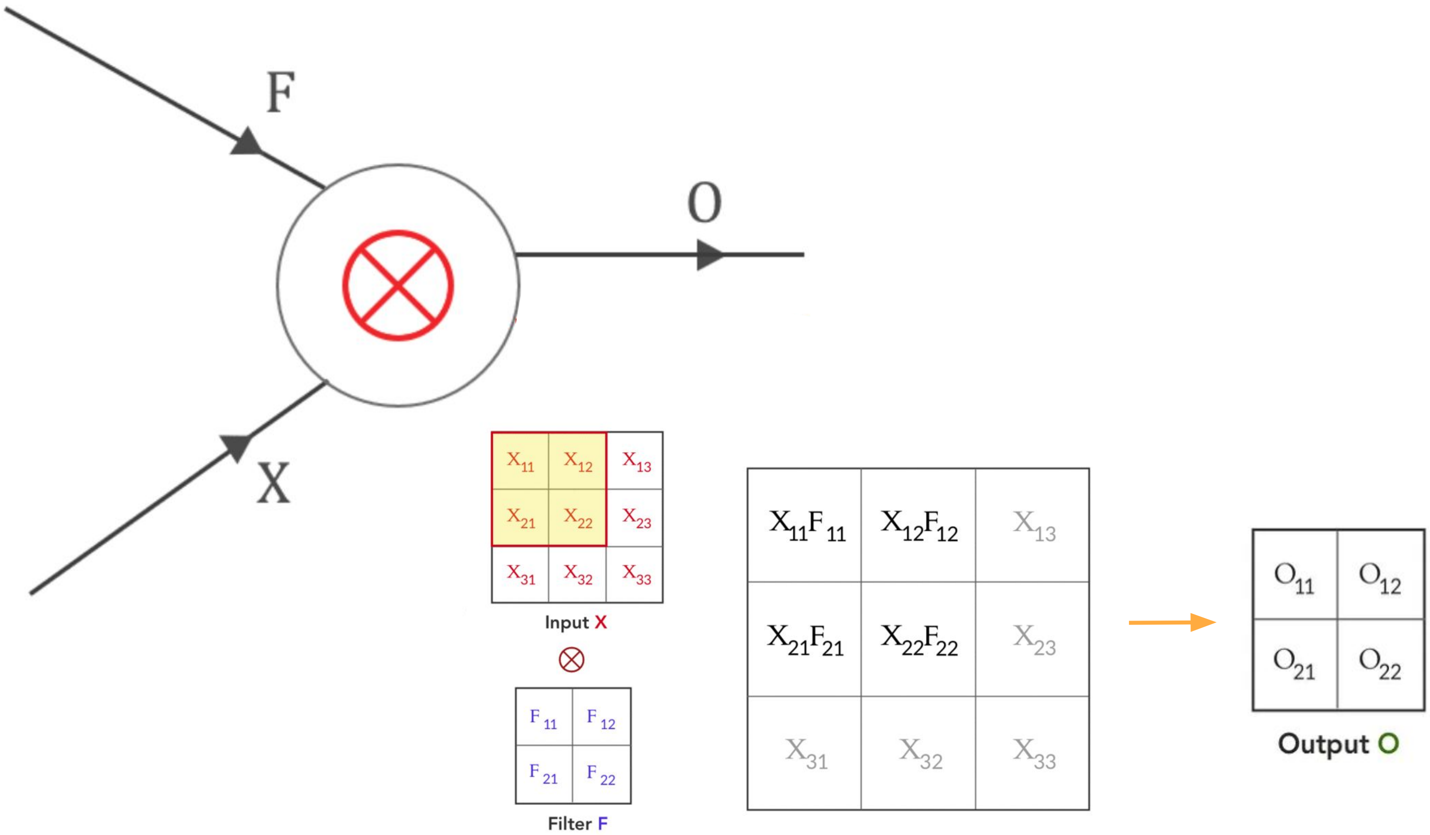




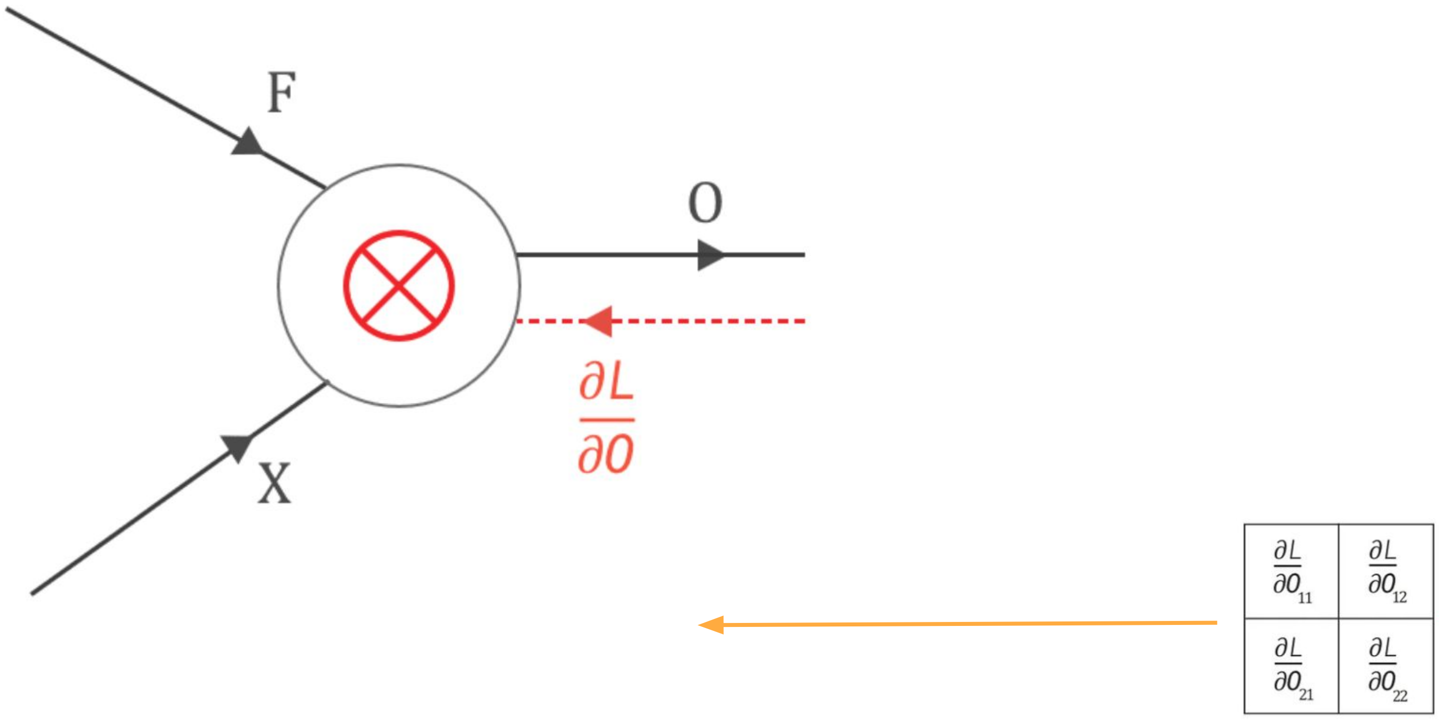
“Eye Glasses”
F
“Vision Test”
O
X
“Eyes”
just an analogy
Convolution
X11 X12 X13
X21 X22 X23
X31 X32 X33
Input X
F 11 F 12
F 21 F 22
Filter F
=
O11 O12
O21 O22
Output O

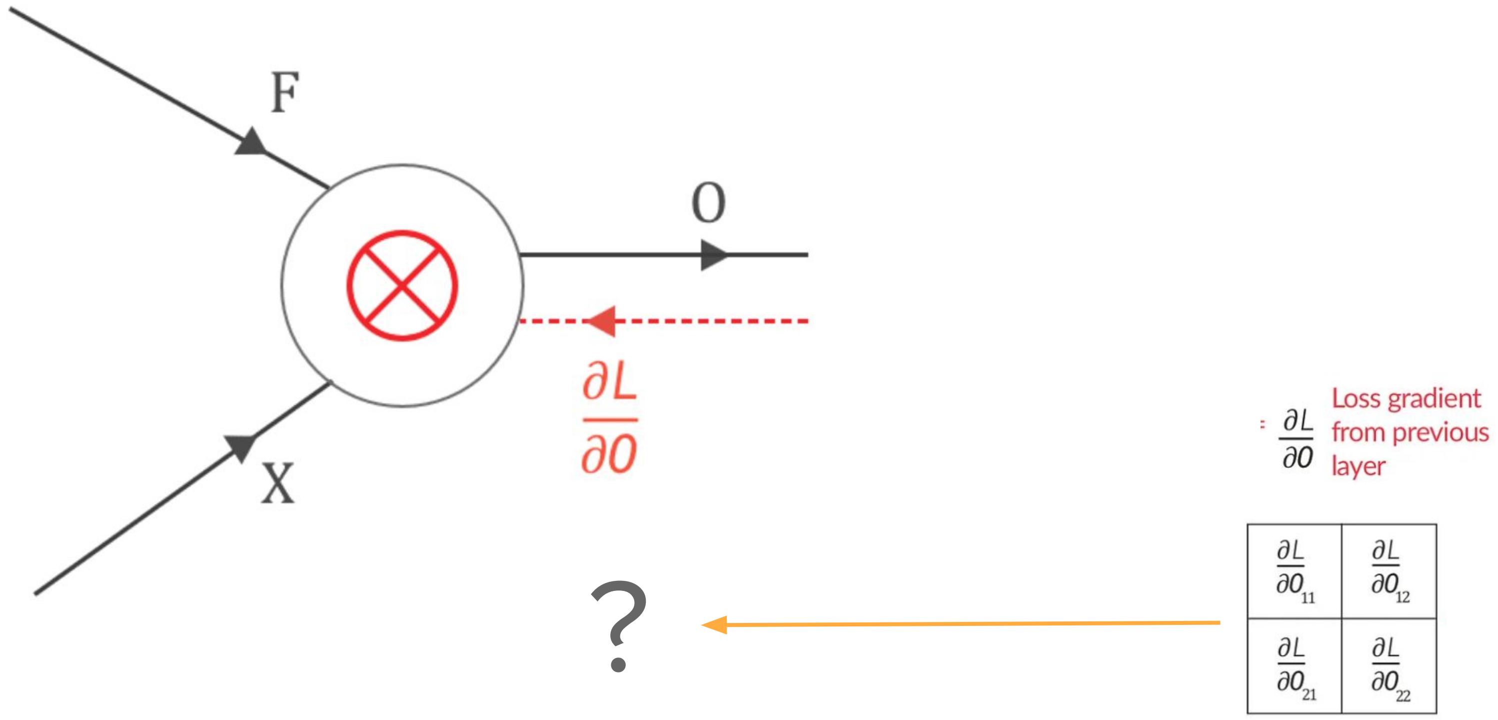


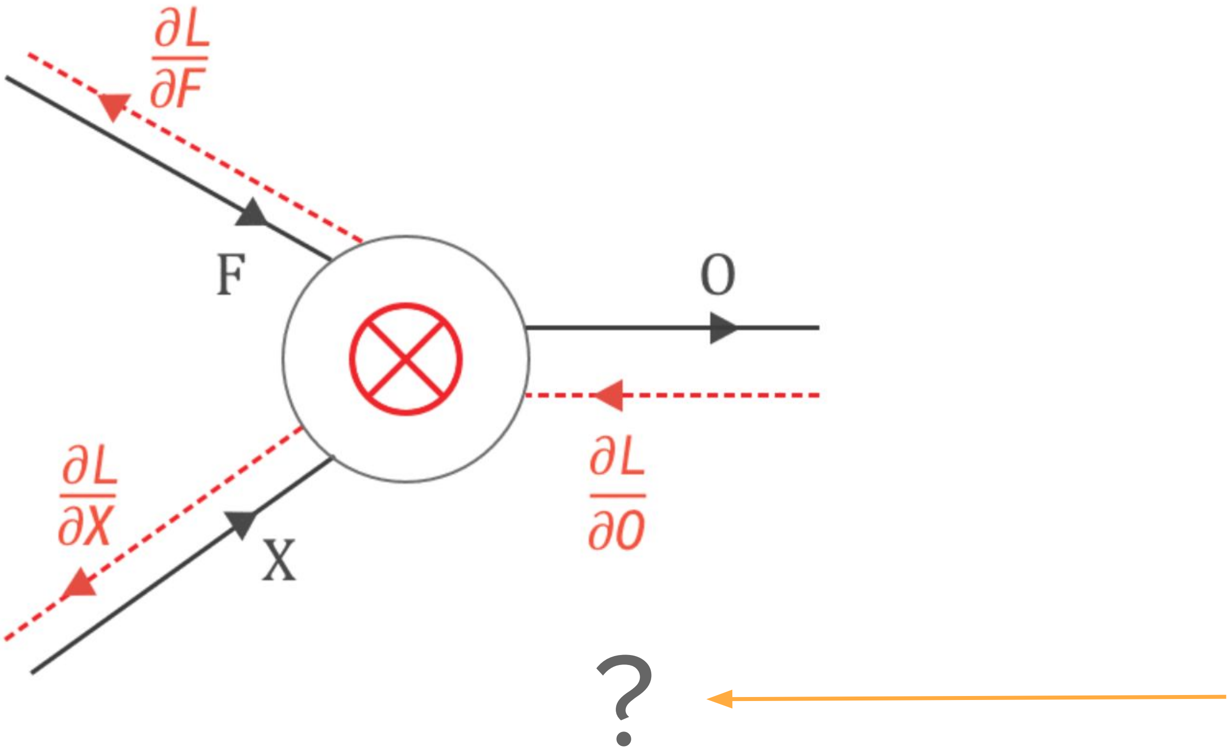
F
O
X
Convolution
X_{11} X_{12} X_{13}
X_{21} X_{22} X_{23}
X_{31} X_{32} X_{33}
Input X
F_{11} F_{12}
F_{21} F_{22}
Filter F
=
O_{11} O_{12}
O_{21} O_{22}
Output O



F
O
X
X11 X12 X13
X21 X22 X23
X31 X32 X33
Input X
F 11 F 12
F 21 F 22
Filter F
X11F11 X12F12 X13
X21F21 X22F22 X23
X31 X32 X33
O11 O12
O21 O22
Output O

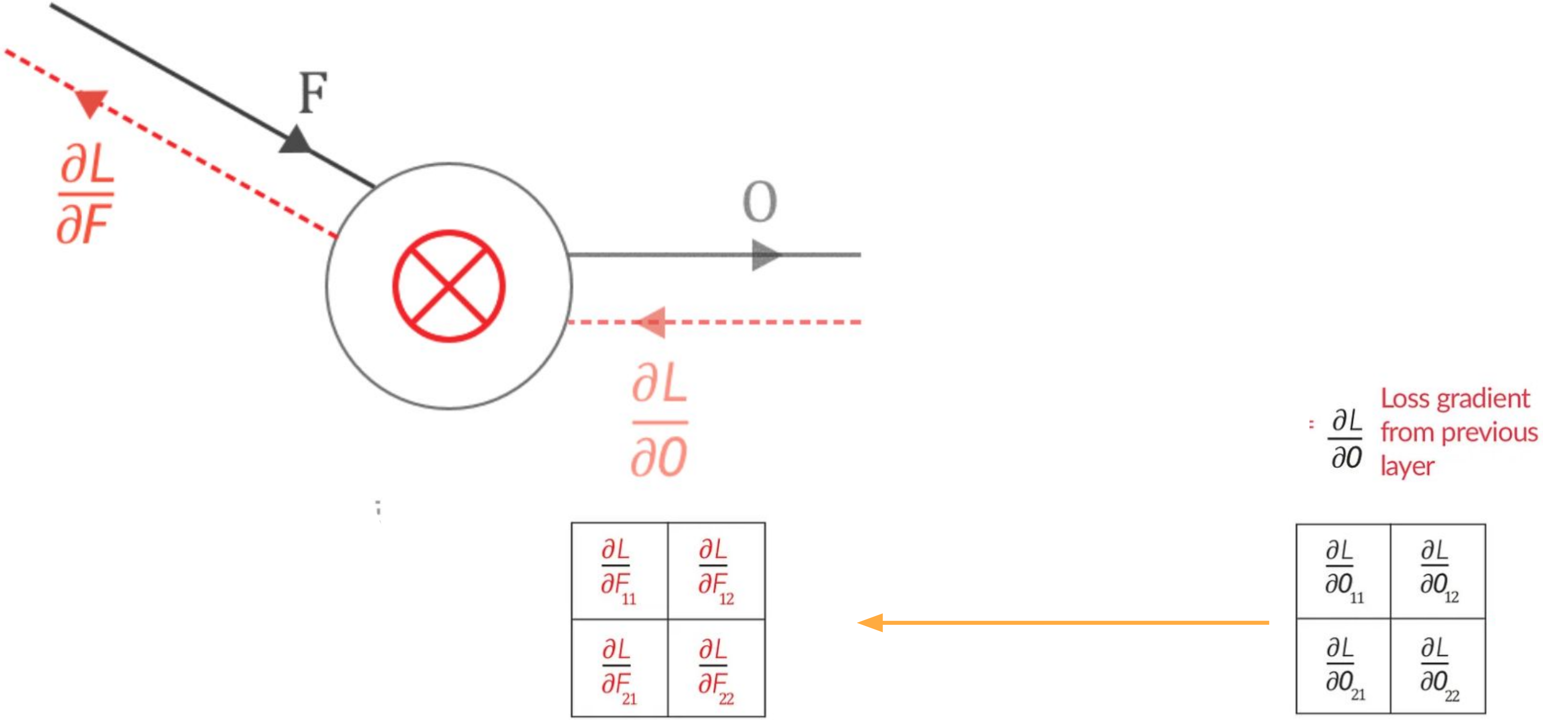

F
X
O
∂L/∂O
∂L/∂O11
∂L/∂O12
∂L/∂O21
∂L/∂O22


F
X
O
∂L/∂O
?
∂L/∂O
Loss gradient from previous layer
∂L/∂O11
∂L/∂O12
∂L/∂O21
∂L/∂O22

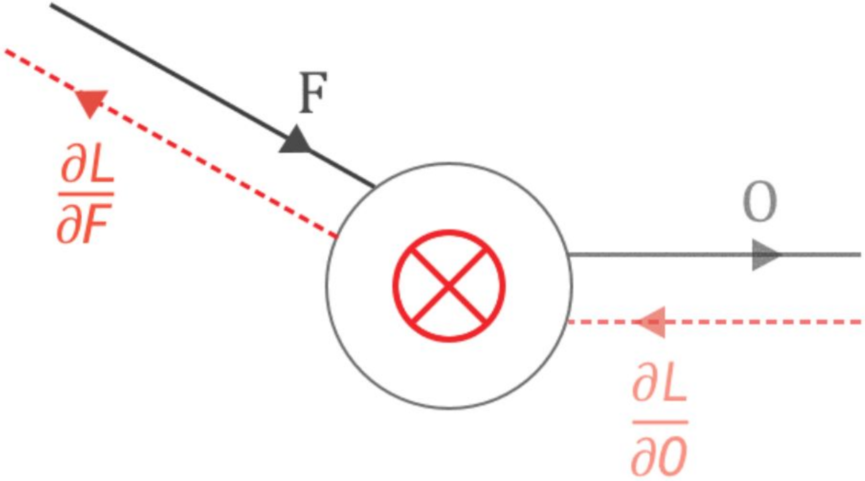


$\frac{\partial L}{\partial O}$: Loss gradient from previous layer

$\frac{\partial L}{\partial O_{11}}$	$\frac{\partial L}{\partial O_{12}}$
$\frac{\partial L}{\partial O_{21}}$	$\frac{\partial L}{\partial O_{22}}$


F
∂L/∂F
O
∂L/∂O
∂L/∂O Loss gradient from previous layer
∂L/∂F11 ∂L/∂F12
∂L/∂F21 ∂L/∂F22
∂L/∂O11 ∂L/∂O12
∂L/∂O21 ∂L/∂O22

$$\frac{\partial L}{\partial F} = \frac{\partial O}{\partial F} * \frac{\partial L}{\partial O}$$



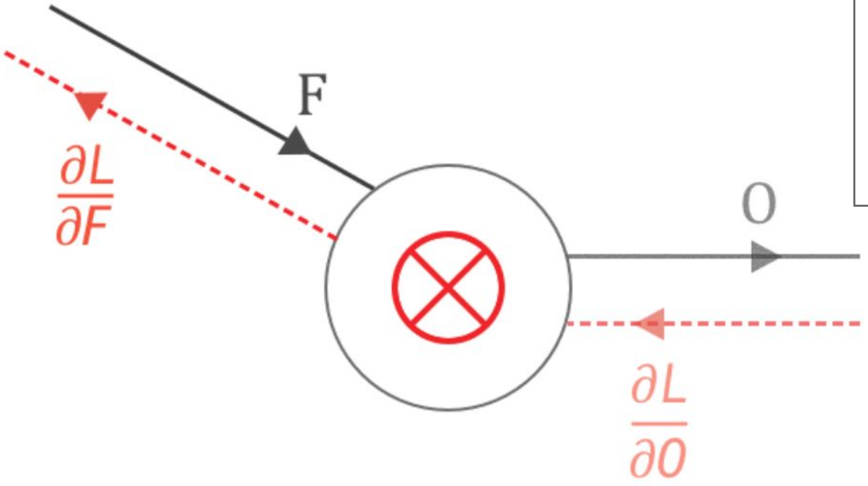
$\frac{\partial L}{\partial O}$: Loss gradient from previous layer

$\frac{\partial L}{\partial F_{11}}$	$\frac{\partial L}{\partial F_{12}}$
$\frac{\partial L}{\partial F_{21}}$	$\frac{\partial L}{\partial F_{22}}$

←

$\frac{\partial L}{\partial O_{11}}$	$\frac{\partial L}{\partial O_{12}}$
$\frac{\partial L}{\partial O_{21}}$	$\frac{\partial L}{\partial O_{22}}$

$$\frac{\partial L}{\partial F} = \frac{\partial O}{\partial F} * \frac{\partial L}{\partial O}$$



For every element of F

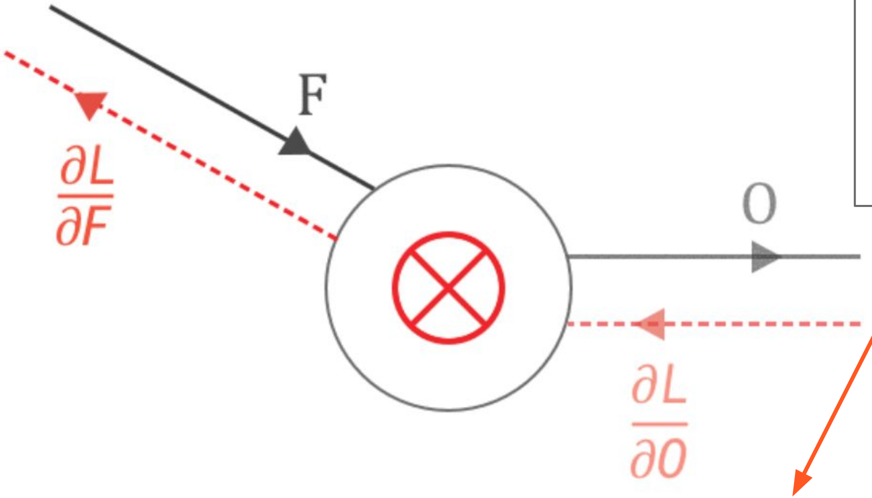
$$\frac{\partial L}{\partial F_i} = \sum_{k=1}^{M} \frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial F_i}$$

$\frac{\partial L}{\partial O}$: Loss gradient from previous layer

$\frac{\partial L}{\partial F_{11}}$	$\frac{\partial L}{\partial F_{12}}$
$\frac{\partial L}{\partial F_{21}}$	$\frac{\partial L}{\partial F_{22}}$

$\frac{\partial L}{\partial O_{11}}$	$\frac{\partial L}{\partial O_{12}}$
$\frac{\partial L}{\partial O_{21}}$	$\frac{\partial L}{\partial O_{22}}$

$$\frac{\partial L}{\partial F} = \frac{\partial O}{\partial F} * \frac{\partial L}{\partial O}$$

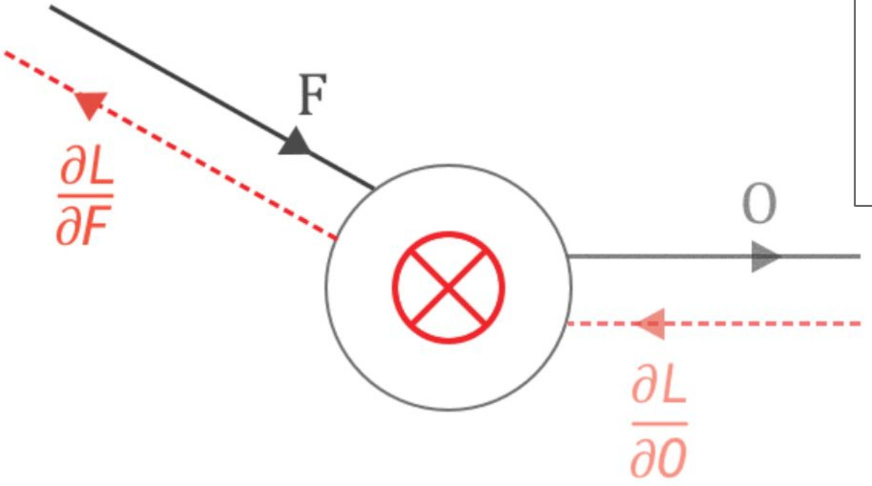


For every element of F

$$\frac{\partial L}{\partial F_i} = \sum_{k=1}^{M} \frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial F_i}$$

$$\begin{bmatrix} \frac{\partial L}{\partial F_{11}} & \frac{\partial L}{\partial F_{12}} \\ \frac{\partial L}{\partial F_{21}} & \frac{\partial L}{\partial F_{22}} \end{bmatrix} = \text{Convolution}\left(\qquad , \begin{bmatrix} \frac{\partial L}{\partial O_{11}} & \frac{\partial L}{\partial O_{12}} \\ \frac{\partial L}{\partial O_{21}} & \frac{\partial L}{\partial O_{22}} \end{bmatrix} \right)$$

$$\frac{\partial L}{\partial F} = \frac{\partial O}{\partial F} * \frac{\partial L}{\partial O}$$

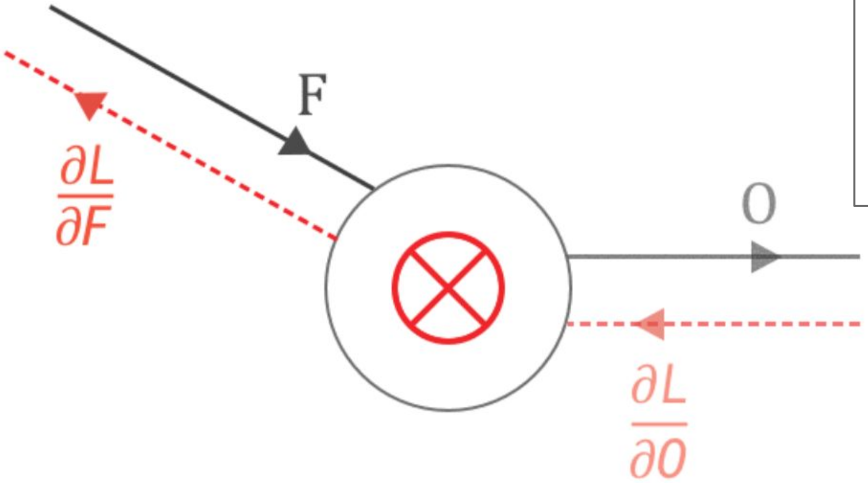


For every element of F

$$\frac{\partial L}{\partial F_i} = \sum_{k=1}^{M} \frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial F_i}$$

$$\begin{bmatrix} \frac{\partial L}{\partial F_{11}} & \frac{\partial L}{\partial F_{12}} \\ \frac{\partial L}{\partial F_{21}} & \frac{\partial L}{\partial F_{22}} \end{bmatrix} = \text{Convolution}\left(? \; , \begin{bmatrix} \frac{\partial L}{\partial O_{11}} & \frac{\partial L}{\partial O_{12}} \\ \frac{\partial L}{\partial O_{21}} & \frac{\partial L}{\partial O_{22}} \end{bmatrix} \right)$$

$$\frac{\partial L}{\partial F} = \frac{\partial O}{\partial F} * \frac{\partial L}{\partial O}$$



For every element of F

$$\frac{\partial L}{\partial F_i} = \sum_{k=1}^{M} \frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial F_i}$$

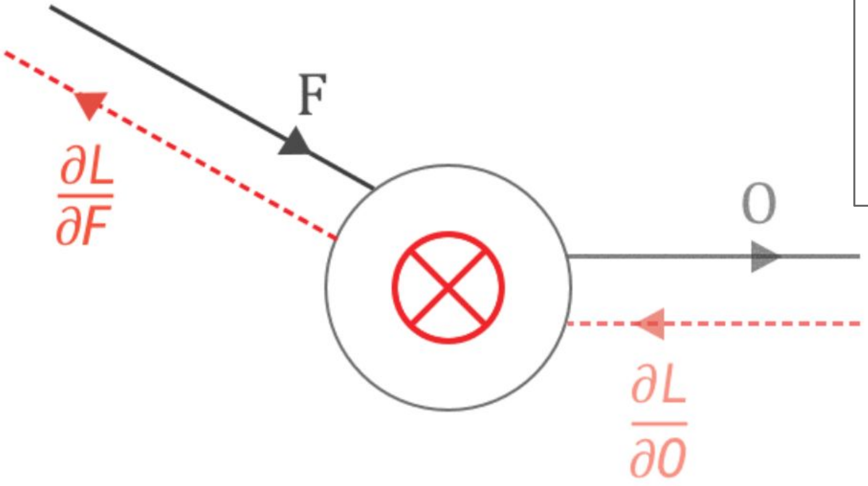
Hint:

$$O_{11} = X_{11}F_{11} + X_{12}F_{12} + X_{21}F_{21} + X_{22}F_{22} \cdots$$

$$\frac{\partial O_{11}}{\partial F_{11}} = X_{11} \quad \frac{\partial O_{11}}{\partial F_{12}} = X_{12} \quad \frac{\partial O_{11}}{\partial F_{21}} = X_{21} \quad \frac{\partial O_{11}}{\partial F_{22}} = X_{22} \cdots$$

$$\begin{bmatrix} \frac{\partial L}{\partial F_{11}} & \frac{\partial L}{\partial F_{12}} \\ \frac{\partial L}{\partial F_{21}} & \frac{\partial L}{\partial F_{22}} \end{bmatrix} = \text{Convolution}\left(\;?\; , \begin{bmatrix} \frac{\partial L}{\partial O_{11}} & \frac{\partial L}{\partial O_{12}} \\ \frac{\partial L}{\partial O_{21}} & \frac{\partial L}{\partial O_{22}} \end{bmatrix} \right)$$

$$\frac{\partial L}{\partial F} = \frac{\partial O}{\partial F} * \frac{\partial L}{\partial O}$$



For every element of F

$$\frac{\partial L}{\partial F_i} = \sum_{k=1}^{M} \frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial F_i}$$

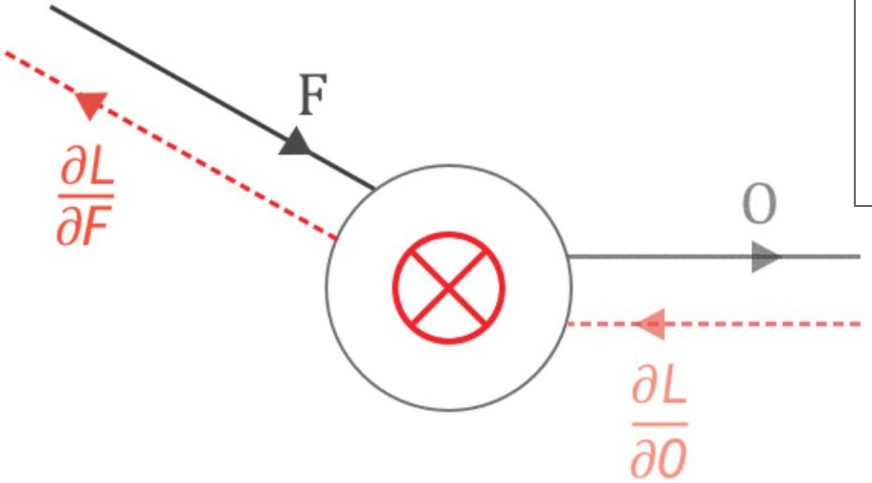
Hint:

$$O_{11} = X_{11}F_{11} + X_{12}F_{12} + X_{21}F_{21} + X_{22}F_{22} \cdots$$

$$\frac{\partial O_{11}}{\partial F_{11}} = X_{11} \quad \frac{\partial O_{11}}{\partial F_{12}} = X_{12} \quad \frac{\partial O_{11}}{\partial F_{21}} = X_{21} \quad \frac{\partial O_{11}}{\partial F_{22}} = X_{22} \cdots$$

$$\begin{bmatrix} \frac{\partial L}{\partial F_{11}} & \frac{\partial L}{\partial F_{12}} \\ \frac{\partial L}{\partial F_{21}} & \frac{\partial L}{\partial F_{22}} \end{bmatrix} = \text{Convolution}\left(X \ , \begin{bmatrix} \frac{\partial L}{\partial O_{11}} & \frac{\partial L}{\partial O_{12}} \\ \frac{\partial L}{\partial O_{21}} & \frac{\partial L}{\partial O_{22}} \end{bmatrix} \right)$$

$$\frac{\partial L}{\partial F} = \frac{\partial O}{\partial F} * \frac{\partial L}{\partial O}$$



For every element of F

$$\frac{\partial L}{\partial F_i} = \sum_{k=1}^{M} \frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial F_i}$$

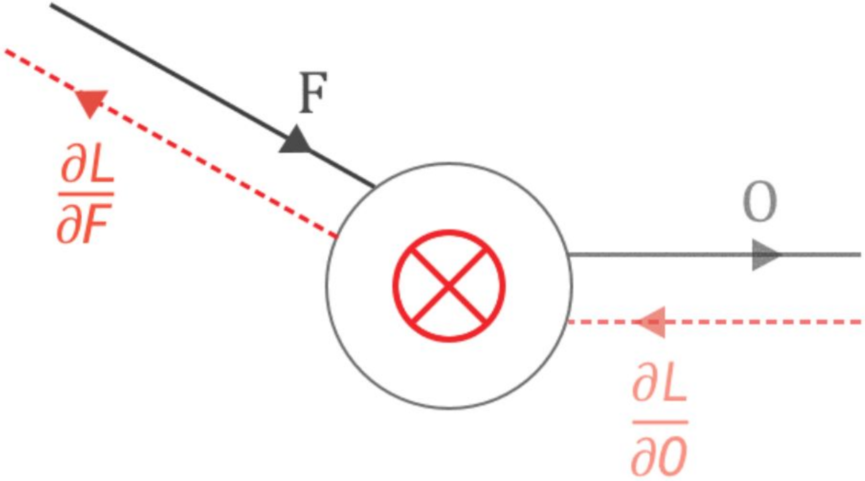
Hint:

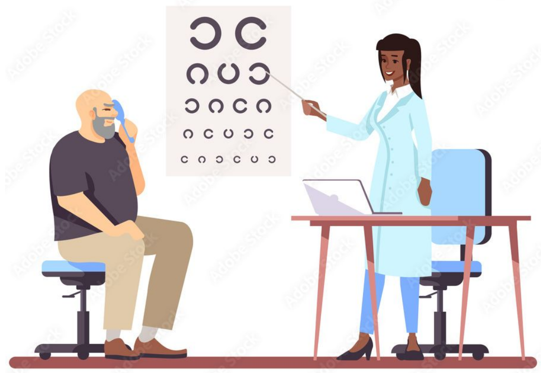
$$O_{11} = X_{11}F_{11} + X_{12}F_{12} + X_{21}F_{21} + X_{22}F_{22} \cdots$$

$$\frac{\partial O_{11}}{\partial F_{11}} = X_{11} \quad \frac{\partial O_{11}}{\partial F_{12}} = X_{12} \quad \frac{\partial O_{11}}{\partial F_{21}} = X_{21} \quad \frac{\partial O_{11}}{\partial F_{22}} = X_{22} \cdots$$

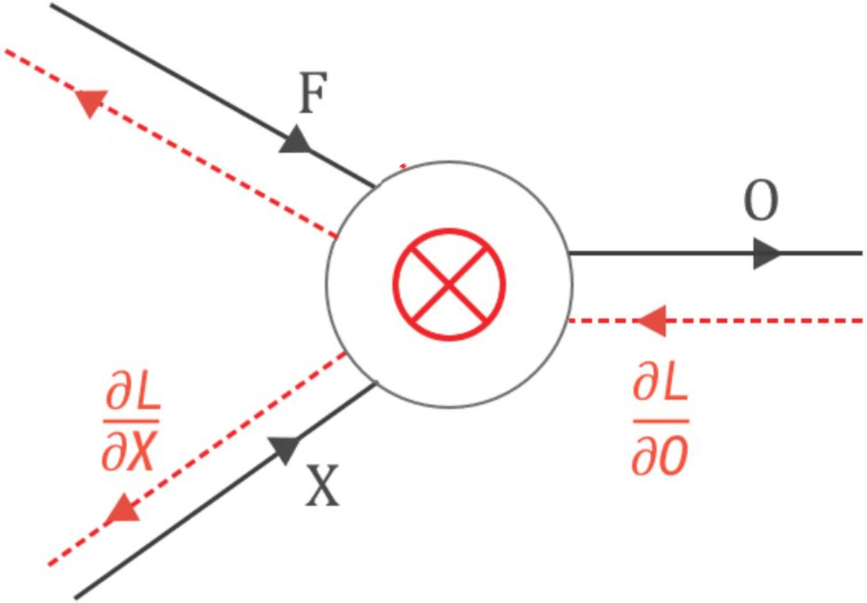
$$\begin{bmatrix} \frac{\partial L}{\partial F_{11}} & \frac{\partial L}{\partial F_{12}} \\ \frac{\partial L}{\partial F_{21}} & \frac{\partial L}{\partial F_{22}} \end{bmatrix} = \text{Convolution}\left(\begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \end{bmatrix}, \begin{bmatrix} \frac{\partial L}{\partial O_{11}} & \frac{\partial L}{\partial O_{12}} \\ \frac{\partial L}{\partial O_{21}} & \frac{\partial L}{\partial O_{22}} \end{bmatrix} \right)$$

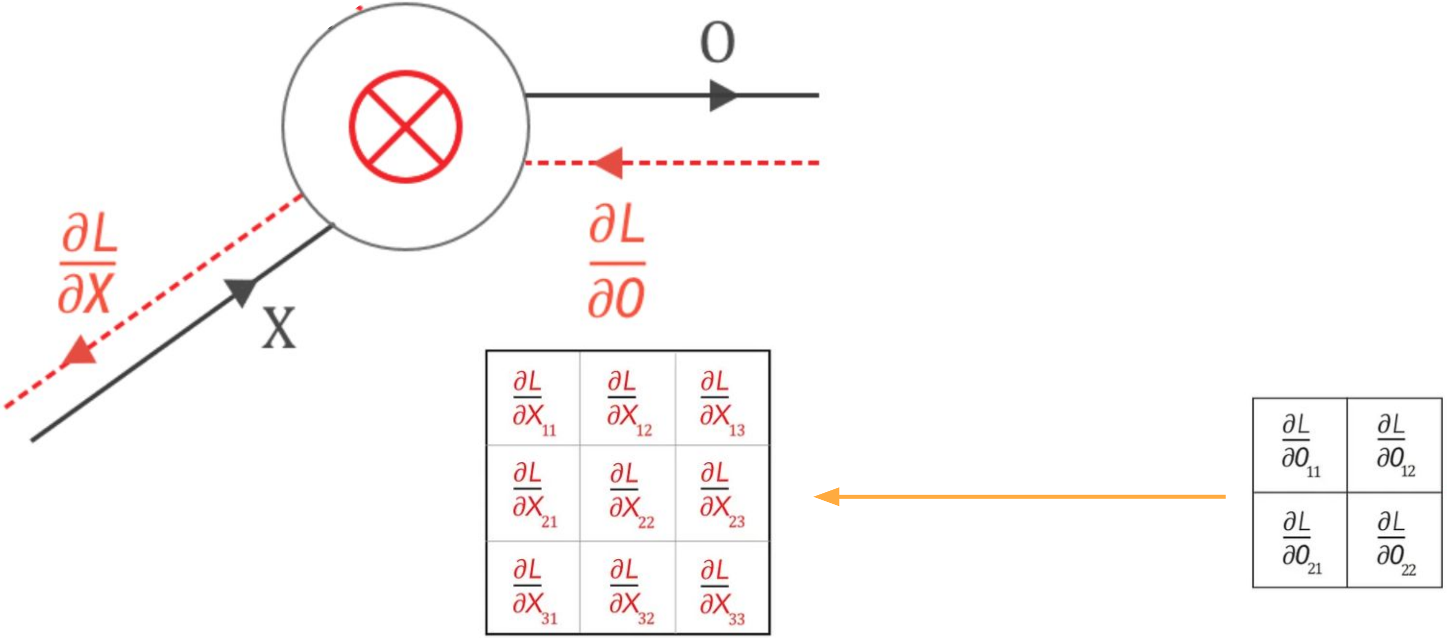
$$\frac{\partial L}{\partial F} = \frac{\partial O}{\partial F} * \frac{\partial L}{\partial O}$$



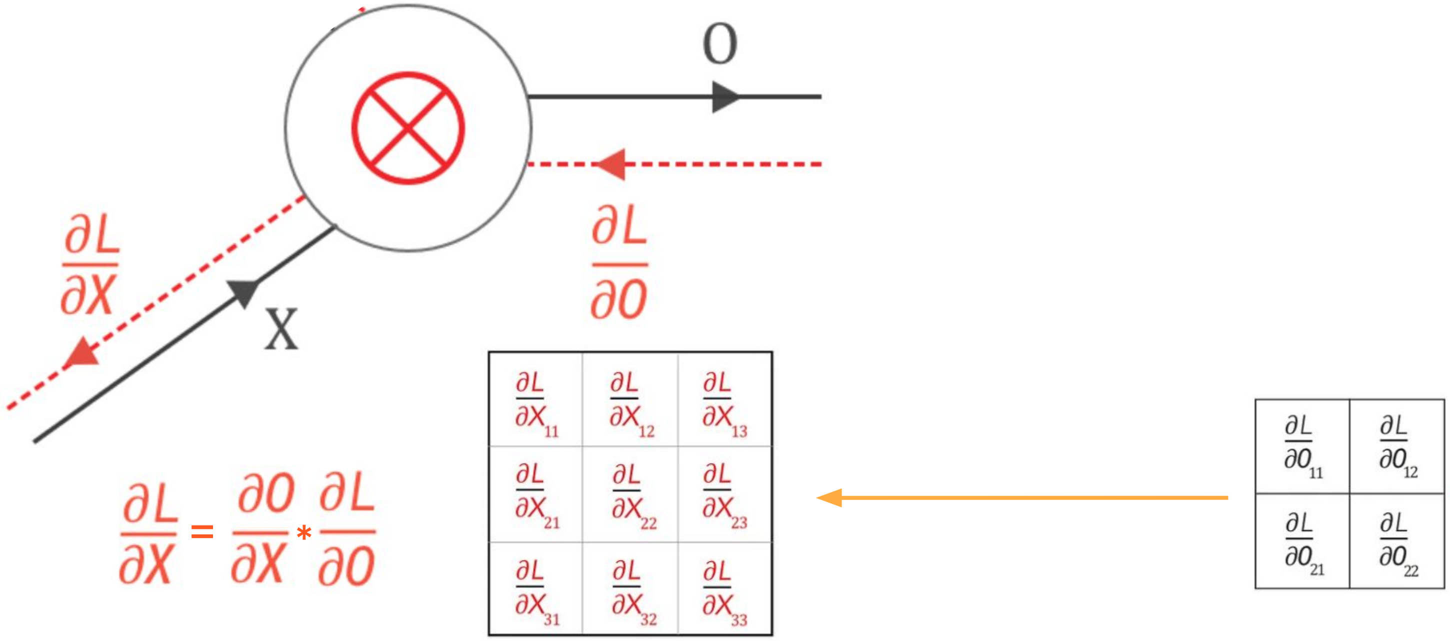


$$\begin{bmatrix} \frac{\partial L}{\partial F_{11}} & \frac{\partial L}{\partial F_{12}} \\ \frac{\partial L}{\partial F_{21}} & \frac{\partial L}{\partial F_{22}} \end{bmatrix} = \text{Convolution}\left(\begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \end{bmatrix}, \begin{bmatrix} \frac{\partial L}{\partial O_{11}} & \frac{\partial L}{\partial O_{12}} \\ \frac{\partial L}{\partial O_{21}} & \frac{\partial L}{\partial O_{22}} \end{bmatrix} \right)$$


F
O
X
$\frac{\partial L}{\partial X}$
$\frac{\partial L}{\partial O}$

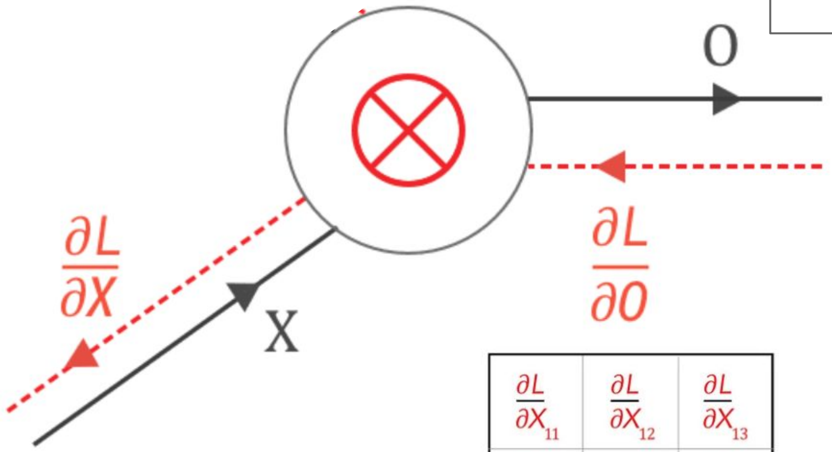


O
X
∂L/∂X
∂L/∂O
∂L/∂X11 ∂L/∂X12 ∂L/∂X13
∂L/∂X21 ∂L/∂X22 ∂L/∂X23
∂L/∂X31 ∂L/∂X32 ∂L/∂X33
∂L/∂O11 ∂L/∂O12
∂L/∂O21 ∂L/∂O22


O
X
$\frac{\partial L}{\partial X}$
$\frac{\partial L}{\partial O}$
$\frac{\partial L}{\partial X} = \frac{\partial O}{\partial X} * \frac{\partial L}{\partial O}$
$\frac{\partial L}{\partial X_{11}}$ $\frac{\partial L}{\partial X_{12}}$ $\frac{\partial L}{\partial X_{13}}$
$\frac{\partial L}{\partial X_{21}}$ $\frac{\partial L}{\partial X_{22}}$ $\frac{\partial L}{\partial X_{23}}$
$\frac{\partial L}{\partial X_{31}}$ $\frac{\partial L}{\partial X_{32}}$ $\frac{\partial L}{\partial X_{33}}$
$\frac{\partial L}{\partial O_{11}}$ $\frac{\partial L}{\partial O_{12}}$
$\frac{\partial L}{\partial O_{21}}$ $\frac{\partial L}{\partial O_{22}}$

For every element of X_i

$$\frac{\partial L}{\partial X_i} = \sum_{k=1}^{M} \frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial X_i}$$

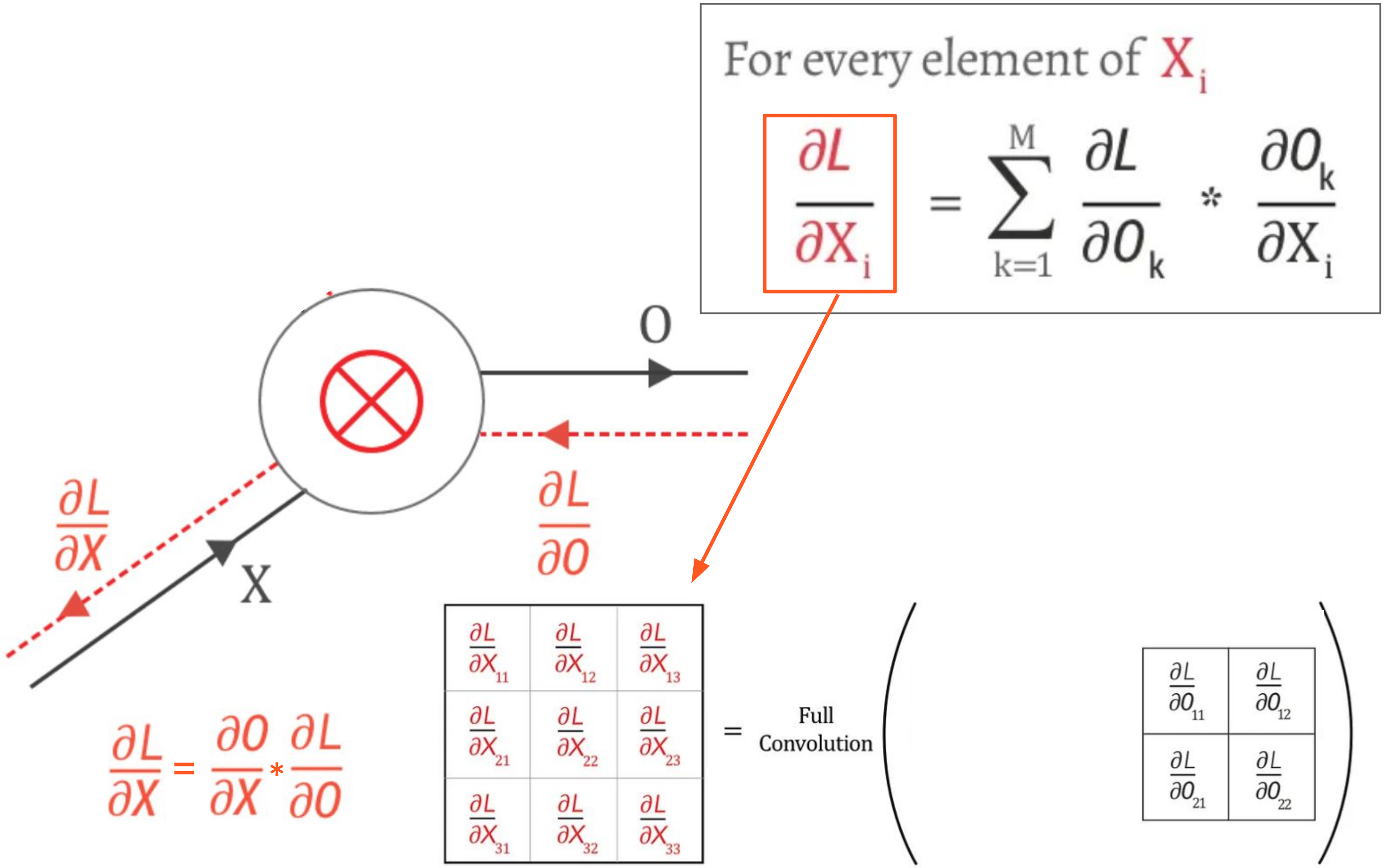


$$\frac{\partial L}{\partial X} = \frac{\partial O}{\partial X} * \frac{\partial L}{\partial O}$$

$\frac{\partial L}{\partial X_{11}}$	$\frac{\partial L}{\partial X_{12}}$	$\frac{\partial L}{\partial X_{13}}$
$\frac{\partial L}{\partial X_{21}}$	$\frac{\partial L}{\partial X_{22}}$	$\frac{\partial L}{\partial X_{23}}$
$\frac{\partial L}{\partial X_{31}}$	$\frac{\partial L}{\partial X_{32}}$	$\frac{\partial L}{\partial X_{33}}$

←

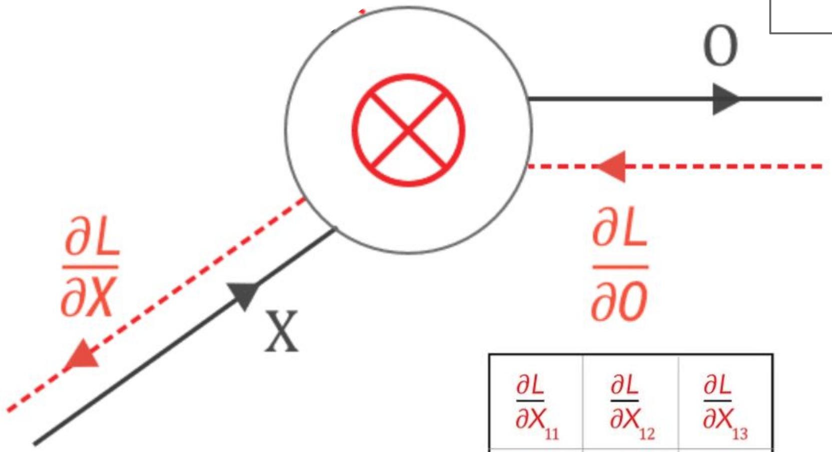
$\frac{\partial L}{\partial O_{11}}$	$\frac{\partial L}{\partial O_{12}}$
$\frac{\partial L}{\partial O_{21}}$	$\frac{\partial L}{\partial O_{22}}$



For every element of X_i
$$\frac{\partial L}{\partial X_i} = \sum_{k=1}^{M} \frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial X_i}$$
O
X
$\frac{\partial L}{\partial O}$
$\frac{\partial L}{\partial X}$
$$\frac{\partial L}{\partial X} = \frac{\partial O}{\partial X} * \frac{\partial L}{\partial O}$$
$\frac{\partial L}{\partial X_{11}}$ $\frac{\partial L}{\partial X_{12}}$ $\frac{\partial L}{\partial X_{13}}$
$\frac{\partial L}{\partial X_{21}}$ $\frac{\partial L}{\partial X_{22}}$ $\frac{\partial L}{\partial X_{23}}$
$\frac{\partial L}{\partial X_{31}}$ $\frac{\partial L}{\partial X_{32}}$ $\frac{\partial L}{\partial X_{33}}$
= Full Convolution
$\frac{\partial L}{\partial O_{11}}$ $\frac{\partial L}{\partial O_{12}}$
$\frac{\partial L}{\partial O_{21}}$ $\frac{\partial L}{\partial O_{22}}$

For every element of X_i

$$\frac{\partial L}{\partial X_i} = \sum_{k=1}^{M} \frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial X_i}$$



$$\frac{\partial L}{\partial X} = \frac{\partial O}{\partial X} * \frac{\partial L}{\partial O}$$

$$\begin{bmatrix} \frac{\partial L}{\partial X_{11}} & \frac{\partial L}{\partial X_{12}} & \frac{\partial L}{\partial X_{13}} \\ \frac{\partial L}{\partial X_{21}} & \frac{\partial L}{\partial X_{22}} & \frac{\partial L}{\partial X_{23}} \\ \frac{\partial L}{\partial X_{31}} & \frac{\partial L}{\partial X_{32}} & \frac{\partial L}{\partial X_{33}} \end{bmatrix} = \text{Full Convolution}\left(? \;,\; \begin{bmatrix} \frac{\partial L}{\partial O_{11}} & \frac{\partial L}{\partial O_{12}} \\ \frac{\partial L}{\partial O_{21}} & \frac{\partial L}{\partial O_{22}} \end{bmatrix} \right)$$

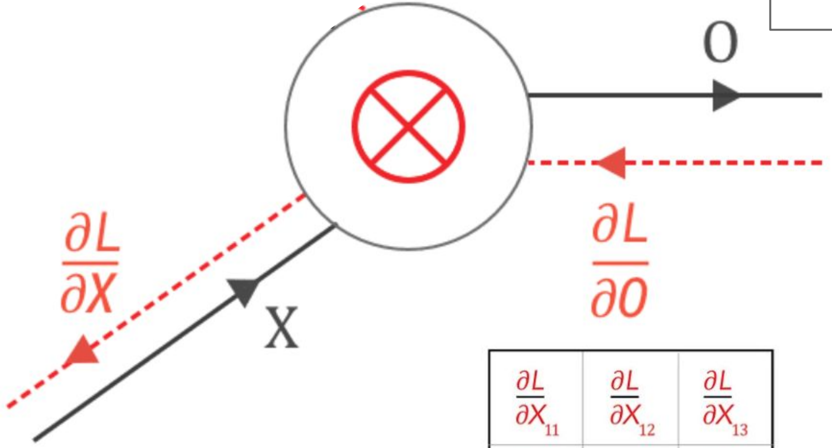
Hint:

$$O_{11} = X_{11}F_{11} + X_{12}F_{12} + X_{21}F_{21} + X_{22}F_{22} \cdots$$

$$\frac{\partial O_{11}}{\partial X_{11}} = F_{11} \quad \frac{\partial O_{11}}{\partial X_{12}} = F_{12} \quad \frac{\partial O_{11}}{\partial X_{21}} = F_{21} \quad \frac{\partial O_{11}}{\partial X_{22}} = F_{22} \quad \cdots$$

For every element of X_i

$$\frac{\partial L}{\partial X_i} = \sum_{k=1}^{M} \frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial X_i}$$



$$\frac{\partial L}{\partial X} = \frac{\partial O}{\partial X} * \frac{\partial L}{\partial O}$$

$$\begin{bmatrix} \frac{\partial L}{\partial X_{11}} & \frac{\partial L}{\partial X_{12}} & \frac{\partial L}{\partial X_{13}} \\ \frac{\partial L}{\partial X_{21}} & \frac{\partial L}{\partial X_{22}} & \frac{\partial L}{\partial X_{23}} \\ \frac{\partial L}{\partial X_{31}} & \frac{\partial L}{\partial X_{32}} & \frac{\partial L}{\partial X_{33}} \end{bmatrix} = \text{Full Convolution}\left(?\ ,\ \begin{bmatrix} \frac{\partial L}{\partial O_{11}} & \frac{\partial L}{\partial O_{12}} \\ \frac{\partial L}{\partial O_{21}} & \frac{\partial L}{\partial O_{22}} \end{bmatrix} \right)$$

Hint:

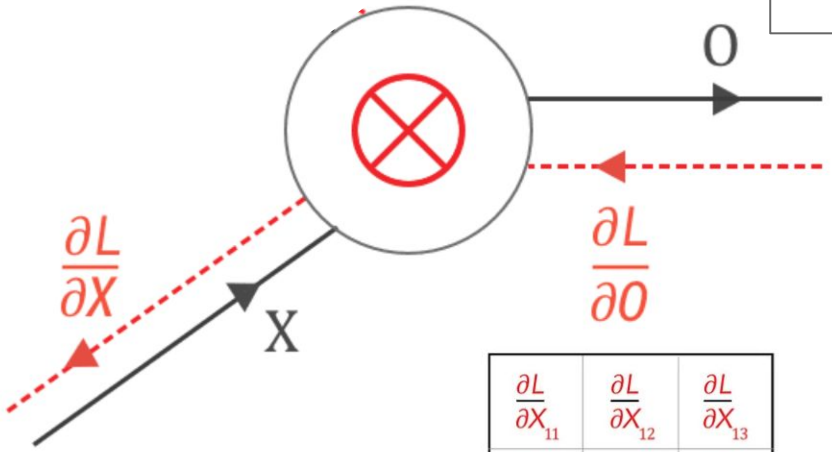
$$O_{11} = X_{11}F_{11} + X_{12}F_{12} + X_{21}F_{21} + X_{22}F_{22} \cdots$$

$$\frac{\partial O_{11}}{\partial X_{11}} = F_{11} \quad \frac{\partial O_{11}}{\partial X_{12}} = F_{12} \quad \frac{\partial O_{11}}{\partial X_{21}} = F_{21} \quad \frac{\partial O_{11}}{\partial X_{22}} = F_{22} \cdots$$

BUT two important things to consider!

For every element of X_i

$$\frac{\partial L}{\partial X_i} = \sum_{k=1}^{M} \frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial X_i}$$



$$\frac{\partial L}{\partial O}$$

$$\frac{\partial L}{\partial X}$$

$$\frac{\partial L}{\partial X} = \frac{\partial O}{\partial X} * \frac{\partial L}{\partial O}$$

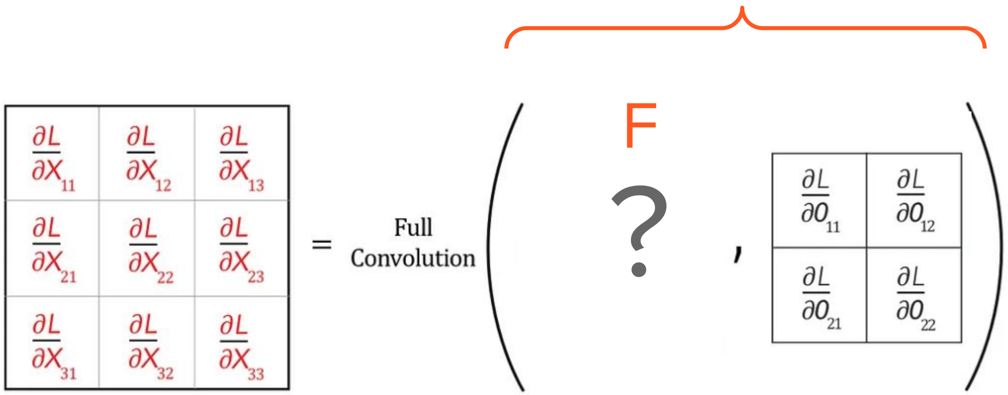
$$\begin{bmatrix} \frac{\partial L}{\partial X_{11}} & \frac{\partial L}{\partial X_{12}} & \frac{\partial L}{\partial X_{13}} \\ \frac{\partial L}{\partial X_{21}} & \frac{\partial L}{\partial X_{22}} & \frac{\partial L}{\partial X_{23}} \\ \frac{\partial L}{\partial X_{31}} & \frac{\partial L}{\partial X_{32}} & \frac{\partial L}{\partial X_{33}} \end{bmatrix} = \text{Full Convolution}\left(\overset{F}{?} \; , \begin{bmatrix} \frac{\partial L}{\partial O_{11}} & \frac{\partial L}{\partial O_{12}} \\ \frac{\partial L}{\partial O_{21}} & \frac{\partial L}{\partial O_{22}} \end{bmatrix} \right)$$

BUT two important things to consider!

- eyeglasses analogy
- spatial dimension

For every element of X_i

$$\frac{\partial L}{\partial X_i} = \sum_{k=1}^{M} \frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial X_i}$$



BUT two important things to consider!

- eyeglasses analogy
- spatial dimension

For every element of X_i

$$\frac{\partial L}{\partial X_i} = \sum_{k=1}^{M} \frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial X_i}$$

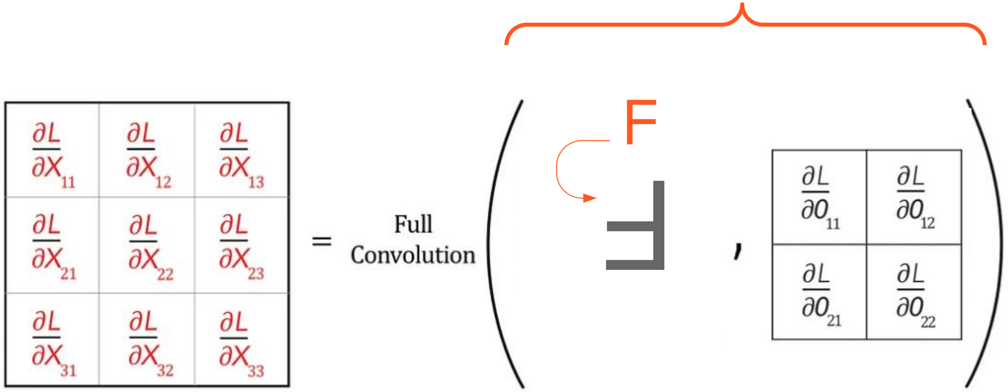
$$\begin{bmatrix} \frac{\partial L}{\partial X_{11}} & \frac{\partial L}{\partial X_{12}} & \frac{\partial L}{\partial X_{13}} \\ \frac{\partial L}{\partial X_{21}} & \frac{\partial L}{\partial X_{22}} & \frac{\partial L}{\partial X_{23}} \\ \frac{\partial L}{\partial X_{31}} & \frac{\partial L}{\partial X_{32}} & \frac{\partial L}{\partial X_{33}} \end{bmatrix} = \text{Full Convolution}\left(\overset{F}{?} \; , \begin{bmatrix} \frac{\partial L}{\partial O_{11}} & \frac{\partial L}{\partial O_{12}} \\ \frac{\partial L}{\partial O_{21}} & \frac{\partial L}{\partial O_{22}} \end{bmatrix} \right)$$

BUT two important things to consider!

- eyeglasses analogy
- spatial dimension

For every element of X_i

$$\frac{\partial L}{\partial X_i} = \sum_{k=1}^{M} \frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial X_i}$$



BUT two important things to consider!

- eyeglasses analogy
- spatial dimension

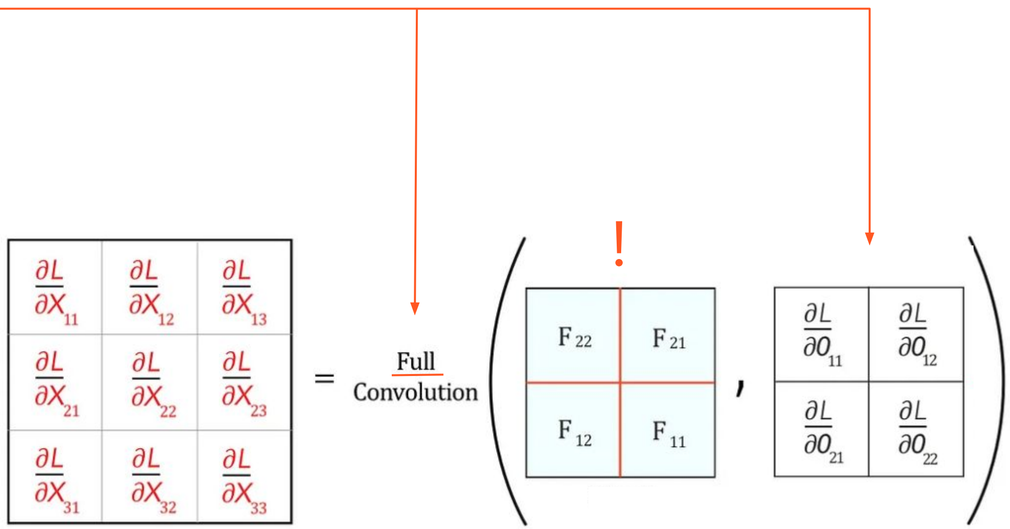
For every element of X_i

$$\frac{\partial L}{\partial X_i} = \sum_{k=1}^{M} \frac{\partial L}{\partial O_k} * \frac{\partial O_k}{\partial X_i}$$

$$\begin{bmatrix} \frac{\partial L}{\partial X_{11}} & \frac{\partial L}{\partial X_{12}} & \frac{\partial L}{\partial X_{13}} \\ \frac{\partial L}{\partial X_{21}} & \frac{\partial L}{\partial X_{22}} & \frac{\partial L}{\partial X_{23}} \\ \frac{\partial L}{\partial X_{31}} & \frac{\partial L}{\partial X_{32}} & \frac{\partial L}{\partial X_{33}} \end{bmatrix} = \text{Full Convolution} \left(\overset{!}{\begin{bmatrix} F_{22} & F_{21} \\ F_{12} & F_{11} \end{bmatrix}}, \begin{bmatrix} \frac{\partial L}{\partial O_{11}} & \frac{\partial L}{\partial O_{12}} \\ \frac{\partial L}{\partial O_{21}} & \frac{\partial L}{\partial O_{22}} \end{bmatrix} \right)$$

BUT two important things to consider!

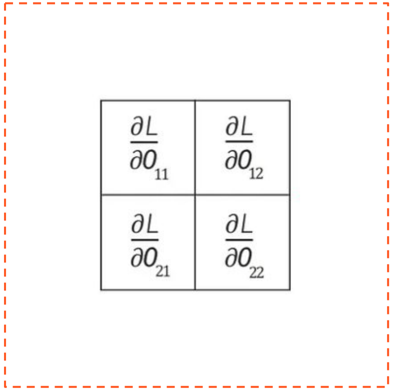
- eyeglasses analogy
- spatial dimension

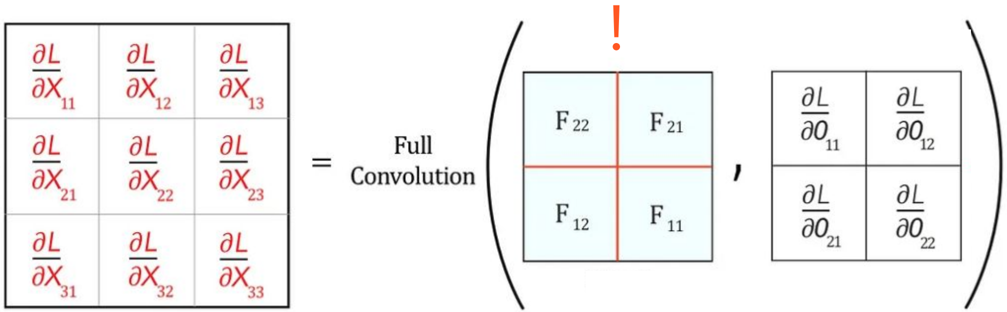


BUT two important things to consider!

- eyeglasses analogy
- spatial dimension

Zero padding with (row length -1) and (column length -1)

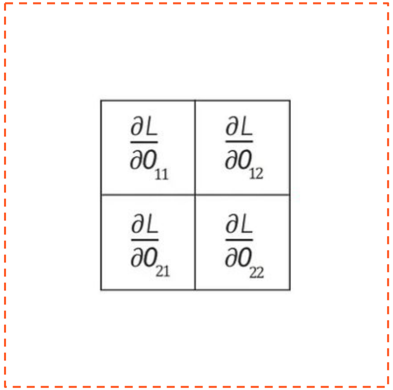


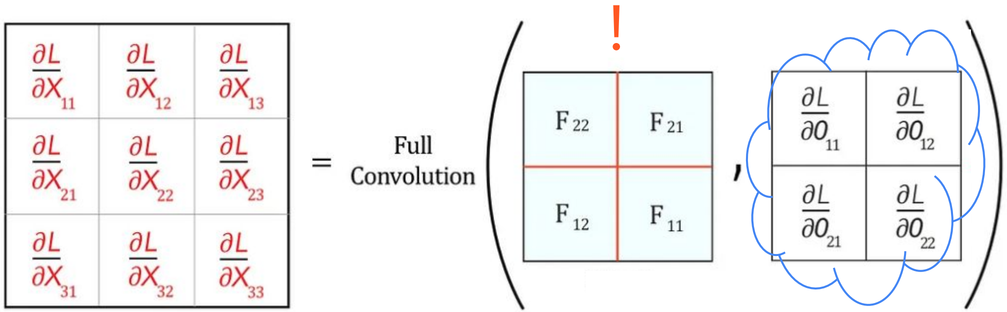


BUT two important things to consider!

- eyeglasses analogy
- spatial dimension

Zero padding with (row length -1) and (column length -1)

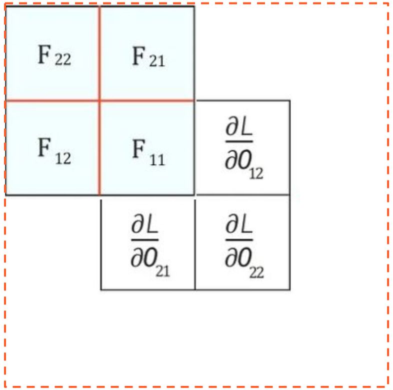


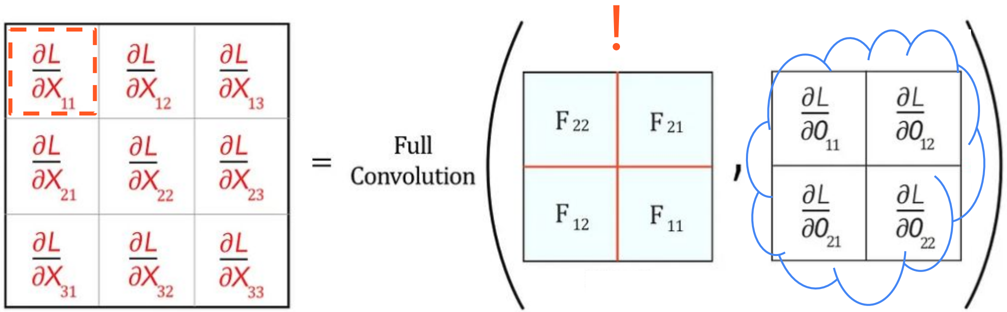


BUT two important things to consider!

- eyeglasses analogy
- spatial dimension

Zero padding with (row length -1) and (column length -1)

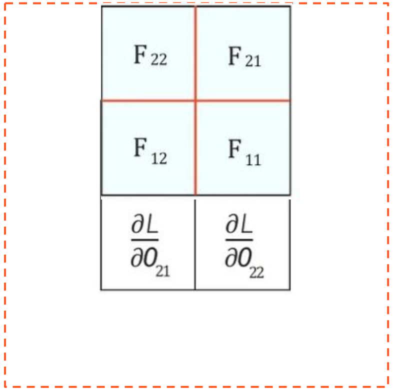


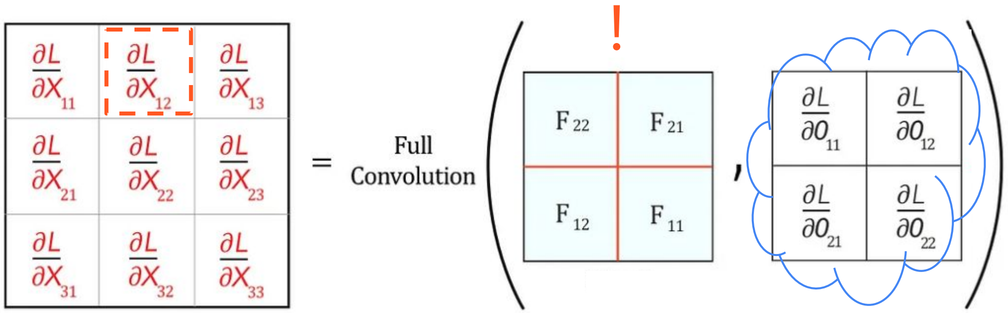


BUT two important things to consider!

- eyeglasses analogy
- spatial dimension

Zero padding with (row length -1) and (column length -1)

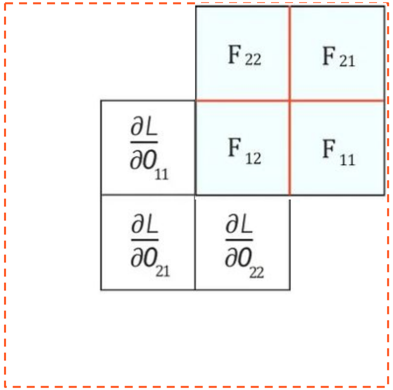


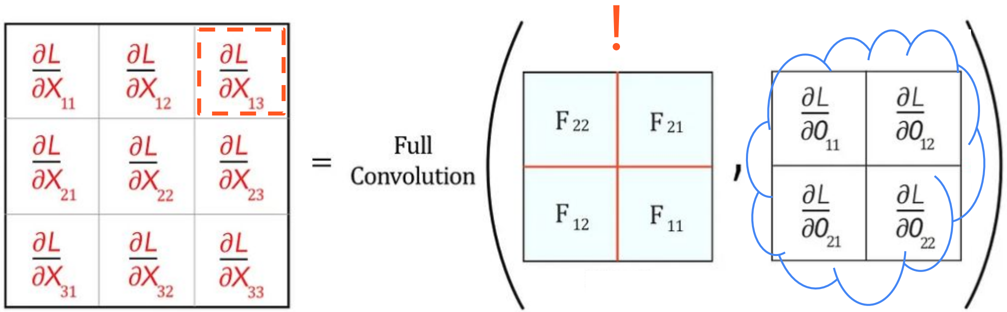


BUT two important things to consider!

- eyeglasses analogy
- spatial dimension

Zero padding with (row length -1) and (column length -1)

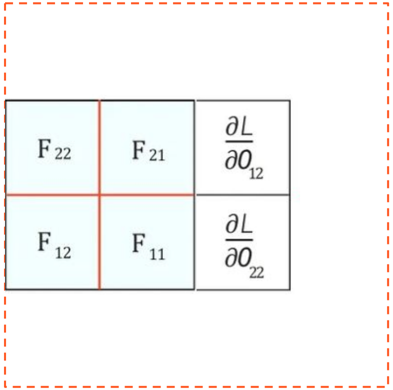


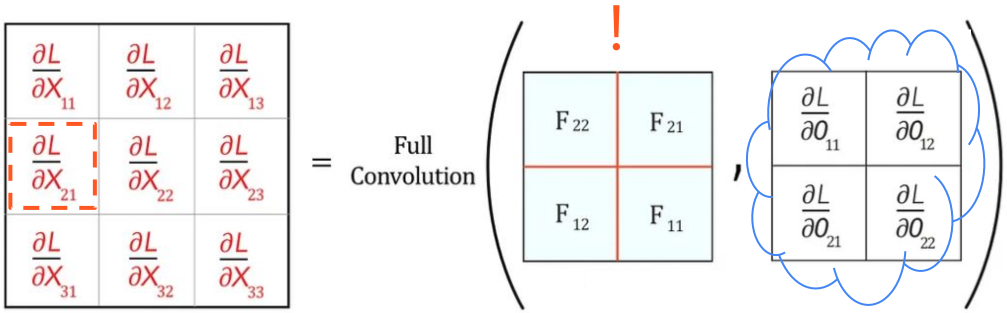


BUT two important things to consider!

- eyeglasses analogy
- spatial dimension

Zero padding with (row length -1) and (column length -1)

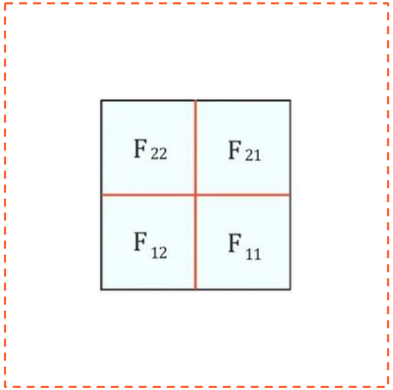


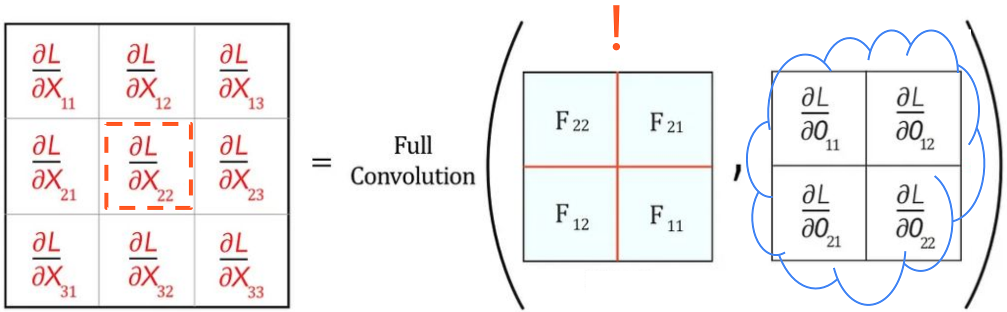


BUT two important things to consider!

- eyeglasses analogy
- spatial dimension

Zero padding with (row length -1) and (column length -1)

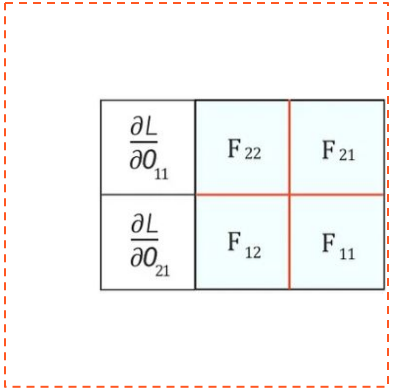


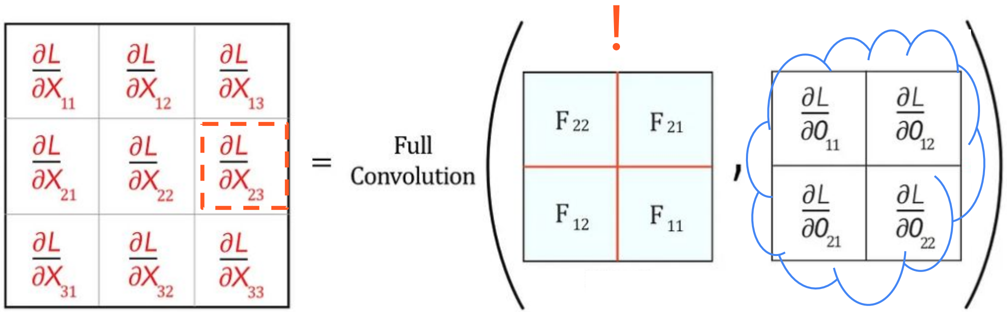


BUT two important things to consider!

- eyeglasses analogy
- spatial dimension

Zero padding with (row length -1) and (column length -1)

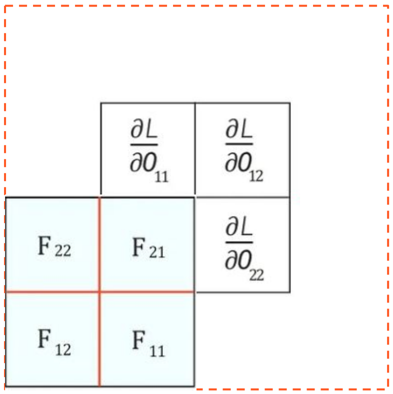


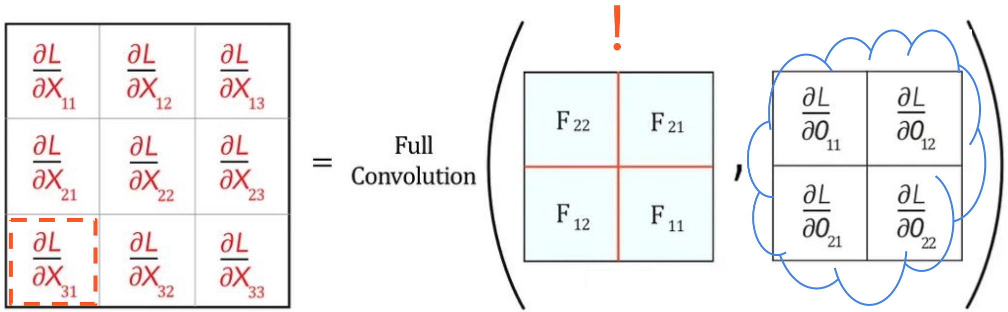


BUT two important things to consider!

- eyeglasses analogy
- spatial dimension

Zero padding with (row length -1) and (column length -1)

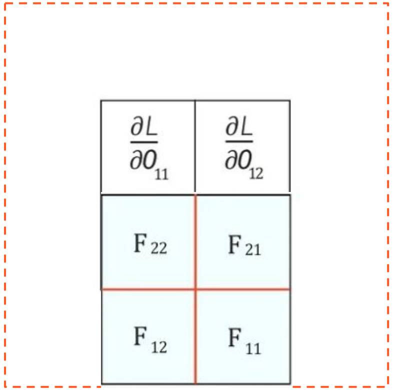


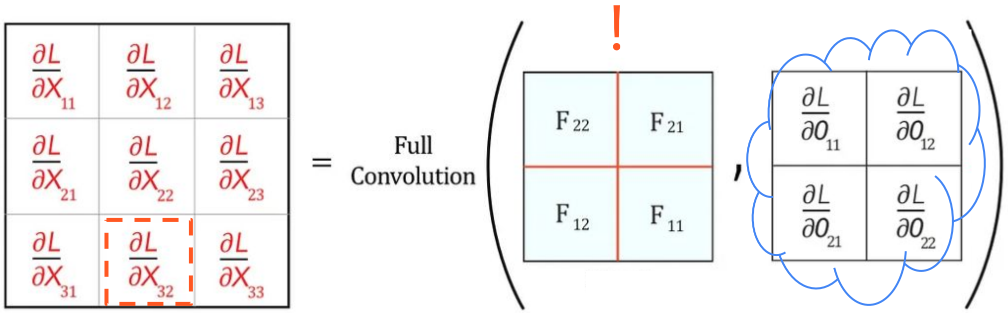


BUT two important things to consider!

- eyeglasses analogy
- spatial dimension

Zero padding with (row length -1) and (column length -1)

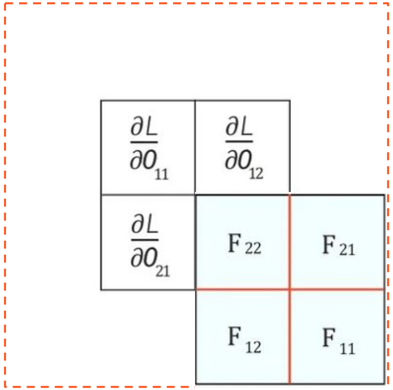


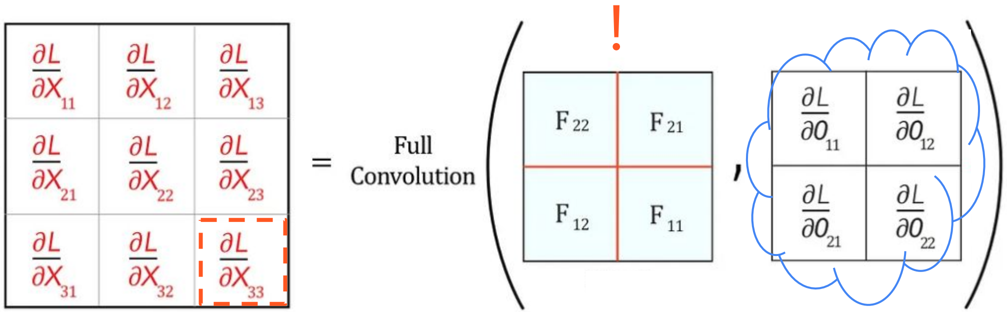


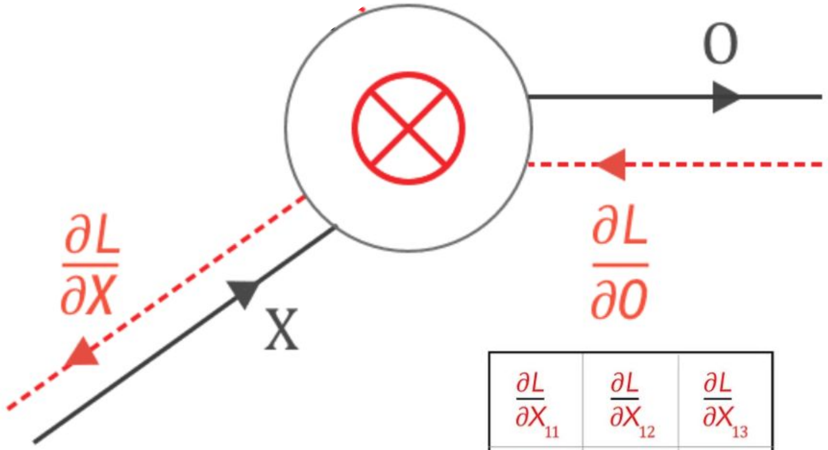
BUT two important things to consider!

- eyeglasses analogy
- spatial dimension

Zero padding with (row length -1) and (column length -1)

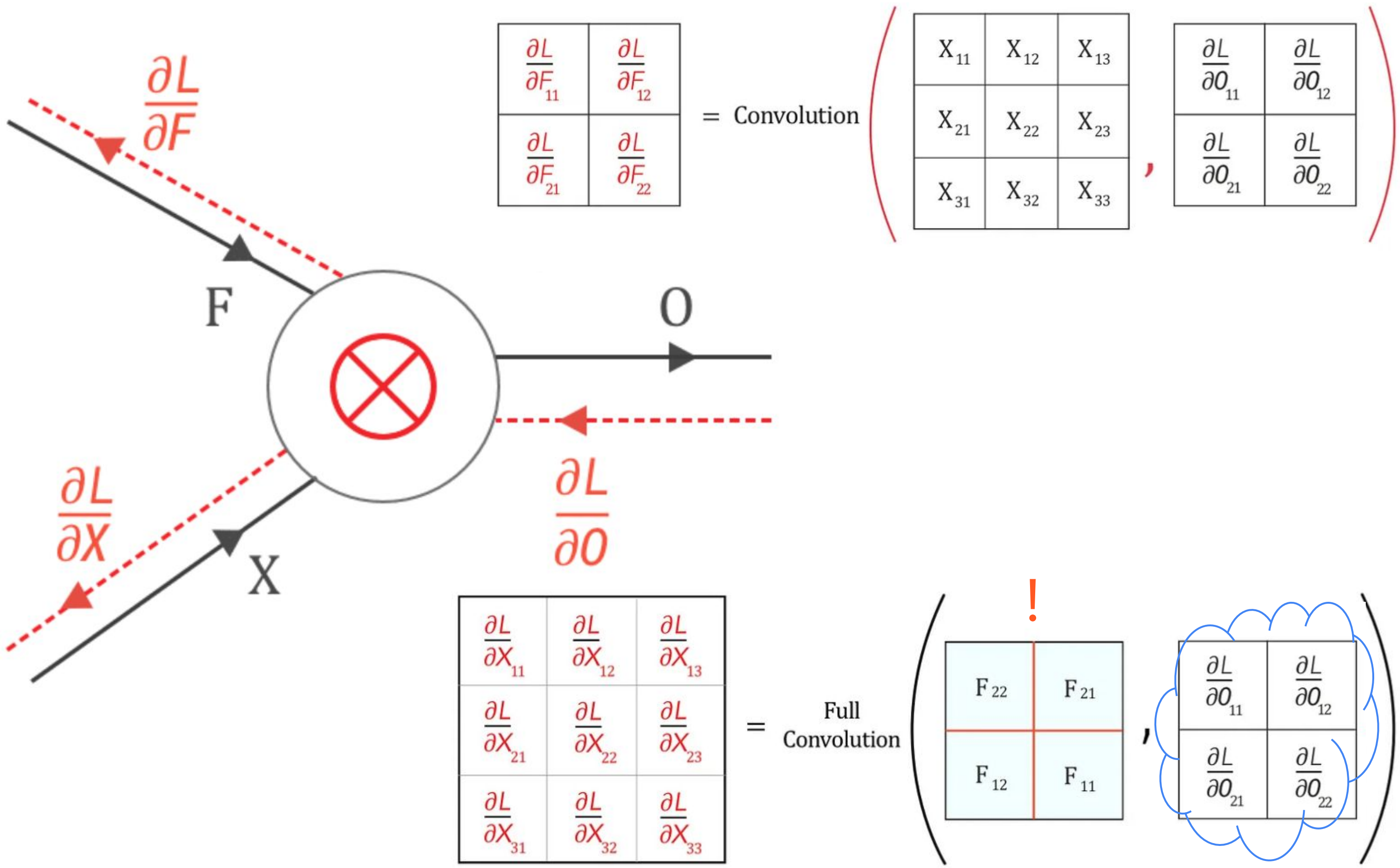




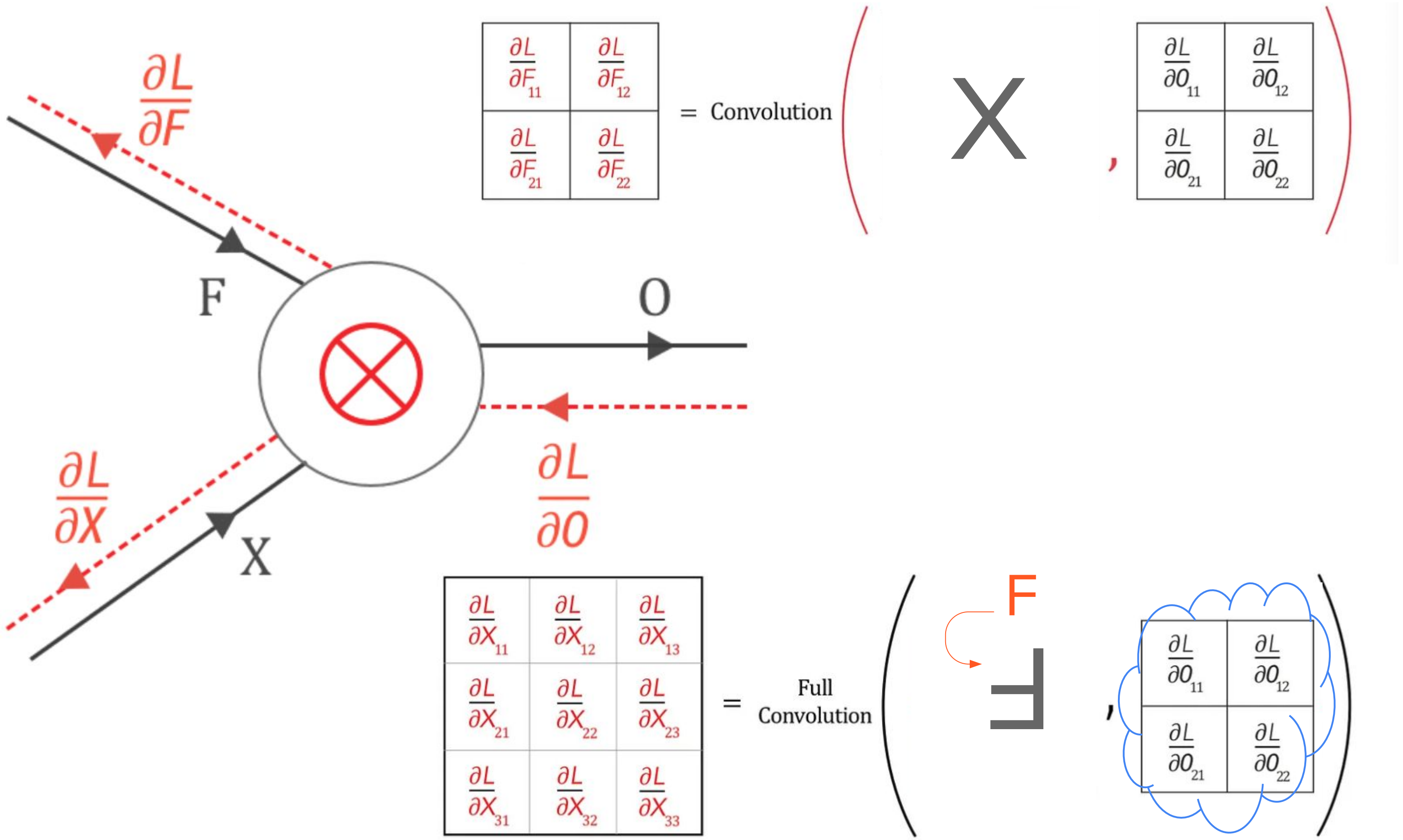


$$\frac{\partial L}{\partial X} = \frac{\partial O}{\partial X} * \frac{\partial L}{\partial O}$$

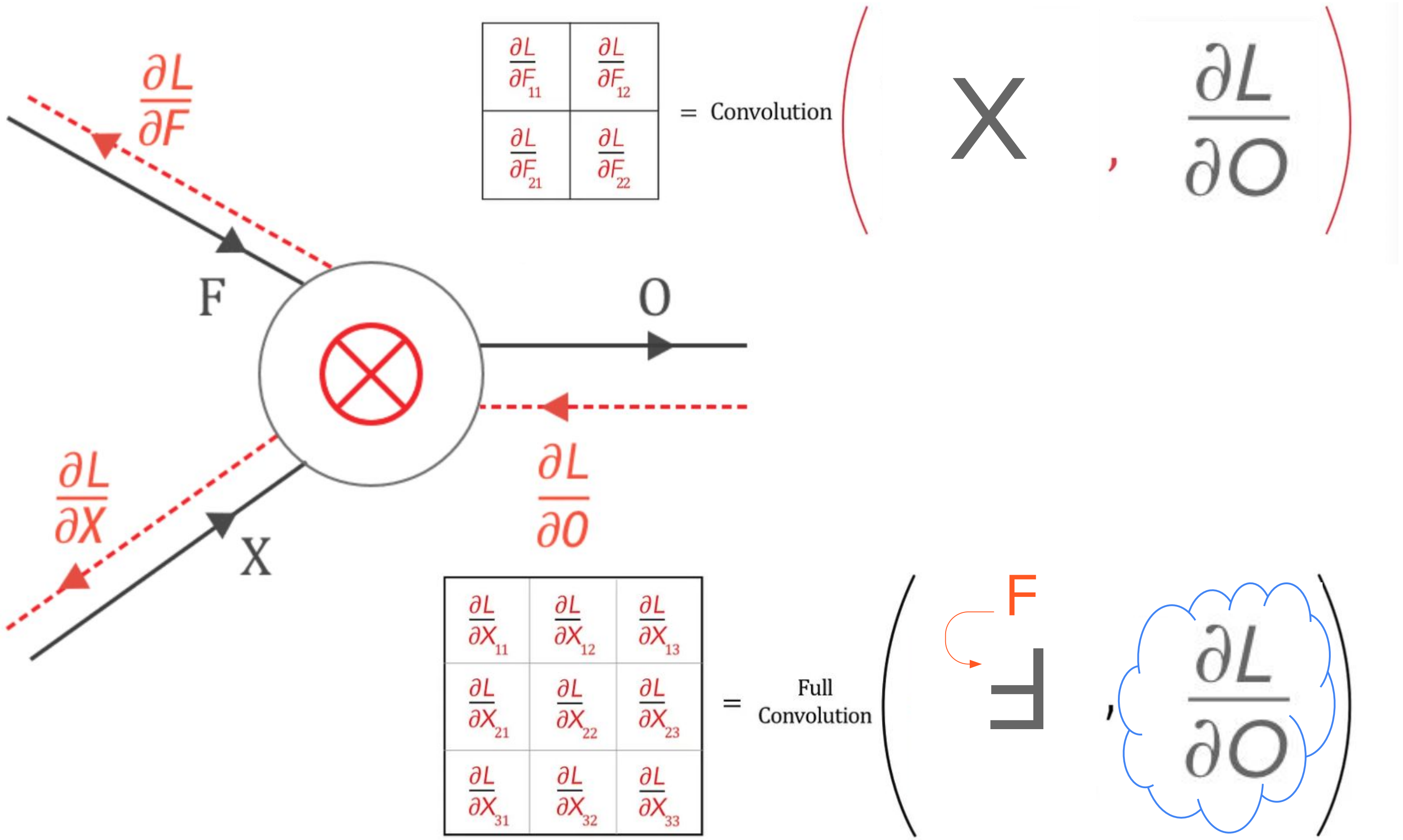
$$\begin{bmatrix} \frac{\partial L}{\partial X_{11}} & \frac{\partial L}{\partial X_{12}} & \frac{\partial L}{\partial X_{13}} \\ \frac{\partial L}{\partial X_{21}} & \frac{\partial L}{\partial X_{22}} & \frac{\partial L}{\partial X_{23}} \\ \frac{\partial L}{\partial X_{31}} & \frac{\partial L}{\partial X_{32}} & \frac{\partial L}{\partial X_{33}} \end{bmatrix} = \text{Full Convolution}\left(\overset{!}{\begin{bmatrix} F_{22} & F_{21} \\ F_{12} & F_{11} \end{bmatrix}}, \begin{bmatrix} \frac{\partial L}{\partial O_{11}} & \frac{\partial L}{\partial O_{12}} \\ \frac{\partial L}{\partial O_{21}} & \frac{\partial L}{\partial O_{22}} \end{bmatrix} \right)$$



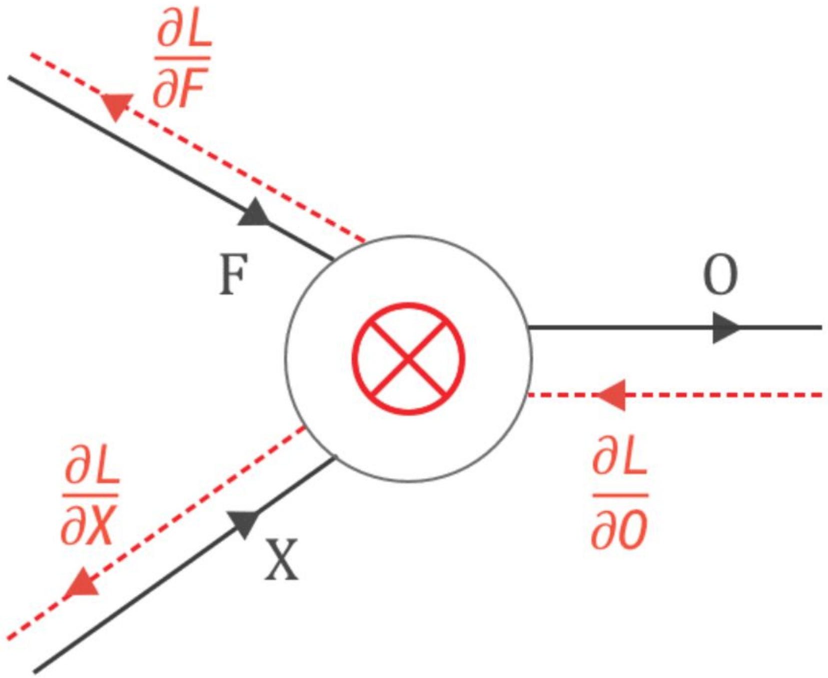
∂L/∂F
F
O
∂L/∂X
X
∂L/∂O
= Convolution
= Full Convolution
!



∂L/∂F
F
O
∂L/∂X
X
∂L/∂O
∂L/∂F11 ∂L/∂F12
∂L/∂F21 ∂L/∂F22
= Convolution (X , ∂L/∂O11 ∂L/∂O12 ∂L/∂O21 ∂L/∂O22)
∂L/∂X11 ∂L/∂X12 ∂L/∂X13
∂L/∂X21 ∂L/∂X22 ∂L/∂X23
∂L/∂X31 ∂L/∂X32 ∂L/∂X33
= Full Convolution (F , ∂L/∂O11 ∂L/∂O12 ∂L/∂O21 ∂L/∂O22)

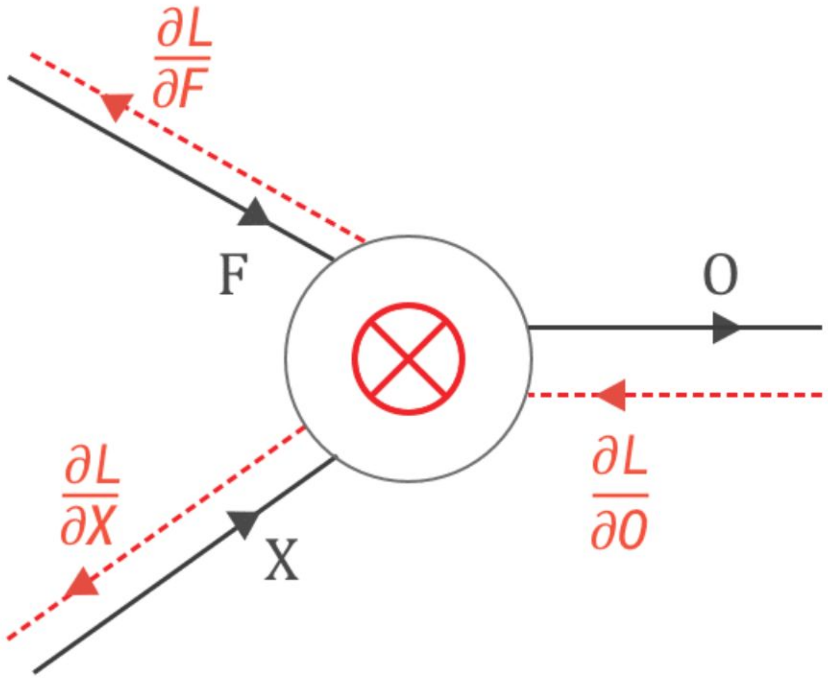


∂L/∂F
F
O
∂L/∂X
X
∂L/∂O
∂L/∂F11 ∂L/∂F12
∂L/∂F21 ∂L/∂F22
= Convolution (X , ∂L/∂O)
∂L/∂X11 ∂L/∂X12 ∂L/∂X13
∂L/∂X21 ∂L/∂X22 ∂L/∂X23
∂L/∂X31 ∂L/∂X32 ∂L/∂X33
= Full Convolution (F , ∂L/∂O)



What about strides > 1 ?

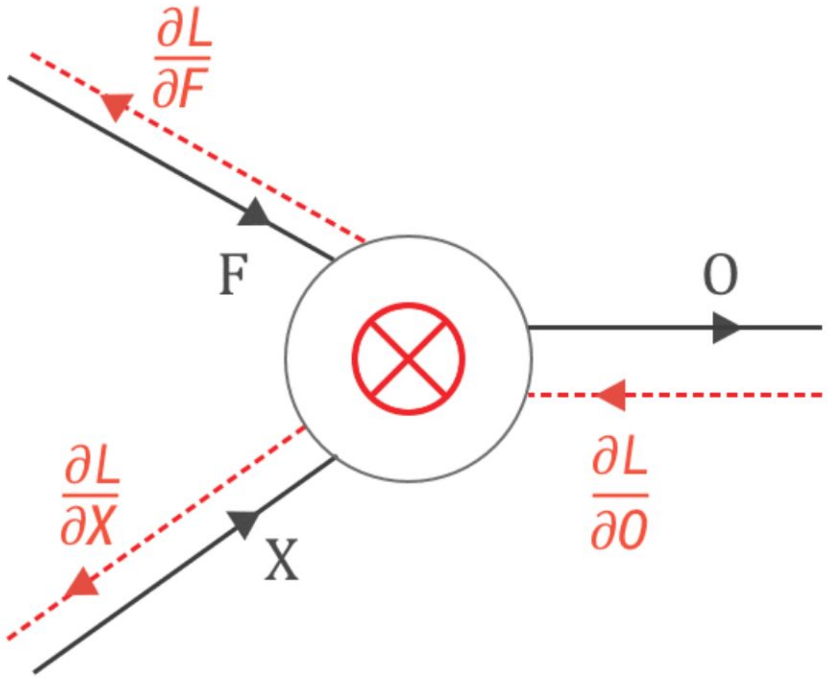
$\frac{\partial L}{\partial O_{11}}$	$\frac{\partial L}{\partial O_{12}}$
$\frac{\partial L}{\partial O_{21}}$	$\frac{\partial L}{\partial O_{22}}$



What about strides > 1 ?

Dilate zeros with (stride_row -1) and (stride_col -1)

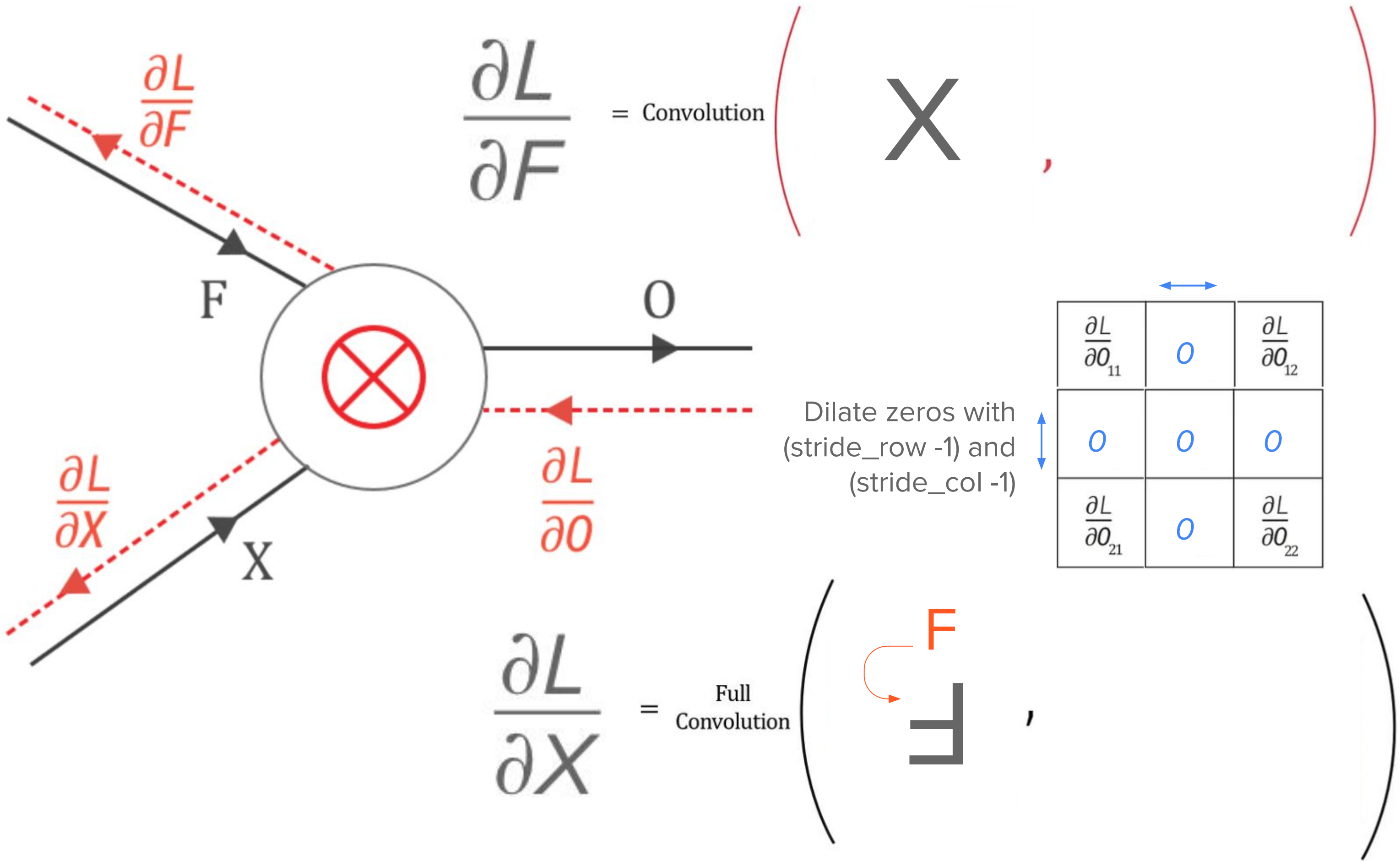
$\frac{\partial L}{\partial O_{11}}$	$\frac{\partial L}{\partial O_{12}}$
$\frac{\partial L}{\partial O_{21}}$	$\frac{\partial L}{\partial O_{22}}$

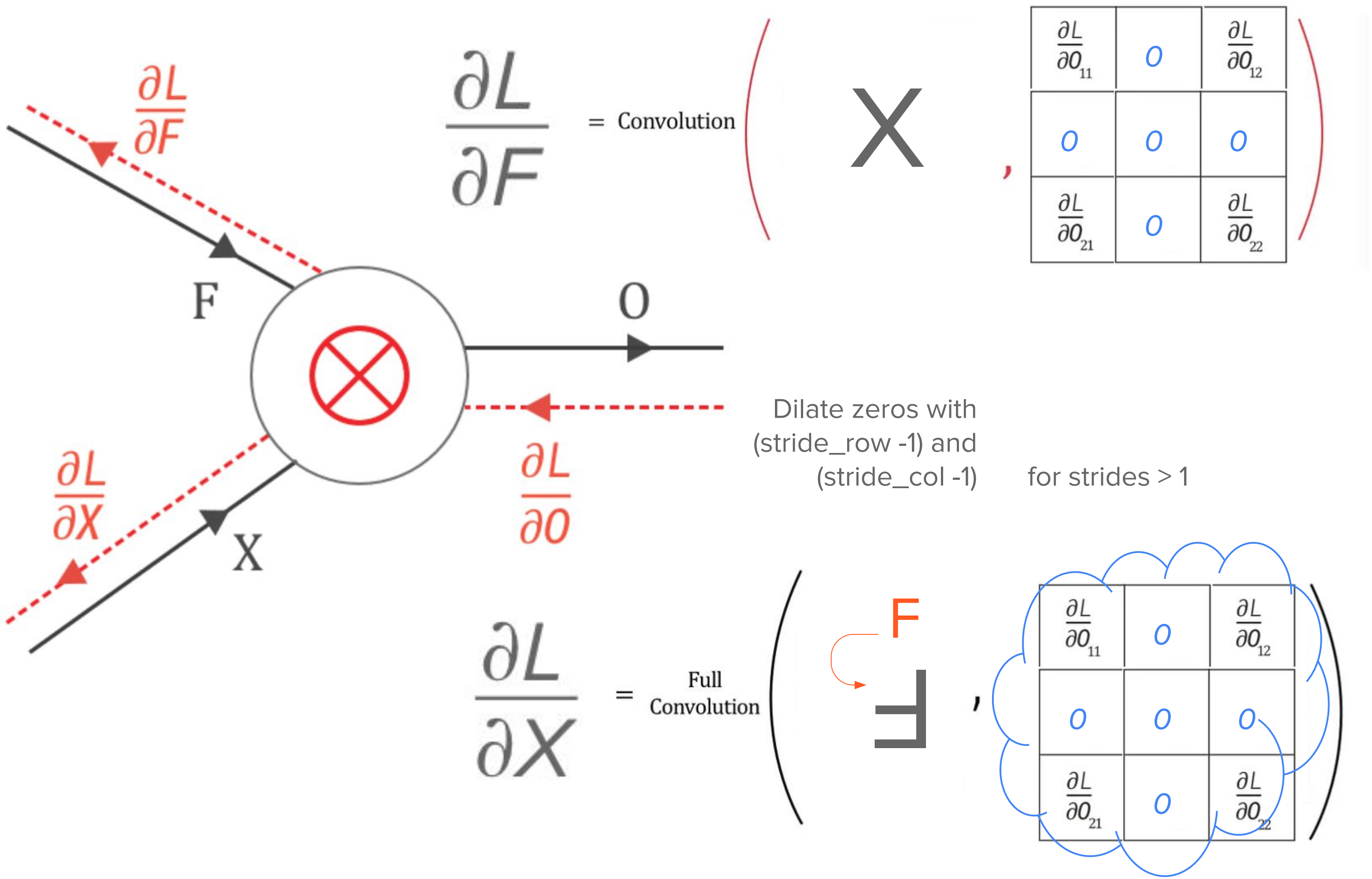


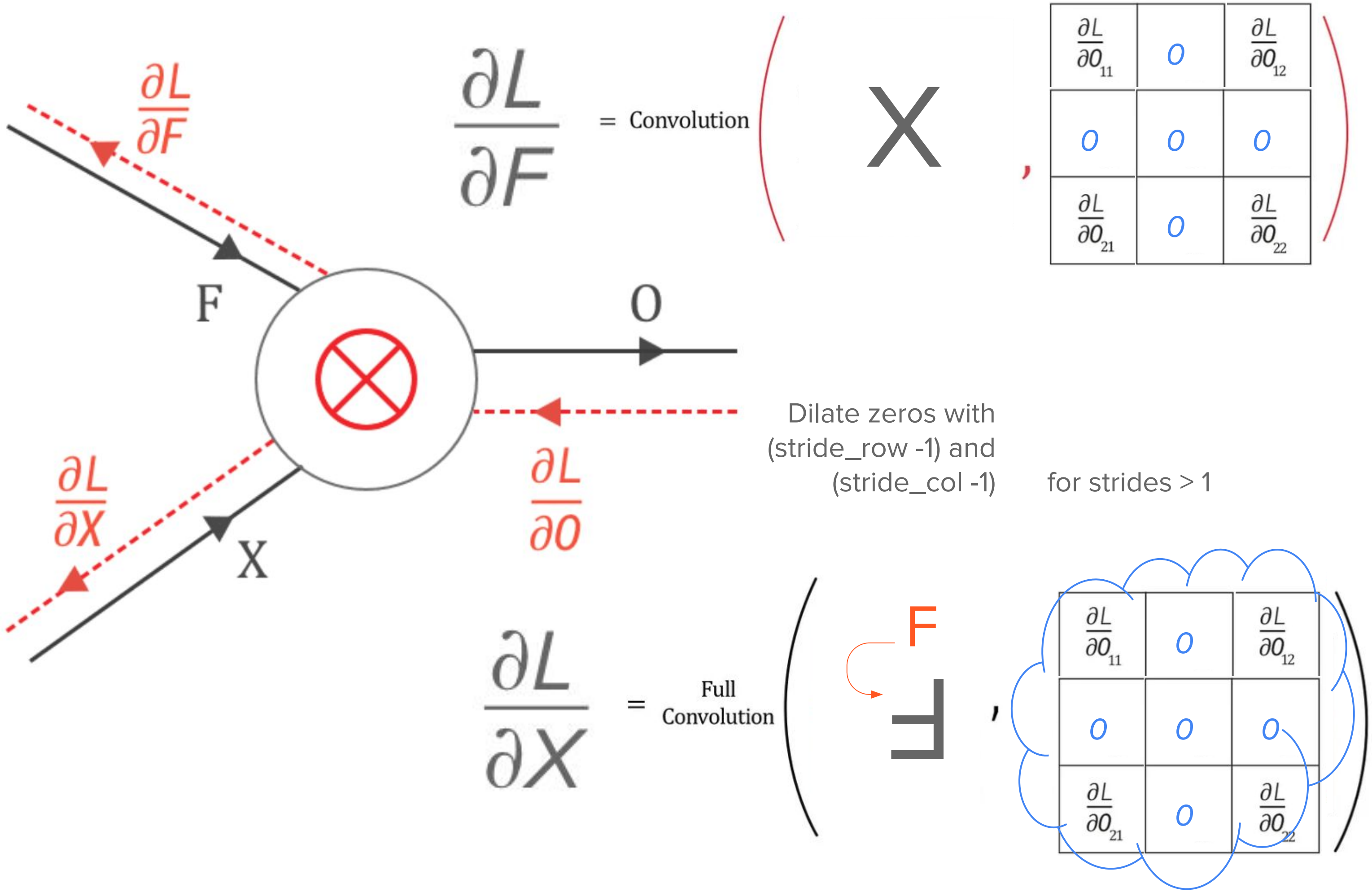
What about strides > 1 ?

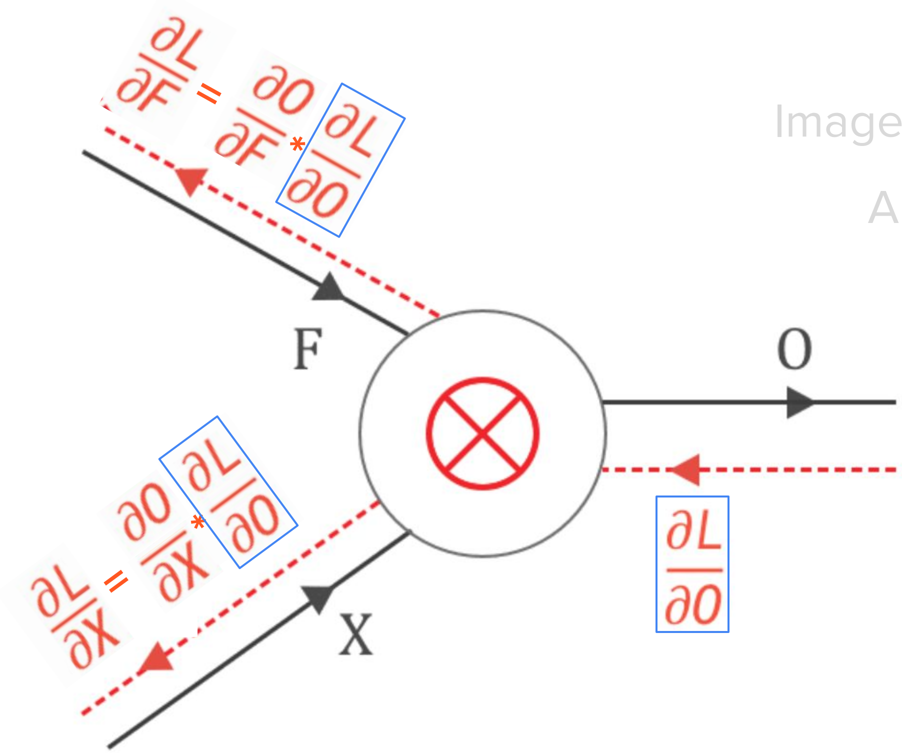
Dilate zeros with (stride_row -1) and (stride_col -1)

$\frac{\partial L}{\partial O_{11}}$	0	$\frac{\partial L}{\partial O_{12}}$
0	0	0
$\frac{\partial L}{\partial O_{21}}$	0	$\frac{\partial L}{\partial O_{22}}$


∂L/∂F
F
O
X
∂L/∂X
∂L/∂O
∂L/∂F = Convolution (X ,)
Dilate zeros with (stride_row -1) and (stride_col -1)
∂L/∂O11 0 ∂L/∂O12
0 0 0
∂L/∂O21 0 ∂L/∂O22
∂L/∂X = Full Convolution (F ,)


∂L/∂F
F
O
∂L/∂X
X
∂L/∂O
∂L/∂F = Convolution (X , [∂L/∂O11, 0, ∂L/∂O12; 0, 0, 0; ∂L/∂O21, 0, ∂L/∂O22])
Dilate zeros with (stride_row -1) and (stride_col -1)
for strides > 1
∂L/∂X = Full Convolution (F , [∂L/∂O11, 0, ∂L/∂O12; 0, 0, 0; ∂L/∂O21, 0, ∂L/∂O22])


∂L/∂F
F
O
∂L/∂X
X
∂L/∂O
∂L/∂F = Convolution (X , [∂L/∂O11, 0, ∂L/∂O12; 0, 0, 0; ∂L/∂O21, 0, ∂L/∂O22])
Dilate zeros with
(stride_row -1) and
(stride_col -1)
for strides > 1
∂L/∂X = Full Convolution (F , [∂L/∂O11, 0, ∂L/∂O12; 0, 0, 0; ∂L/∂O21, 0, ∂L/∂O22])



Images credit: pavisj.medium.com

And now, onwards to Kahoot!