Learning for vision III
Convolutional networks

Karel Zimmermann

http://cmp.felk.cvut.cz/~zimmerk/

Vision for Robotics and Autonomous Systems
https://cyber.felk.cvut.cz/vras/

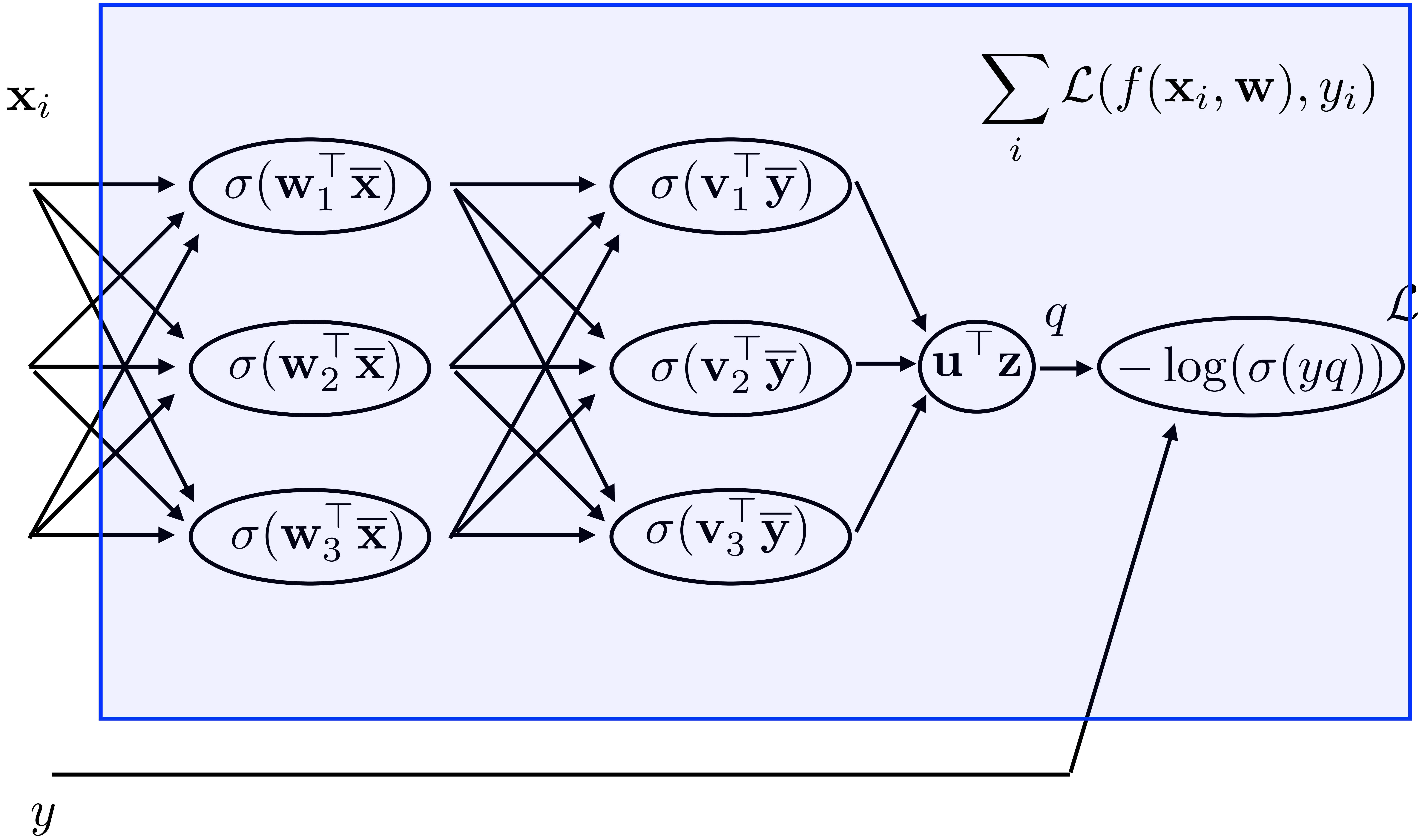
Center for Machine Perception
https://cmp.felk.cvut.cz

Department for Cybernetics
Faculty of Electrical Engineering
Czech Technical University in Prague

Outline

- Fully connected neural network
- Avoid overfitting by search for the NN model suitable for image processing [Hubel and Wiesel 1960].
- Feedforward and Backprop in ConvNets.

Fully connected neural network

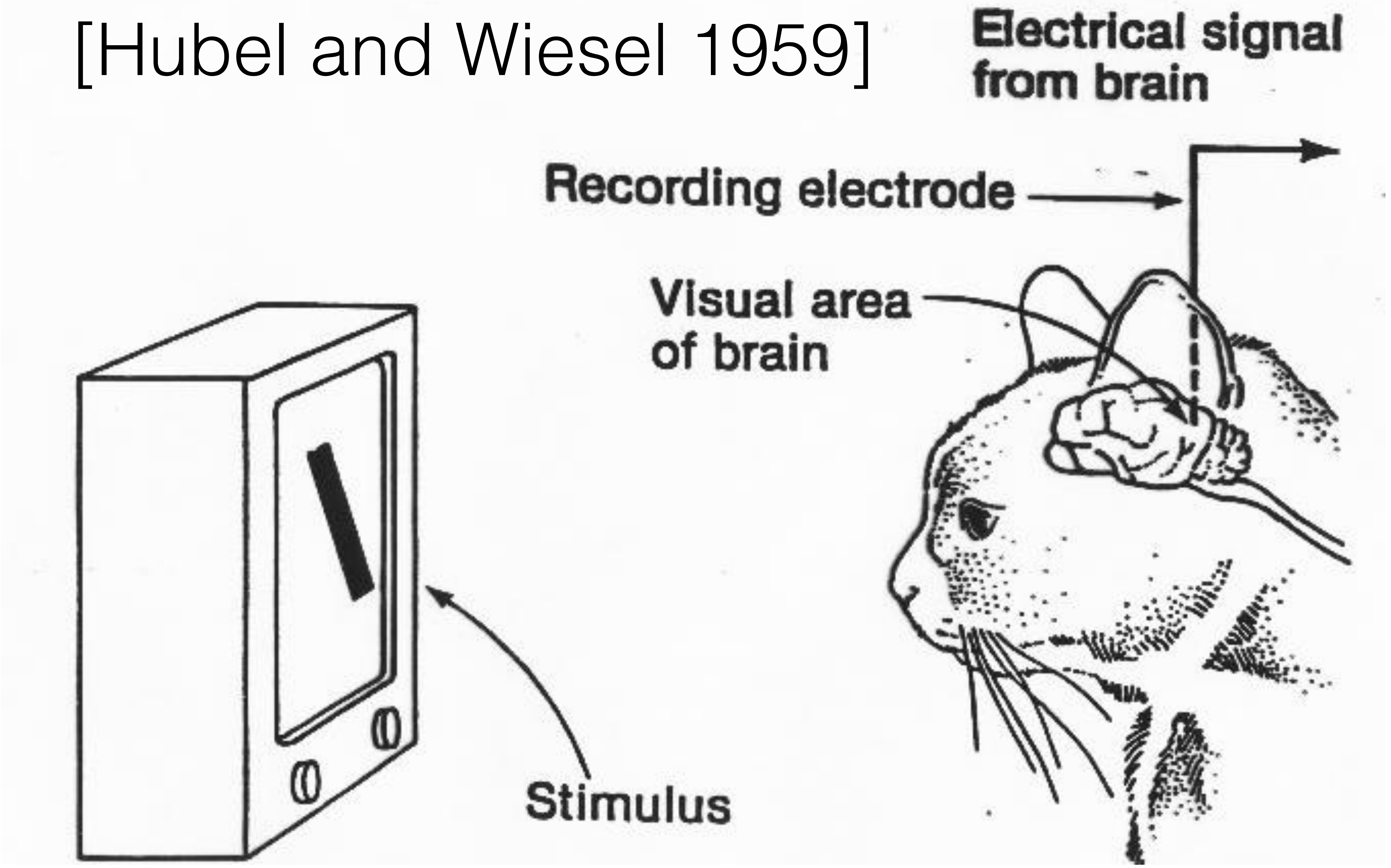


The Tungsten Electrode [Hubel-Science-1957]

http://braintour.harvard.edu/archives/portfolio-items/hubel-and-wiesel

- Device capable to record signal from a single neuron

[Hubel and Wiesel 1959]



- Experiment with anaesthetised paralysed cat

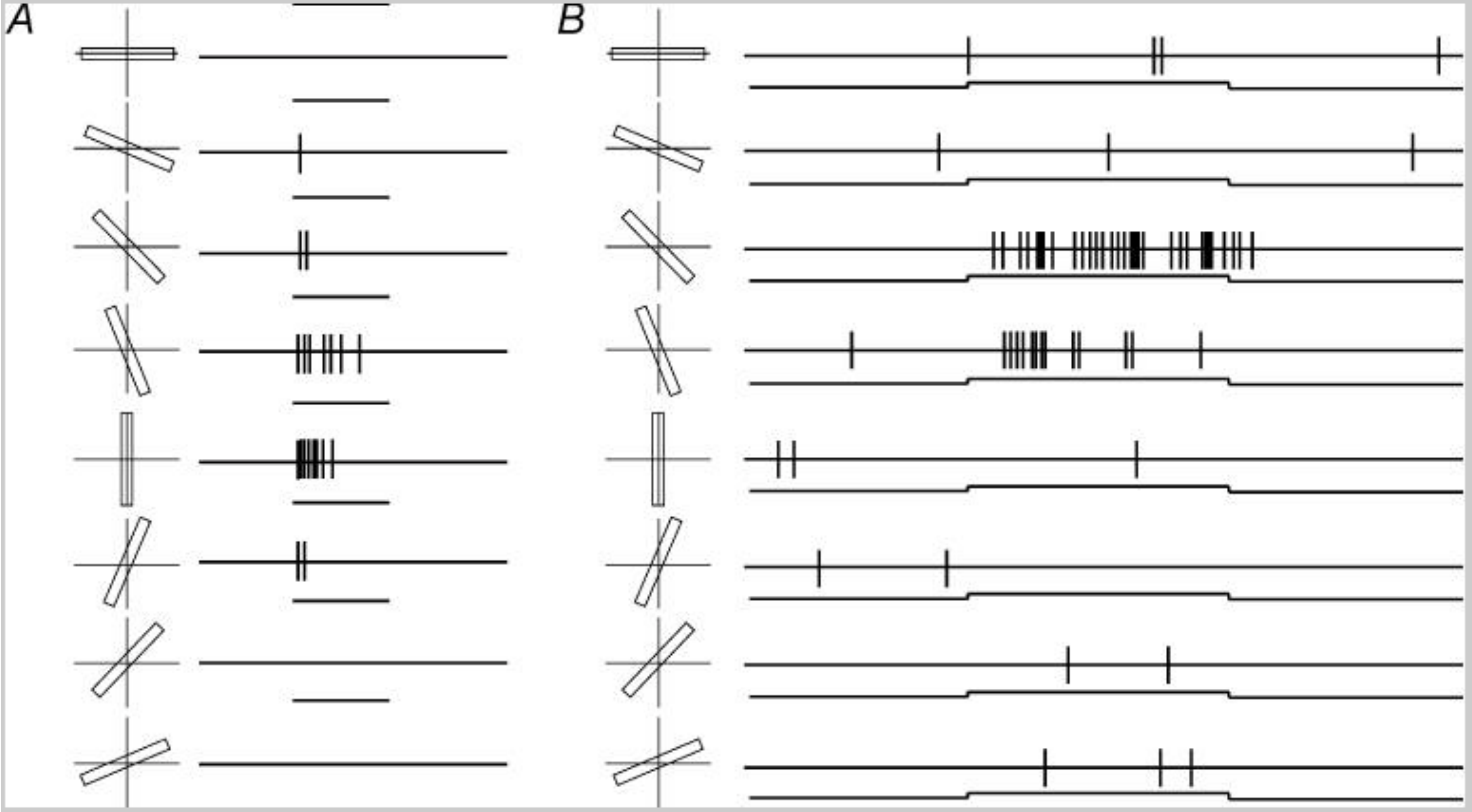
[Hubel and Wiesel 1960]

https://knowingneurons.com/2014/10/29/hubel-and-wiesel-the-neural-basis-of-visual-perception/

[Hubel and Wiesel 1960]

paralysed cat

awake monkey



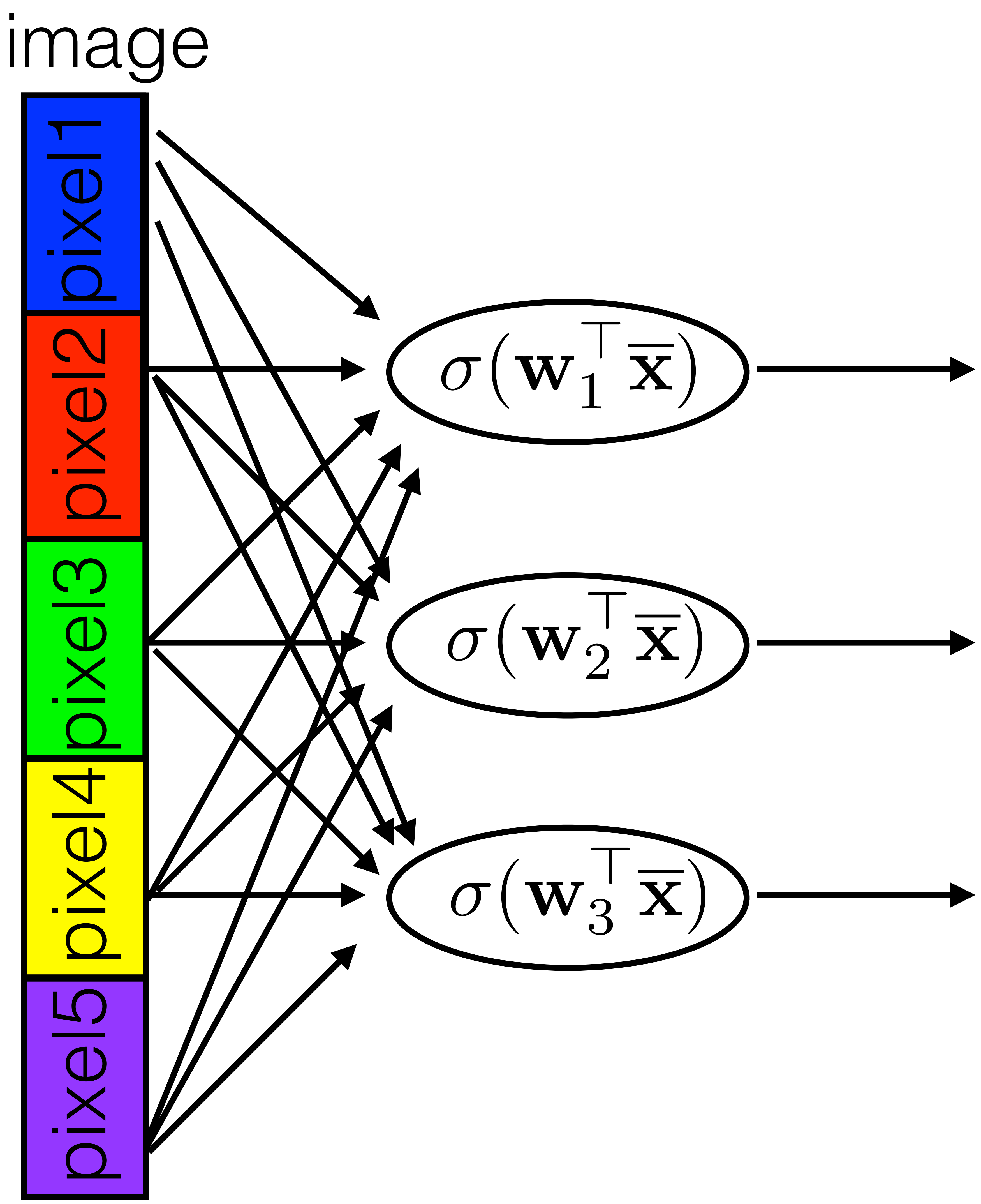
https://knowingneurons.com/2014/10/29/hubel-and-wiesel-the-neural-basis-of-visual-perception/

Hubel and Wiesel experiments in 1950s and 1960s

- Nobel Prize in Physiology and Medicine in 1981
- Dr. Hubel: “There has been a myth that the brain cannot understand itself. It is compared to a man trying to lift himself by his own bootstraps. We feel that is nonsense. The brain can be studied just as the kidney can.”

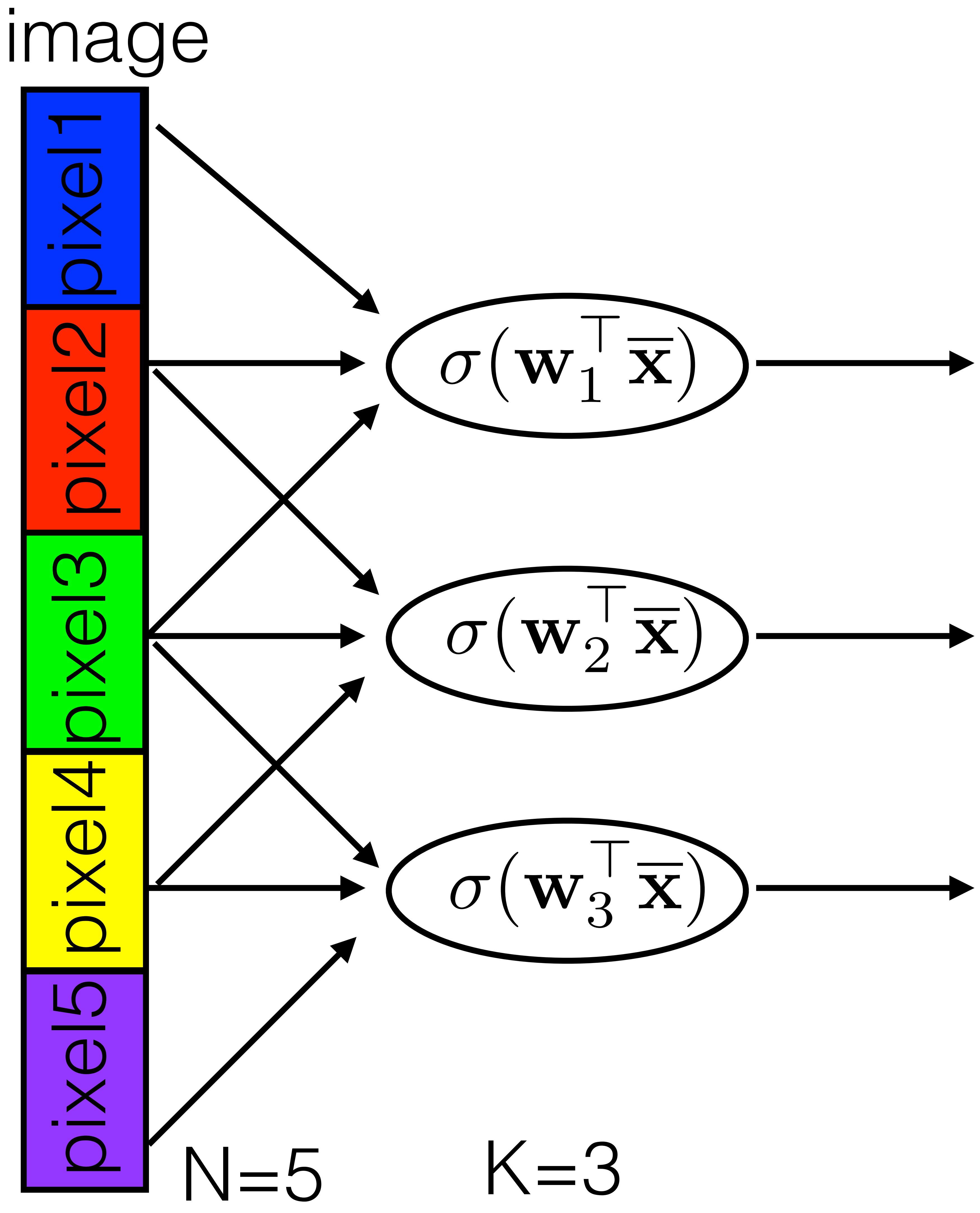
https://knowingneurons.com/2014/10/29/hubel-and-wiesel-the-neural-basis-of-visual-perception/

1. **Topographical map:** nearby neurons process information from nearby visual fields



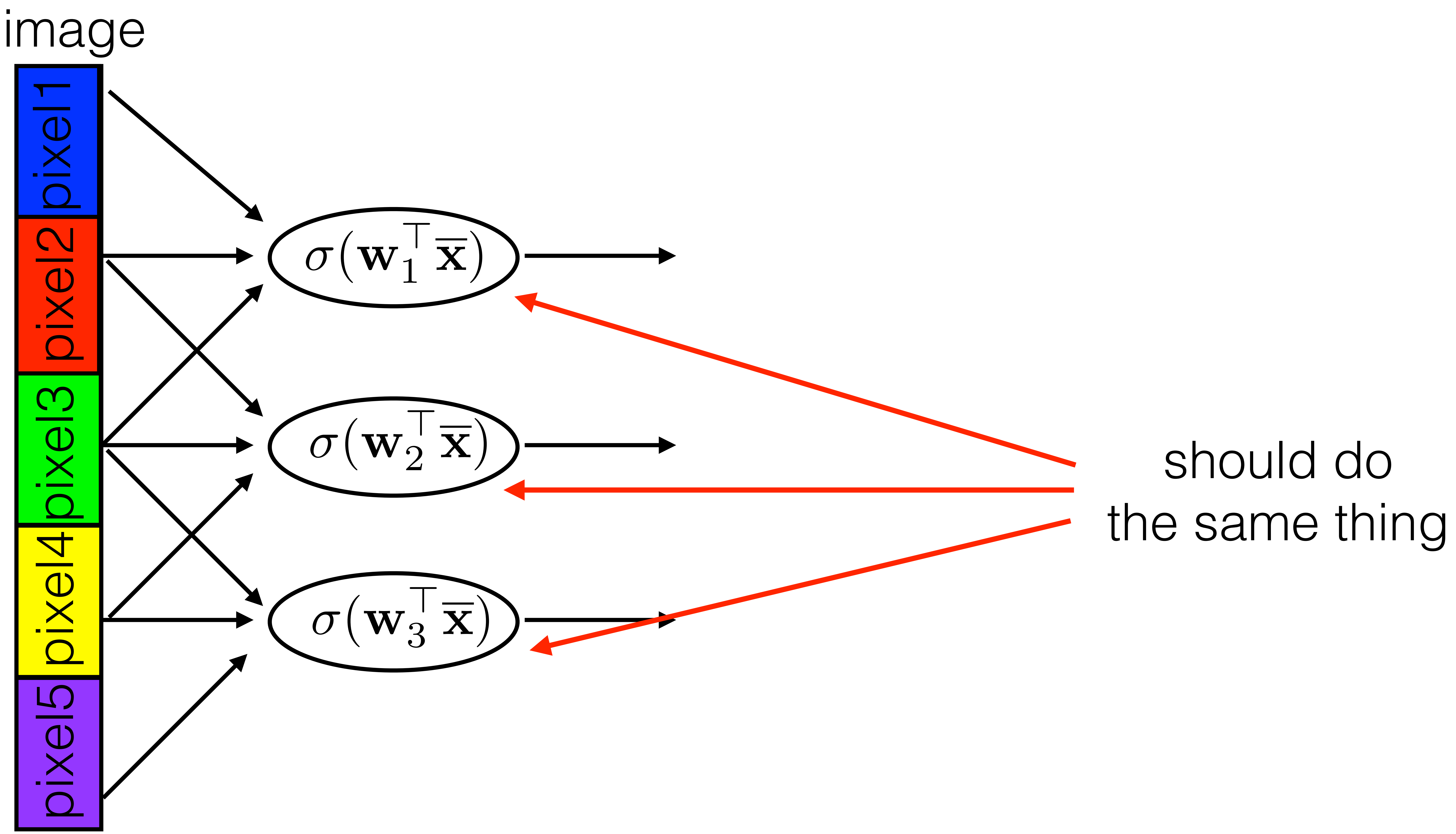
- Processing of visual information in cortex is not fully connected.

1. **Topographical map:** nearby neurons process information from nearby visual fields

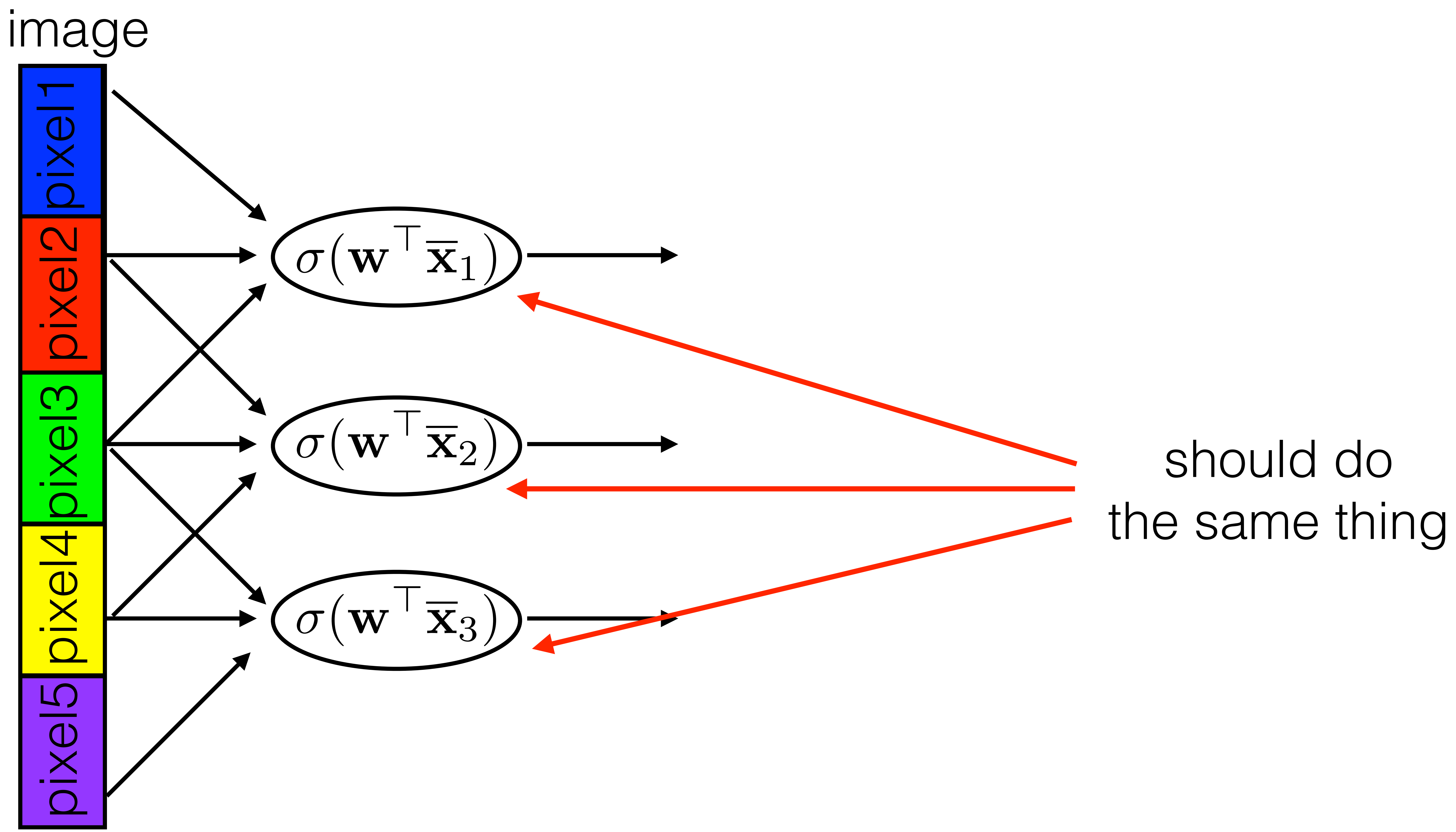


- What is the weight dimensionality reduction for N-pixel image and (n=3)-dimensional spatial neighbourhood?

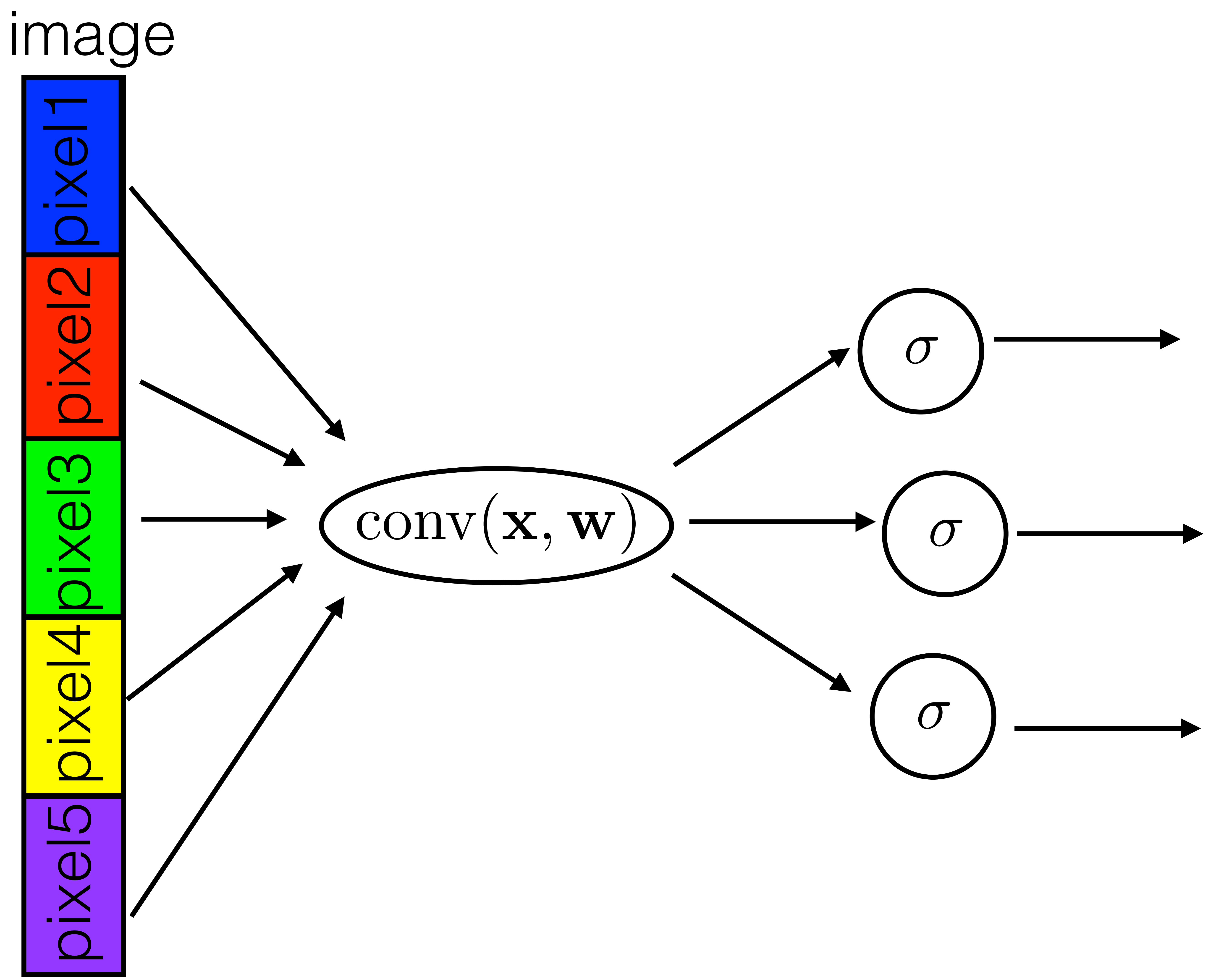
2. **Translation invariance:** the same edge is detected at all positions



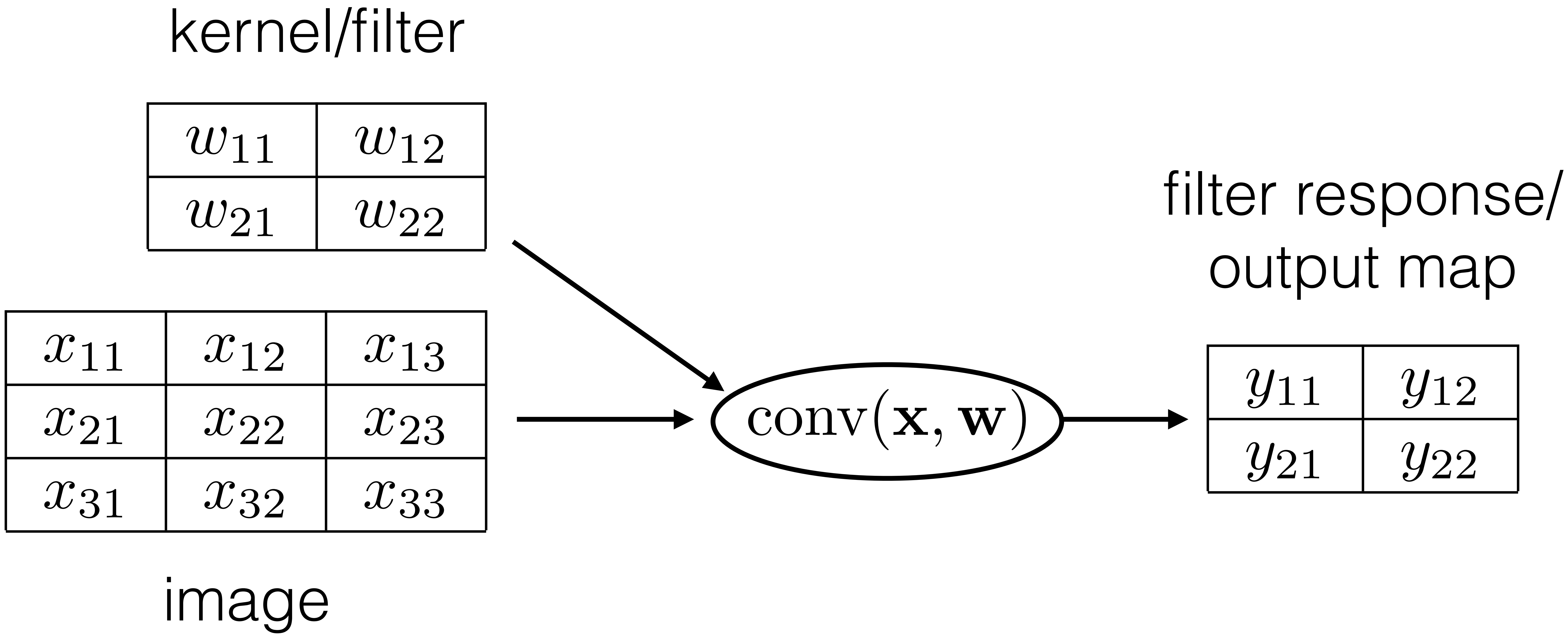
2. **Translation invariance:** the same edge is detected at all positions



2. **Translation invariance:** the same edge is detected at all positions



Convolution forward pass $\mathbf{y} = \mathrm{conv}(\mathbf{x}, \mathbf{w})$



Convolution forward pass $\mathbf{y} = \mathrm{conv}(\mathbf{x}, \mathbf{w})$

$$\begin{array}{|c|c|}\hline y_{11} & y_{12} \\ \hline y_{21} & y_{22} \\ \hline\end{array} = \mathrm{conv}\left(\begin{array}{|c|c|c|}\hline x_{11} & x_{12} & x_{13} \\ \hline x_{21} & x_{22} & x_{23} \\ \hline x_{31} & x_{32} & x_{33} \\ \hline\end{array} , \begin{array}{|c|c|}\hline w_{11} & w_{12} \\ \hline w_{21} & w_{22} \\ \hline\end{array} \right)$$

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

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

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

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

Convolution forward pass $\mathbf{y} = \mathrm{conv}(\mathbf{x}, \mathbf{w})$

$$\begin{bmatrix} \boxed{y_{11}} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} = \mathrm{conv}\left(\begin{bmatrix} \boxed{x_{11}} & \boxed{x_{12}} & x_{13} \\ \boxed{x_{21}} & \boxed{x_{22}} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}, \begin{bmatrix} \boxed{w_{11}} & \boxed{w_{12}} \\ \boxed{w_{21}} & \boxed{w_{22}} \end{bmatrix} \right)$$

$$\boxed{y_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}}$$

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

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

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

Convolution forward pass $\mathbf{y} = \mathrm{conv}(\mathbf{x}, \mathbf{w})$

$$\begin{array}{|c|c|} \hline y_{11} & y_{12} \\ \hline y_{21} & y_{22} \\ \hline \end{array} = \mathrm{conv}\left(\begin{array}{|c|c|c|} \hline x_{11} & x_{12} & x_{13} \\ \hline x_{21} & x_{22} & x_{23} \\ \hline x_{31} & x_{32} & x_{33} \\ \hline \end{array}, \begin{array}{|c|c|} \hline w_{11} & w_{12} \\ \hline w_{21} & w_{22} \\ \hline \end{array} \right)$$

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

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

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

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

Convolution forward pass $\mathbf{y} = \text{conv}(\mathbf{x}, \mathbf{w})$

$$\begin{array}{|c|c|}\hline y_{11} & y_{12} \\ \hline y_{21} & y_{22} \\ \hline \end{array} = \text{conv}\left(\begin{array}{|c|c|c|}\hline x_{11} & x_{12} & x_{13} \\ \hline x_{21} & x_{22} & x_{23} \\ \hline x_{31} & x_{32} & x_{33} \\ \hline \end{array}, \begin{array}{|c|c|}\hline w_{11} & w_{12} \\ \hline w_{21} & w_{22} \\ \hline \end{array} \right)$$

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

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

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

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

Convolution forward pass $\mathbf{y} = \mathrm{conv}(\mathbf{x}, \mathbf{w})$

$$\begin{array}{|c|c|}\hline y_{11} & y_{12} \\ \hline y_{21} & y_{22} \\ \hline \end{array} = \mathrm{conv}\left(\begin{array}{|c|c|c|}\hline x_{11} & x_{12} & x_{13} \\ \hline x_{21} & x_{22} & x_{23} \\ \hline x_{31} & x_{32} & x_{33} \\ \hline \end{array}, \begin{array}{|c|c|}\hline w_{11} & w_{12} \\ \hline w_{21} & w_{22} \\ \hline \end{array} \right)$$

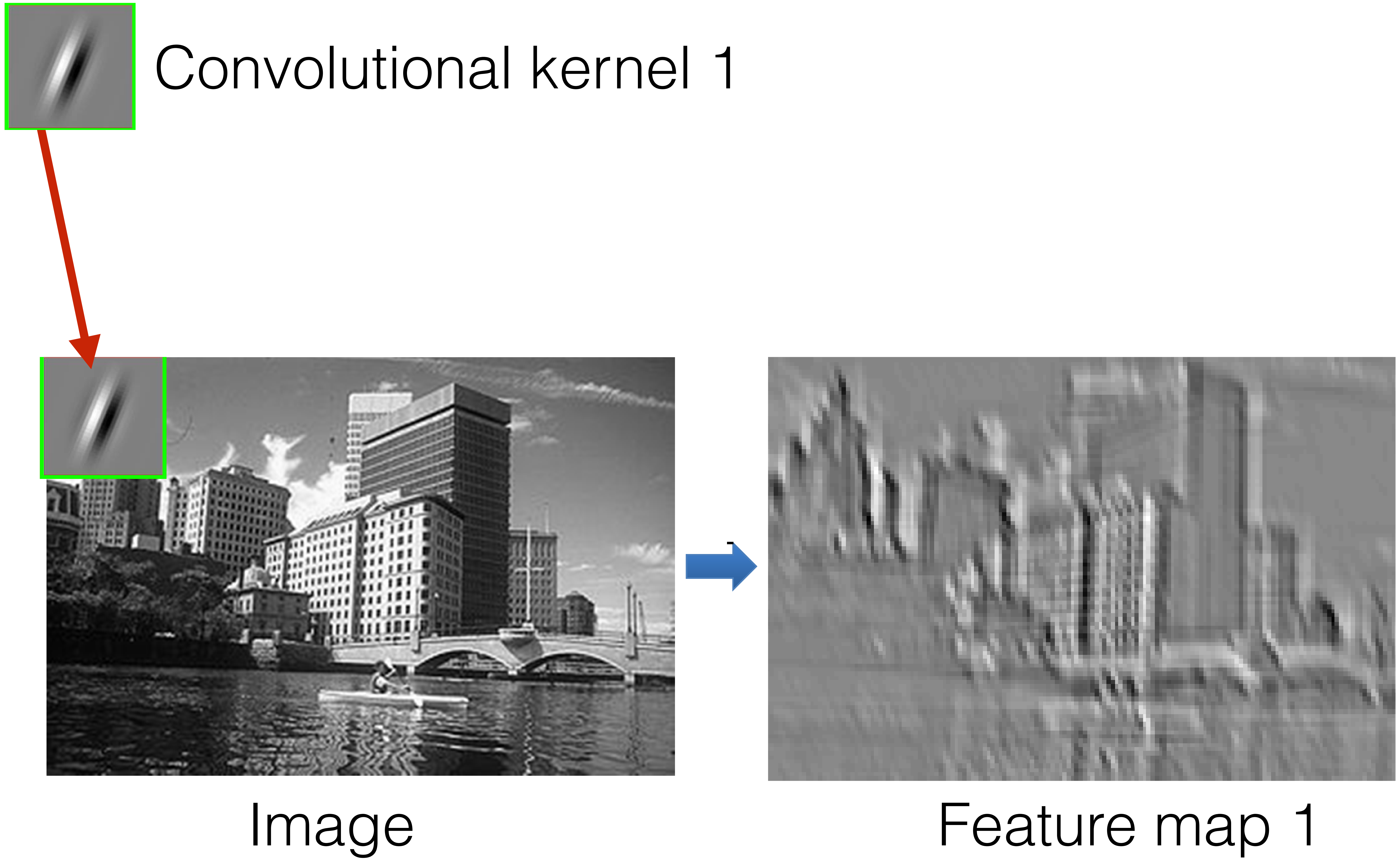
$$y_{11} = w_{11}x_{11} + w_{12}x_{12} + w_{21}x_{21} + w_{22}x_{22}$$

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

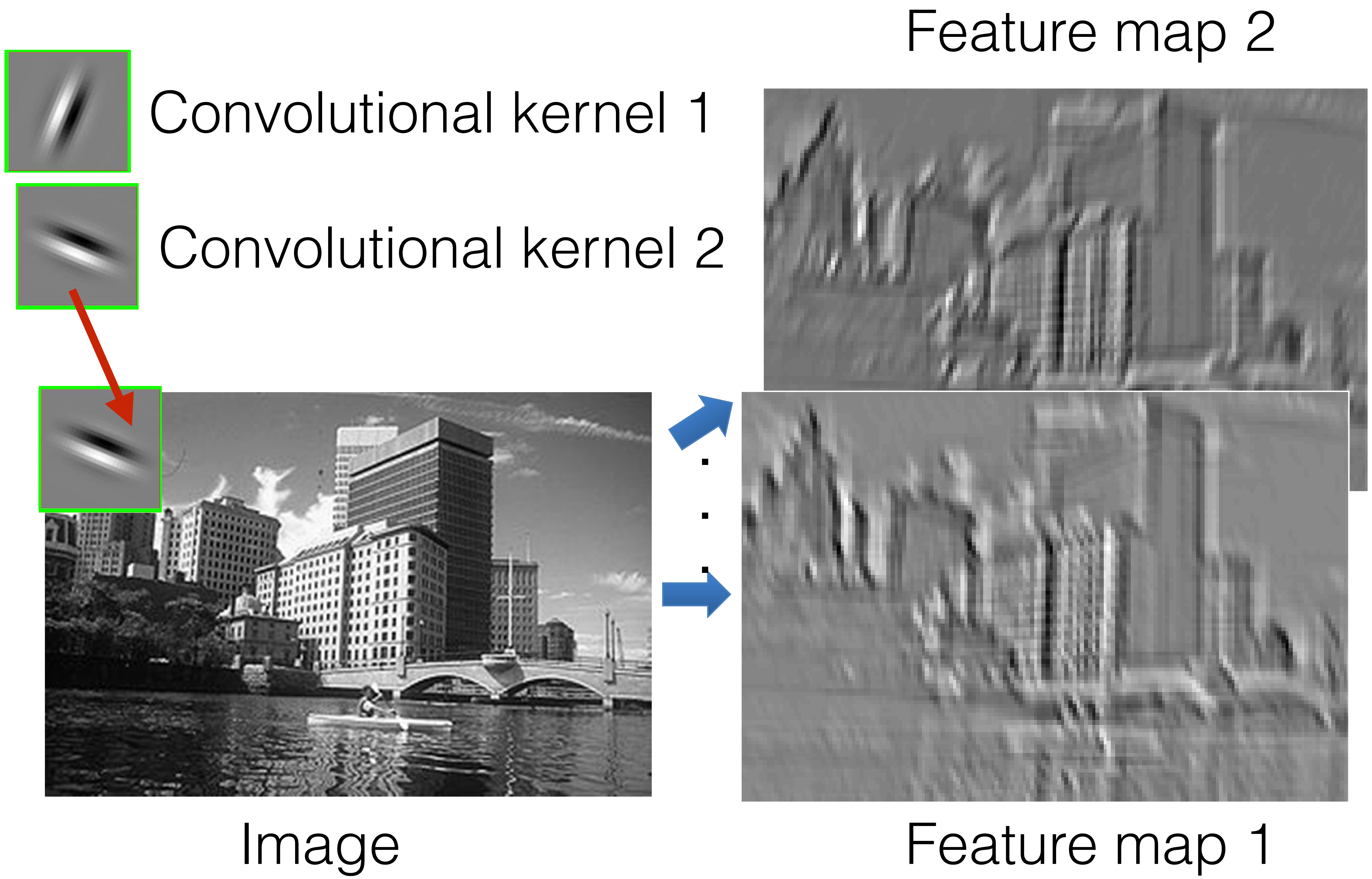
$$y_{21} = w_{11}x_{21} + w_{12}x_{22} + w_{21}x_{31} + w_{22}x_{32}$$

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

Feature maps



Feature maps



Convolution layer properties - output size

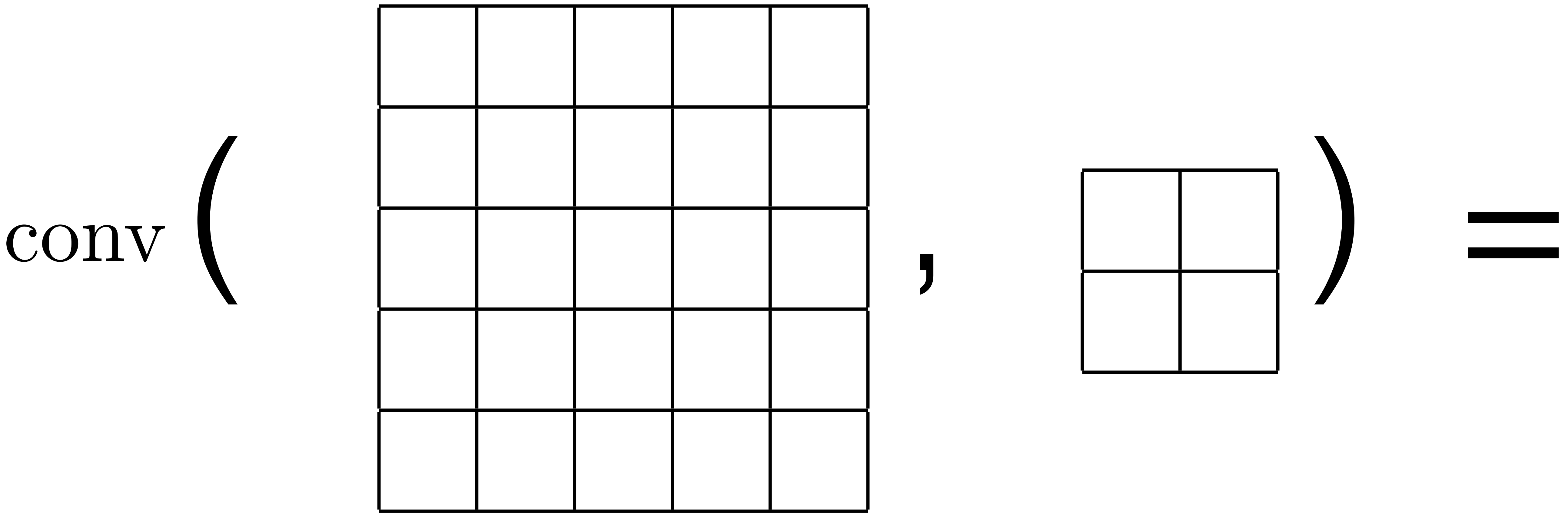
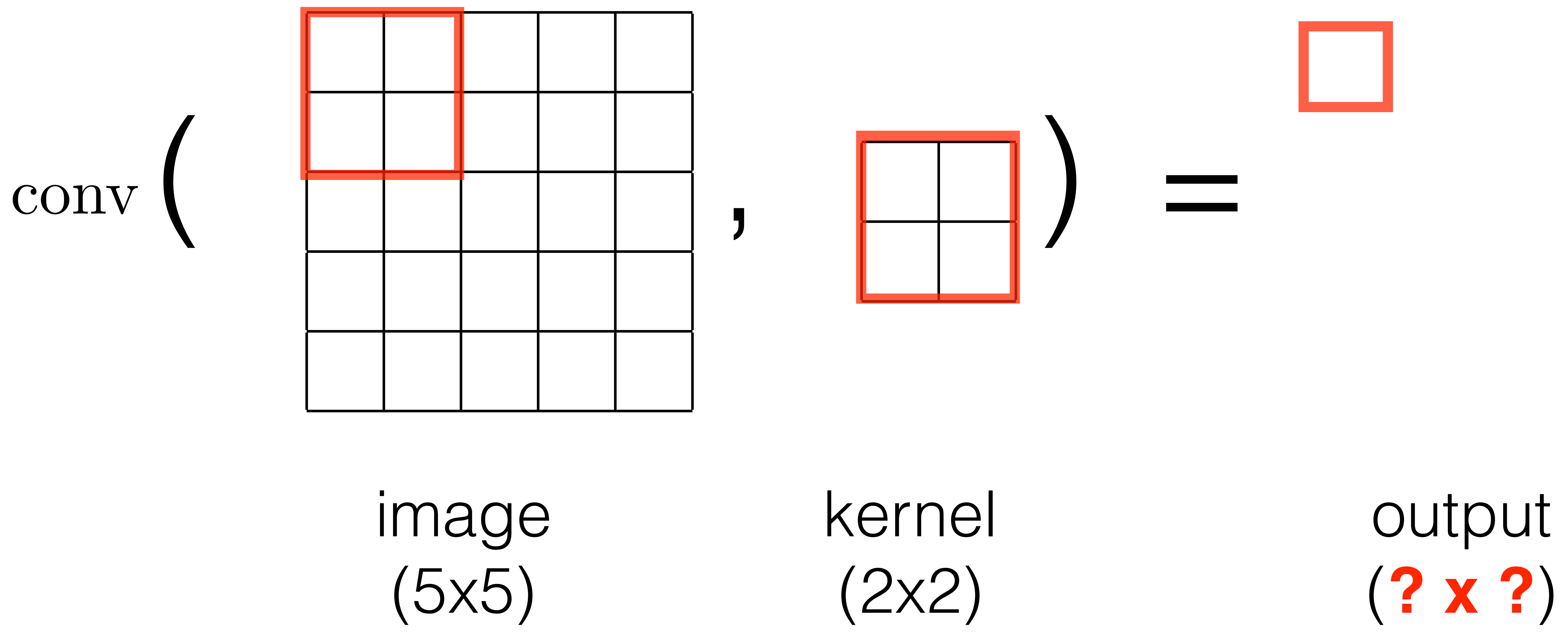


image (5x5) | kernel (2x2) | output (**? x ?**)

Convolution layer properties - output size



Convolution layer properties - output size

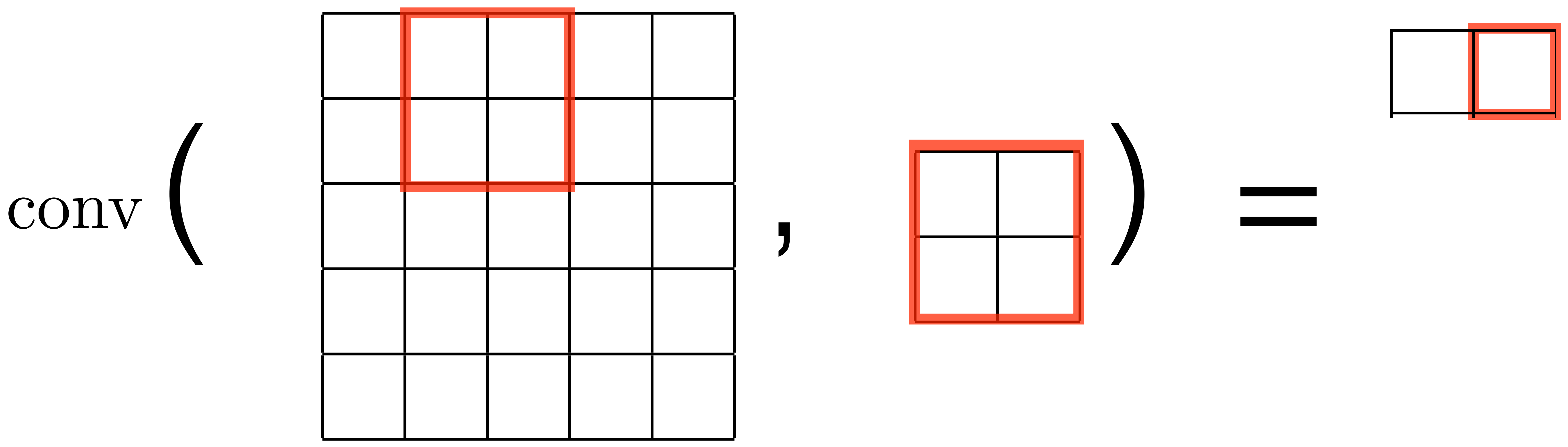


image (5x5)

kernel (2x2)

output (**? x ?**)

Convolution layer properties - output size

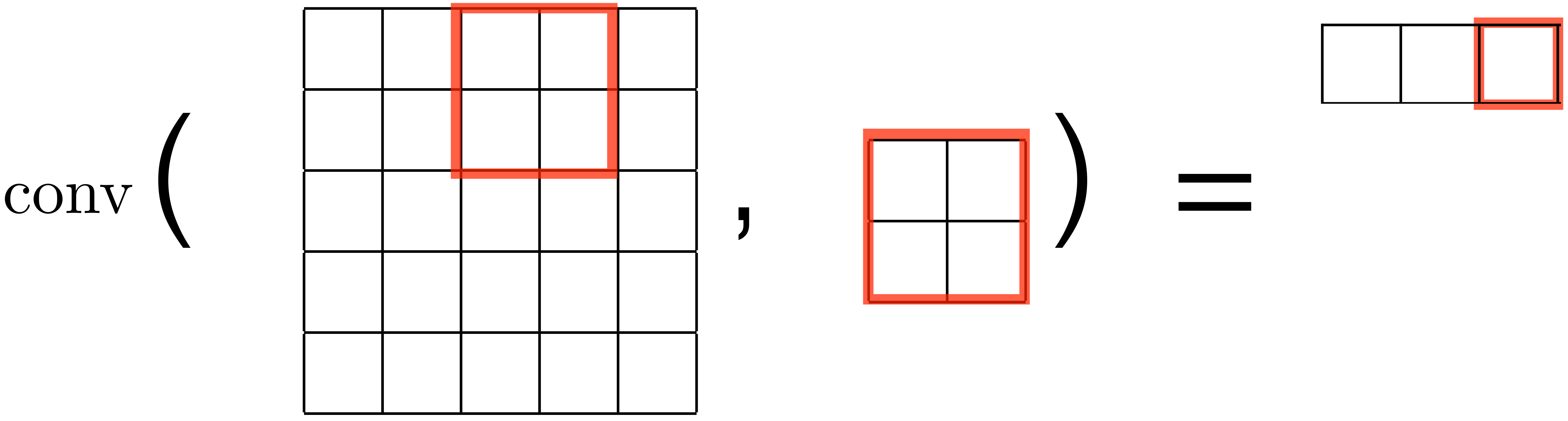


image (5x5)

kernel (2x2)

output (**? x ?**)

Convolution layer properties - output size

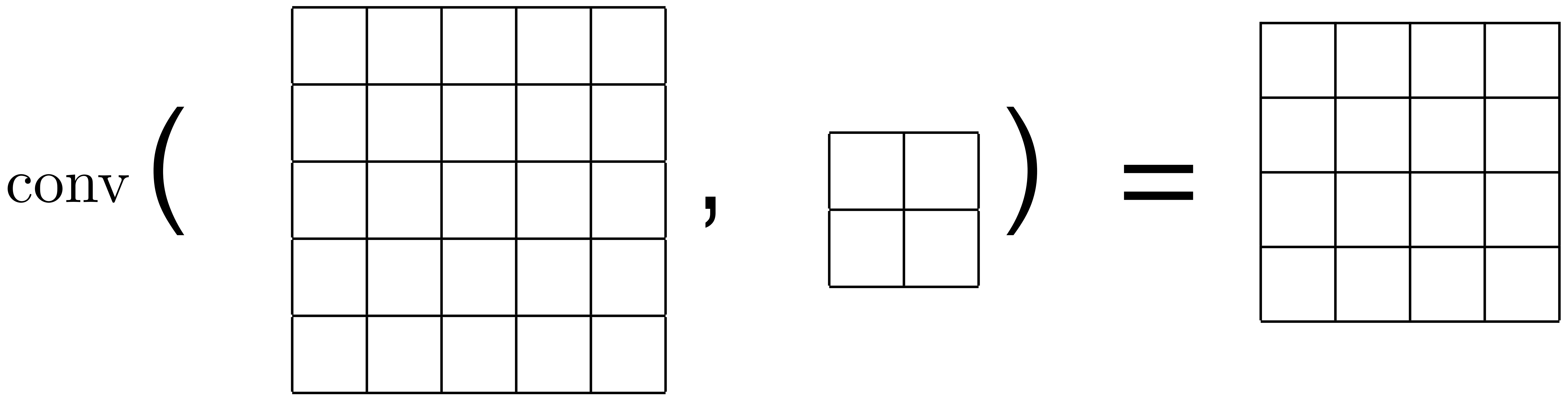


image
(5x5)

kernel
(2x2)

output
(4x4)

Convolution layer properties - output size

$$M = N - K + 1$$

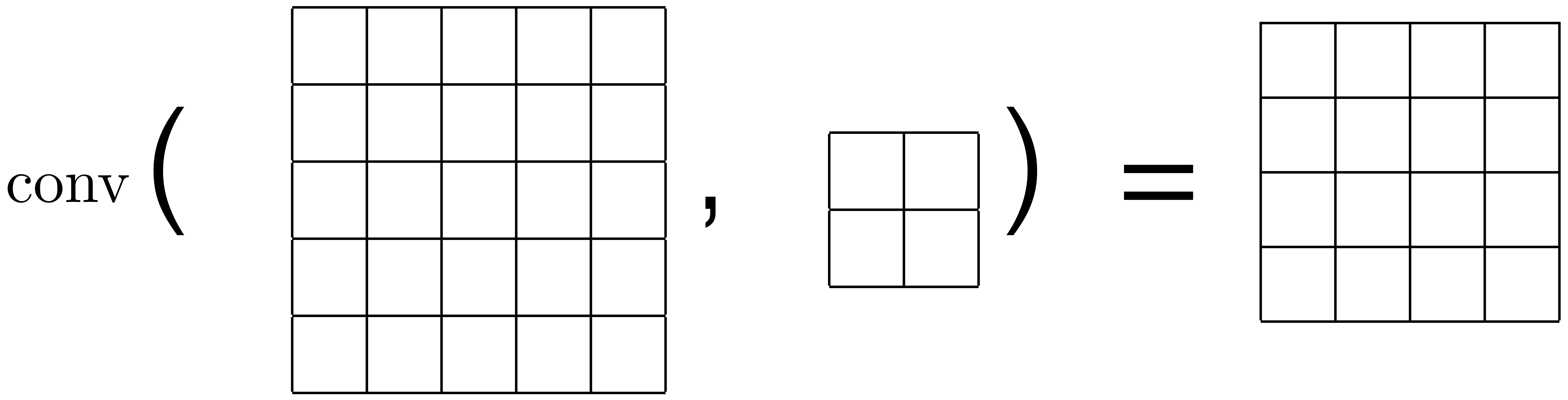


image (NxN)

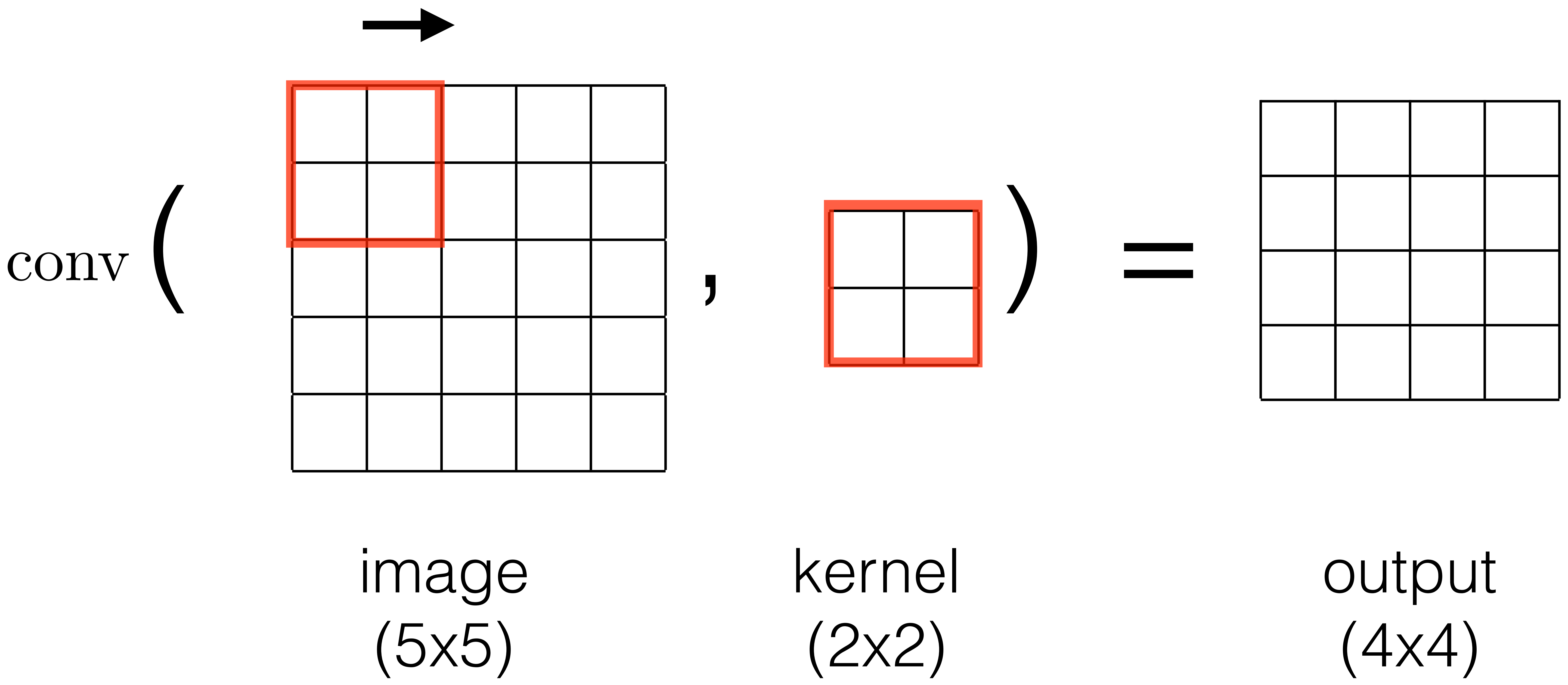
kernel (KxK)

output (MxM)

Convolution layer properties - stride

stride = 1

kernel moves by 1 pixel



Convolution layer properties - stride

stride = 3

kernel moves by 3 pixels

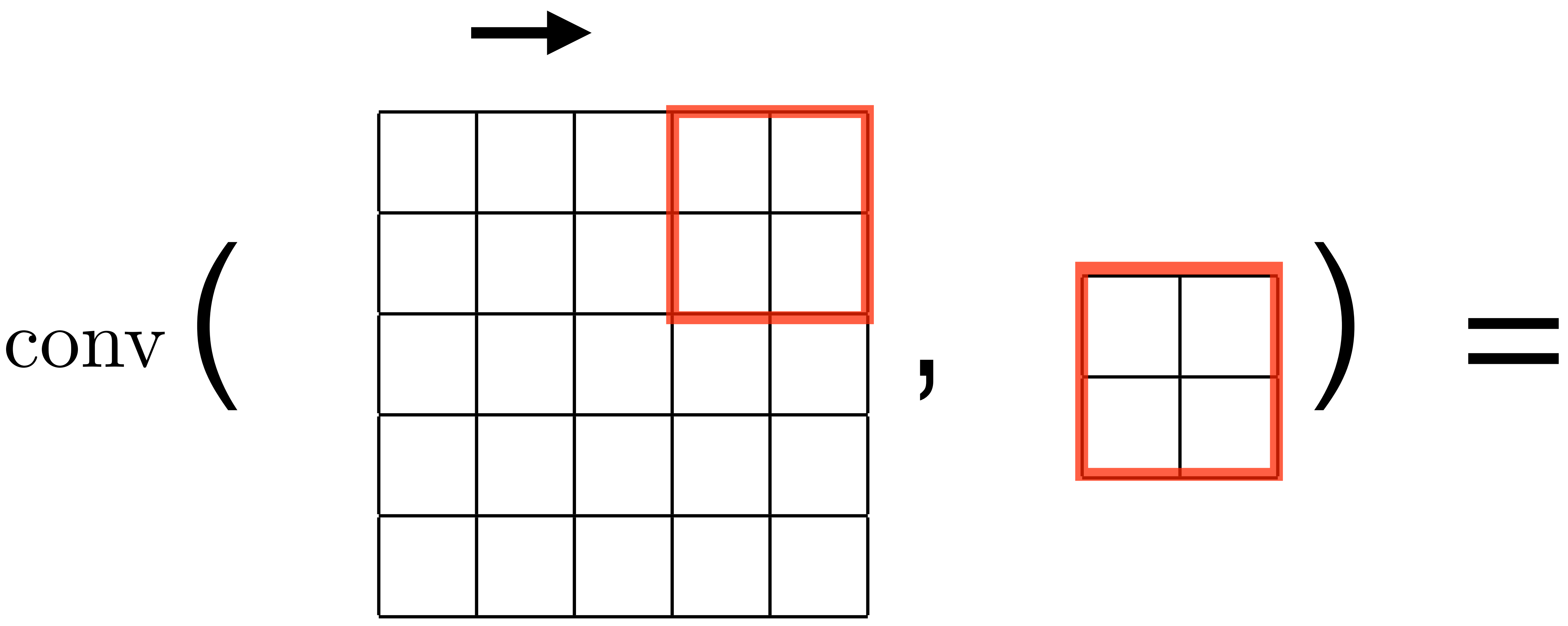


image
(5x5)

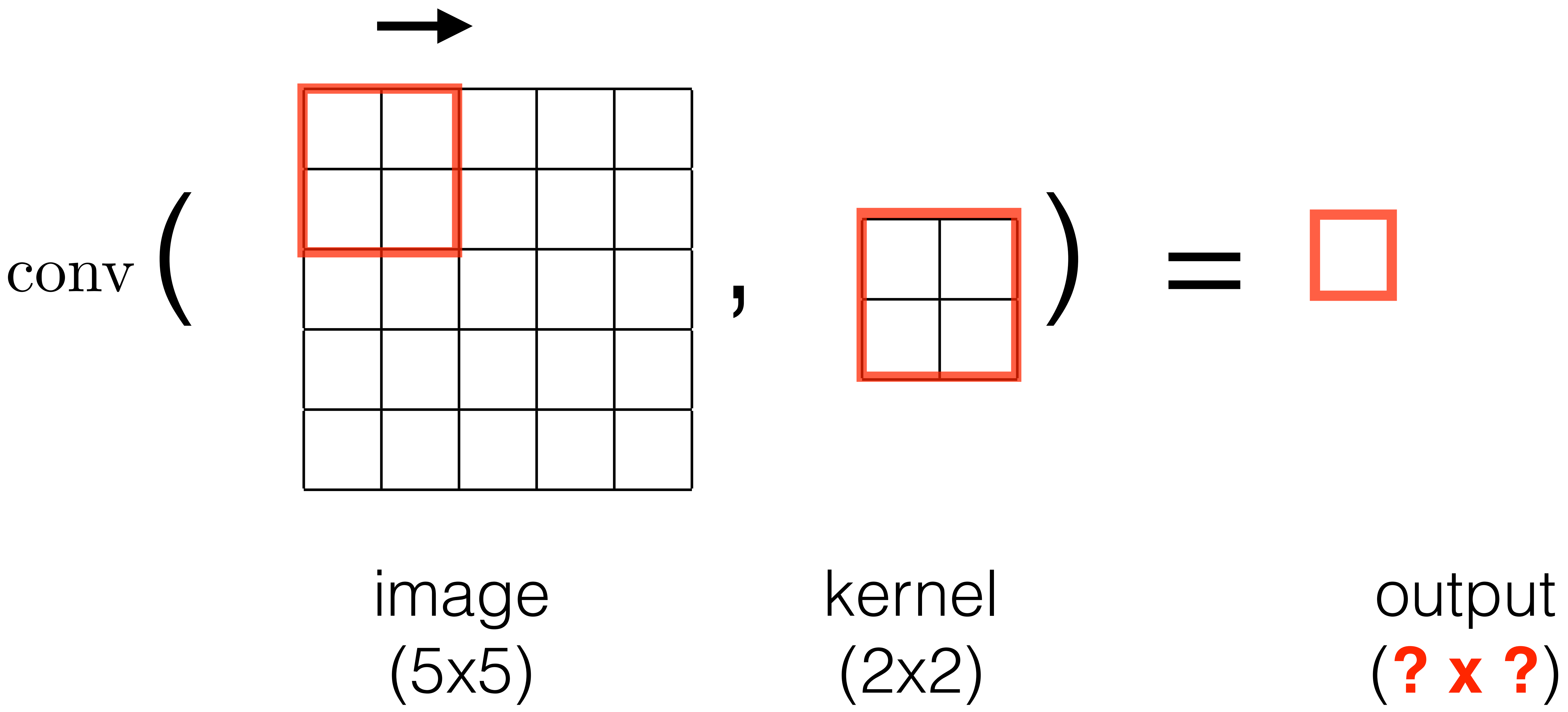
kernel
(2x2)

output
(**? x ?**)

Convolution layer properties - stride

stride = 3

kernel moves by 3 pixels



Convolution layer properties - stride

stride = 3

kernel moves by 3 pixels

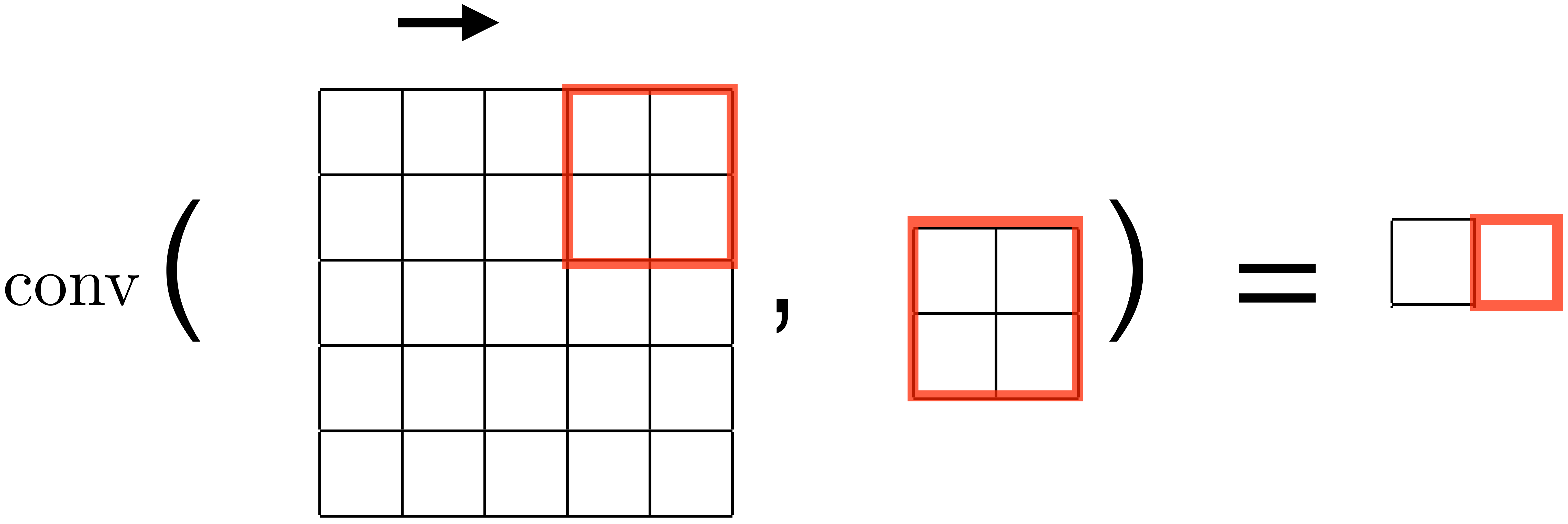
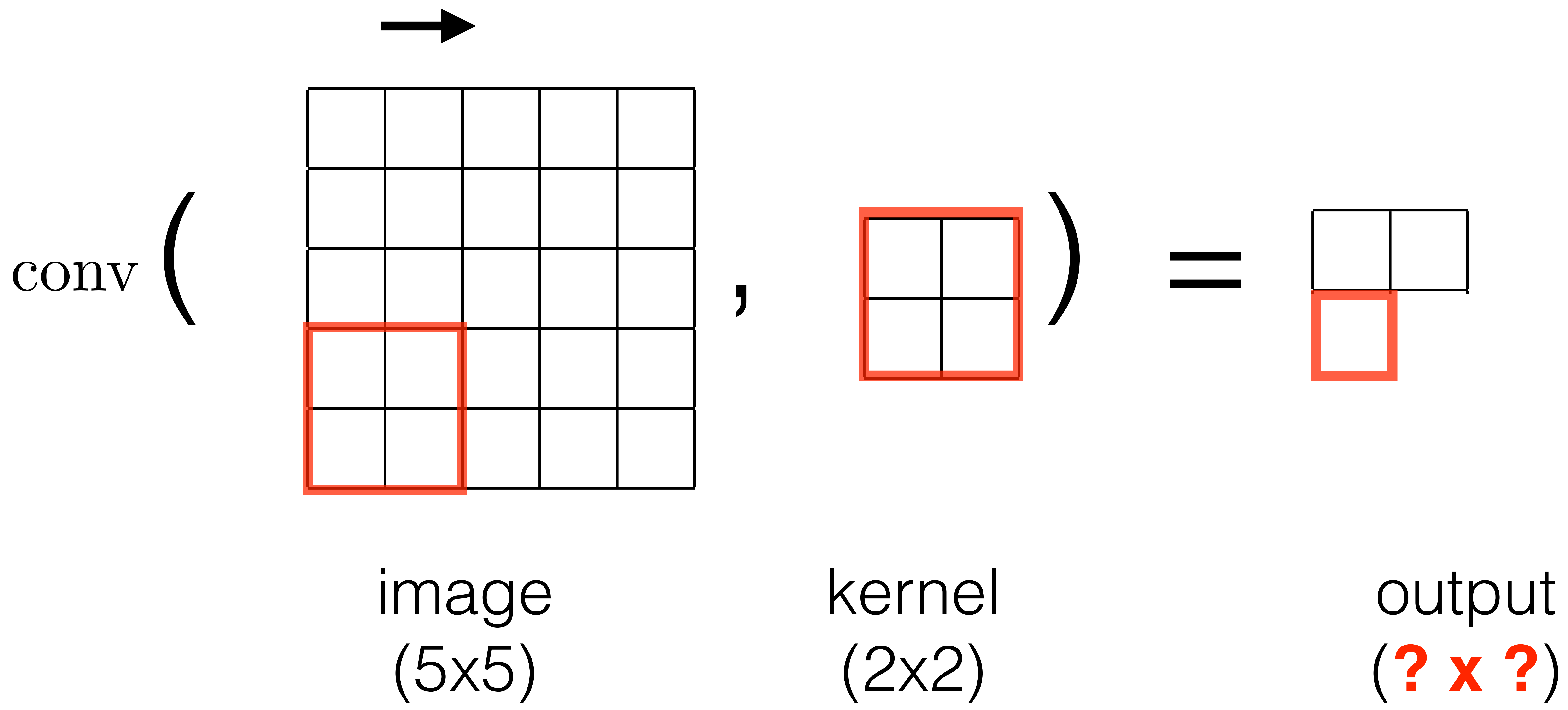


image	kernel	output
(5x5)	(2x2)	(? x ?)

Convolution layer properties - stride

stride = 3

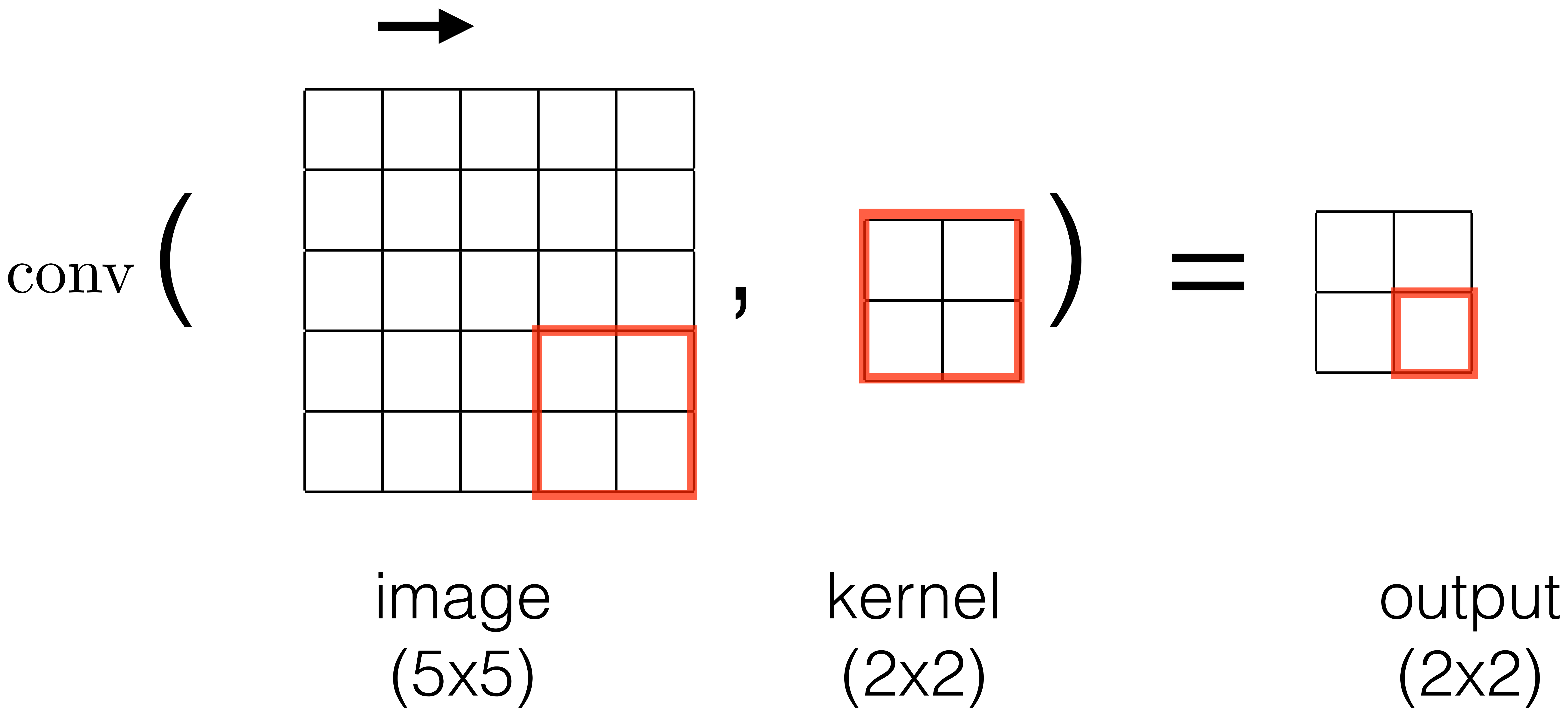
kernel moves by 3 pixels



Convolution layer properties - stride

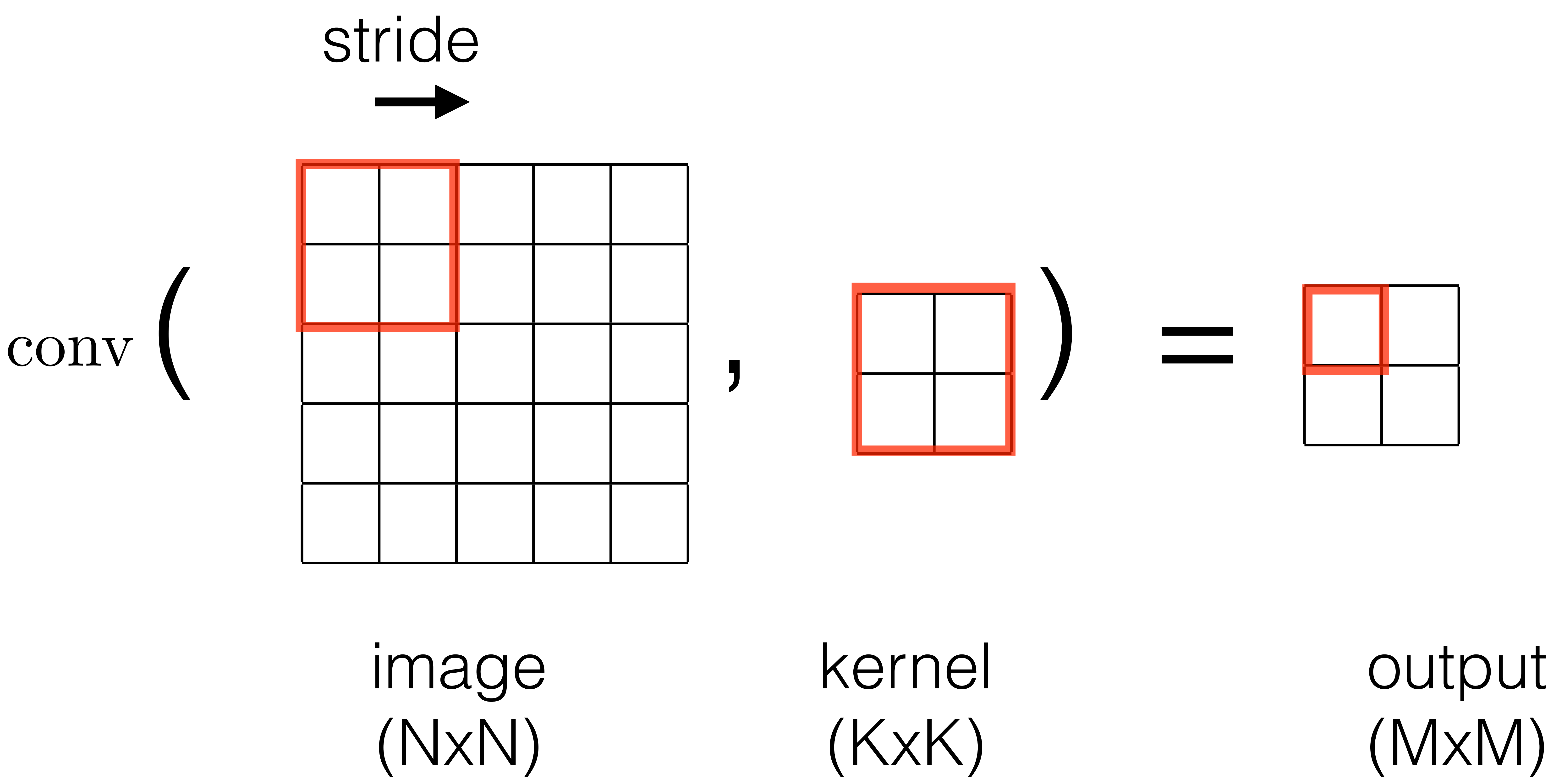
stride = 3

kernel moves by 3 pixels



Convolution layer properties - stride

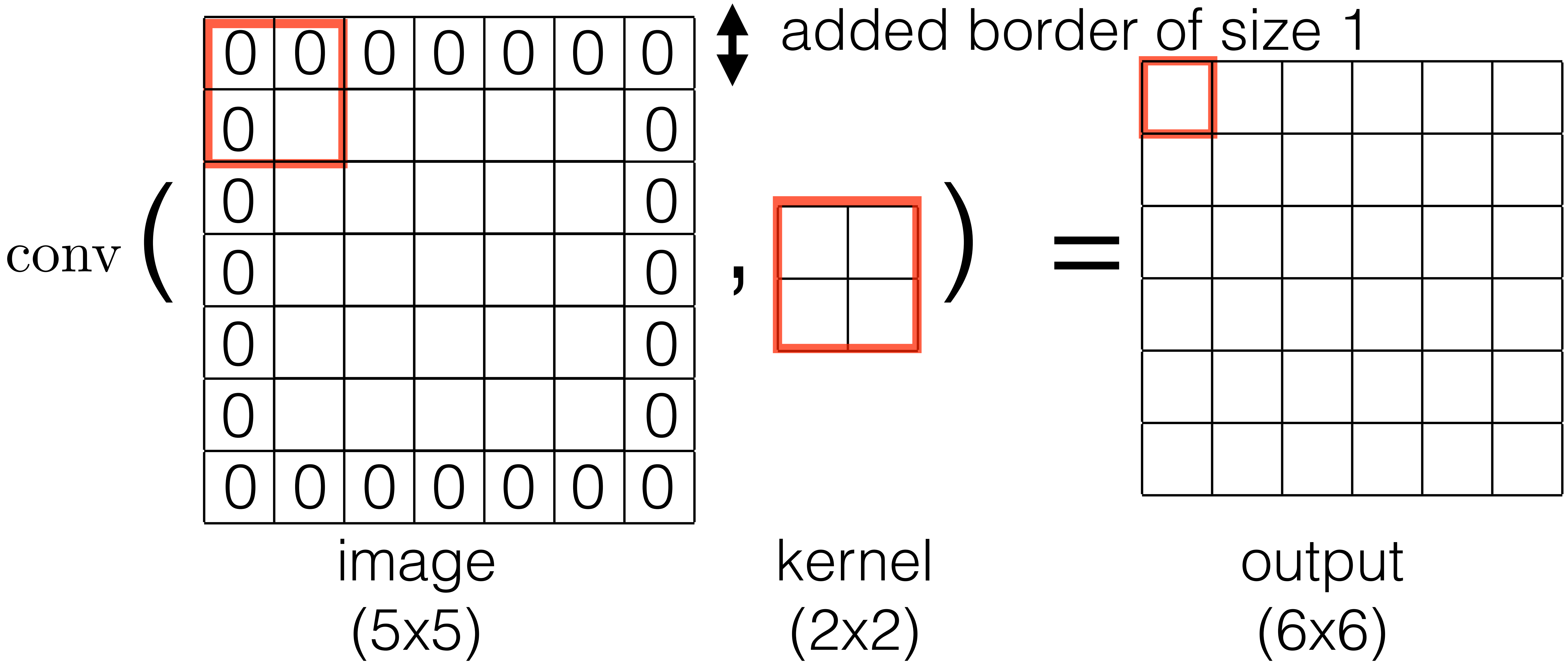
M = **floor**((N-K) / stride + 1)



e.g. M = (5-2) / 3 +1 = 2

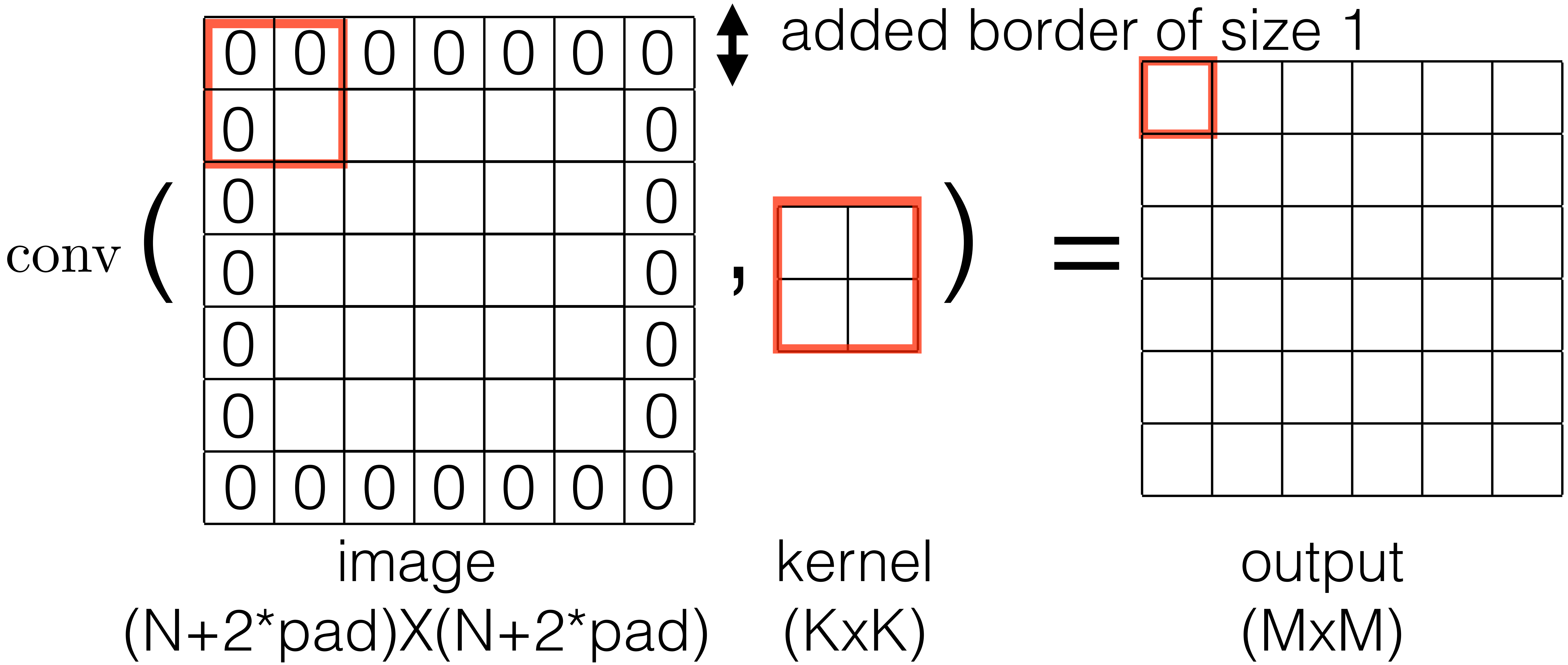
Convolution layer properties - pad

pad = 1



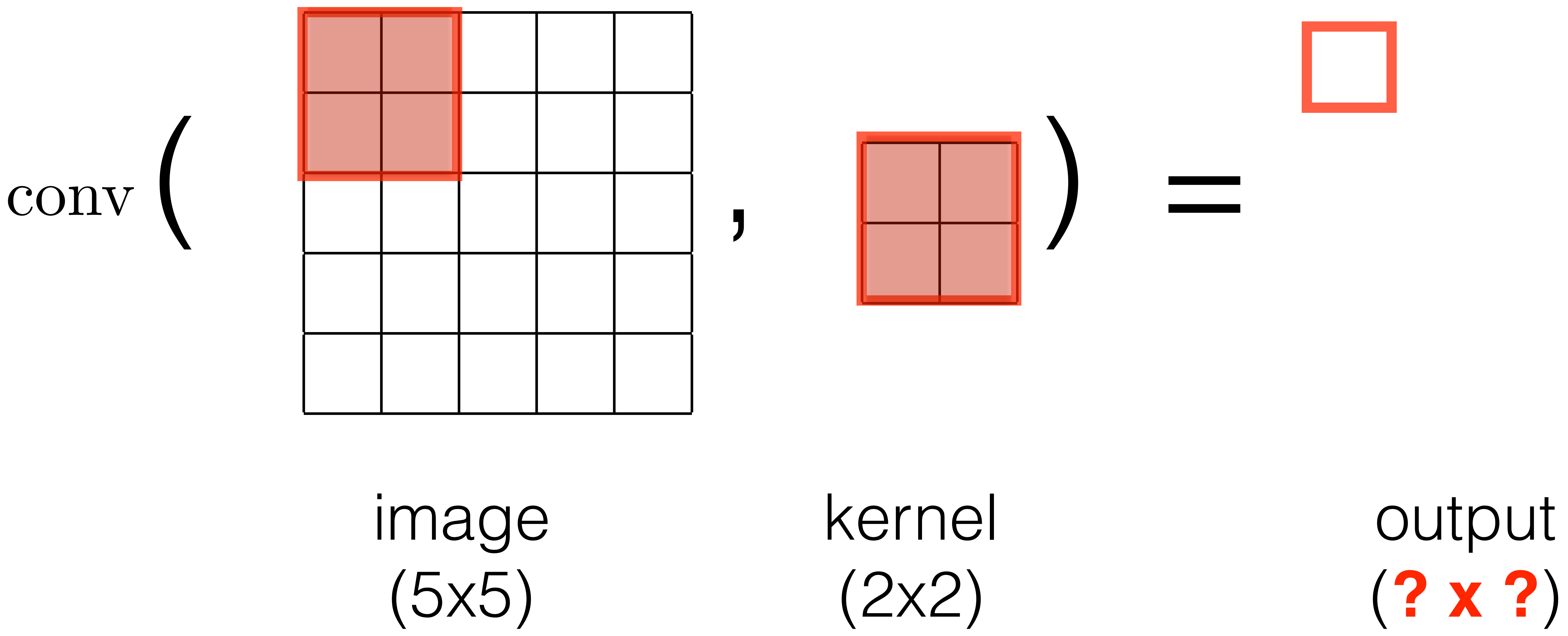
Convolution layer properties - pad

M = **`floor`**((N+2*pad-K) / stride + 1)



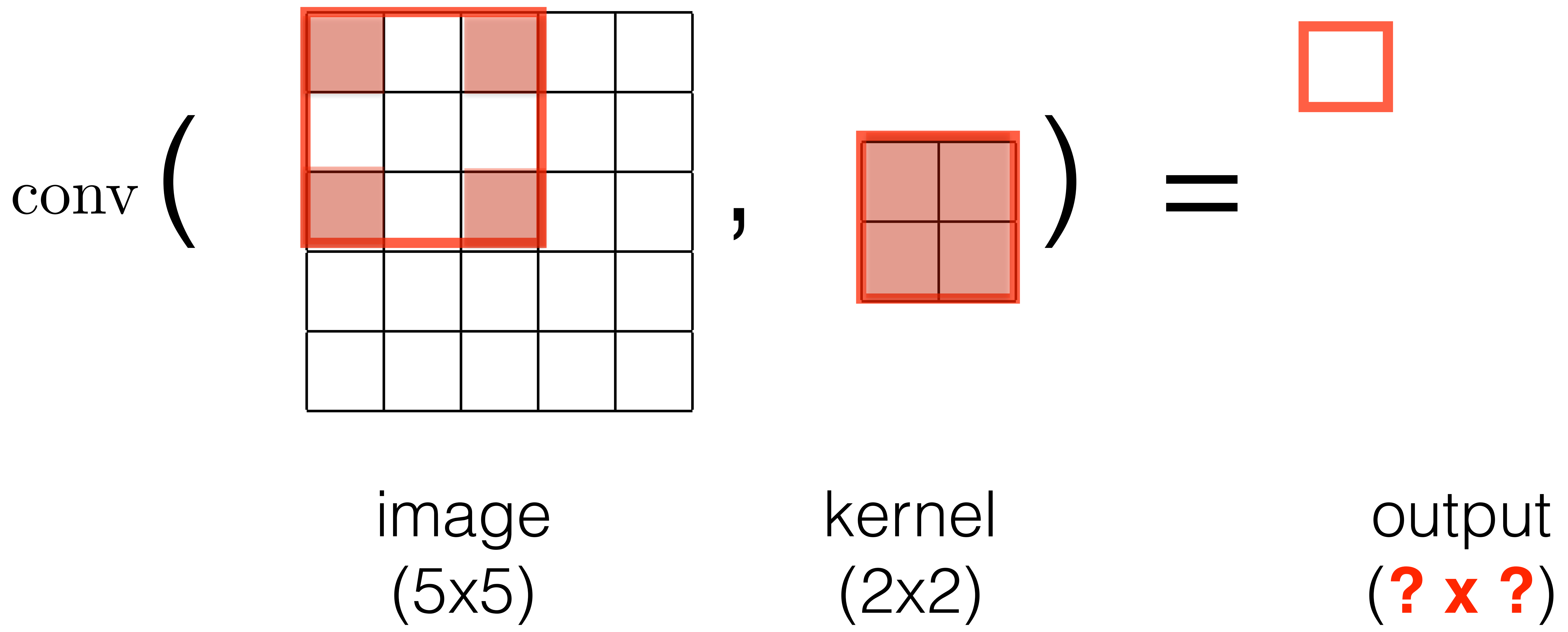
Convolution layer

Dilatation rate = 1



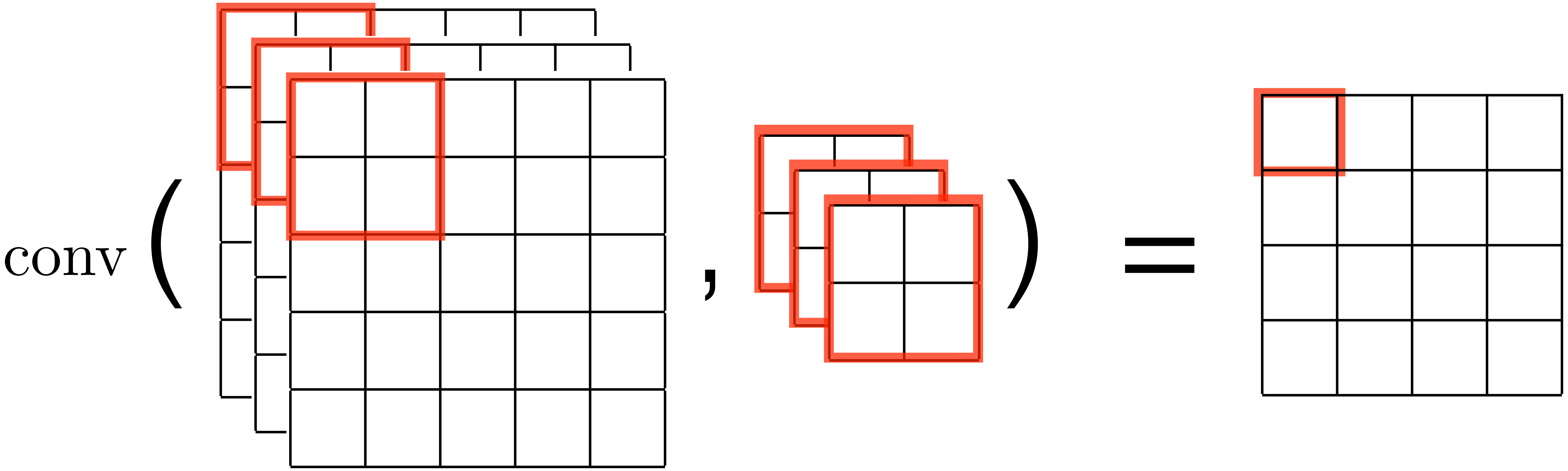
Atrous convolution layer

Dilatation rate = 2



Show python code

Multi-channel convolution

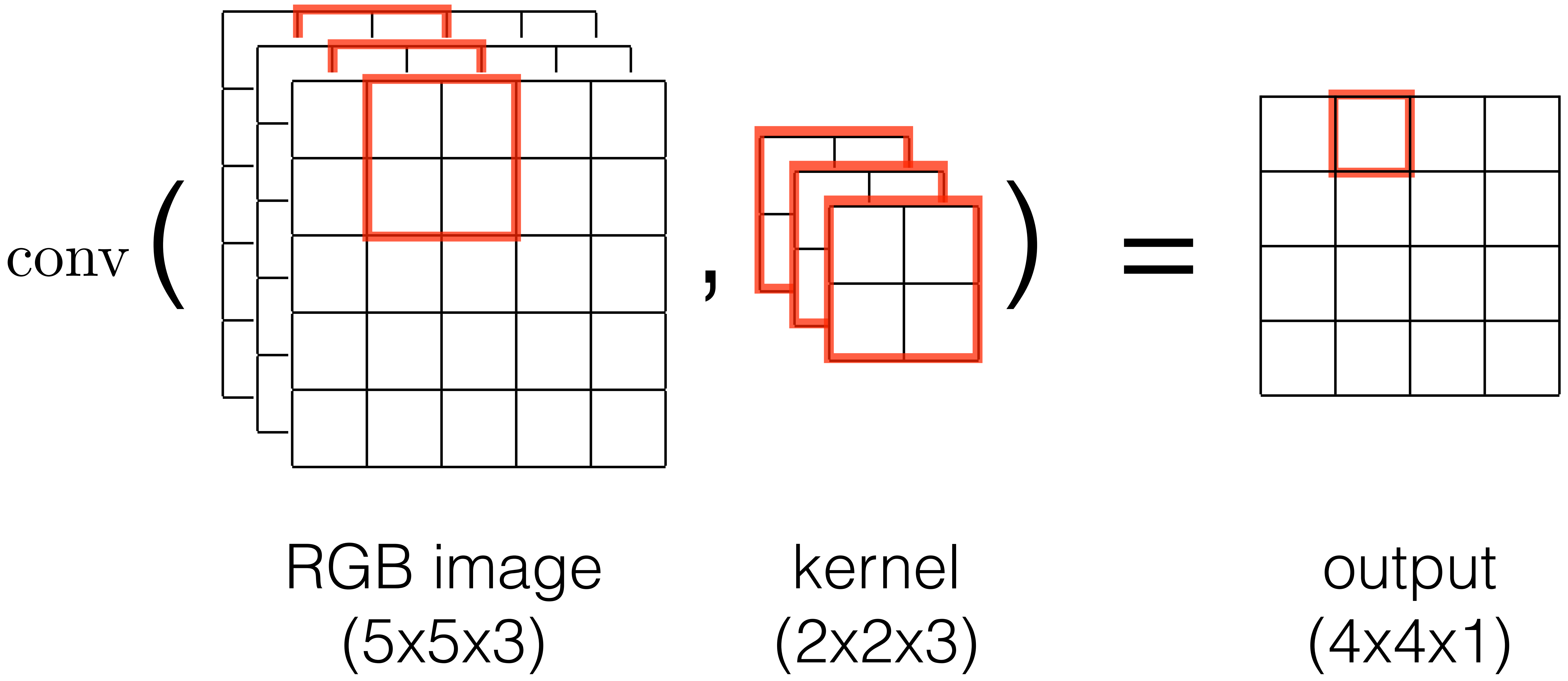


RGB image
(5x5x3)

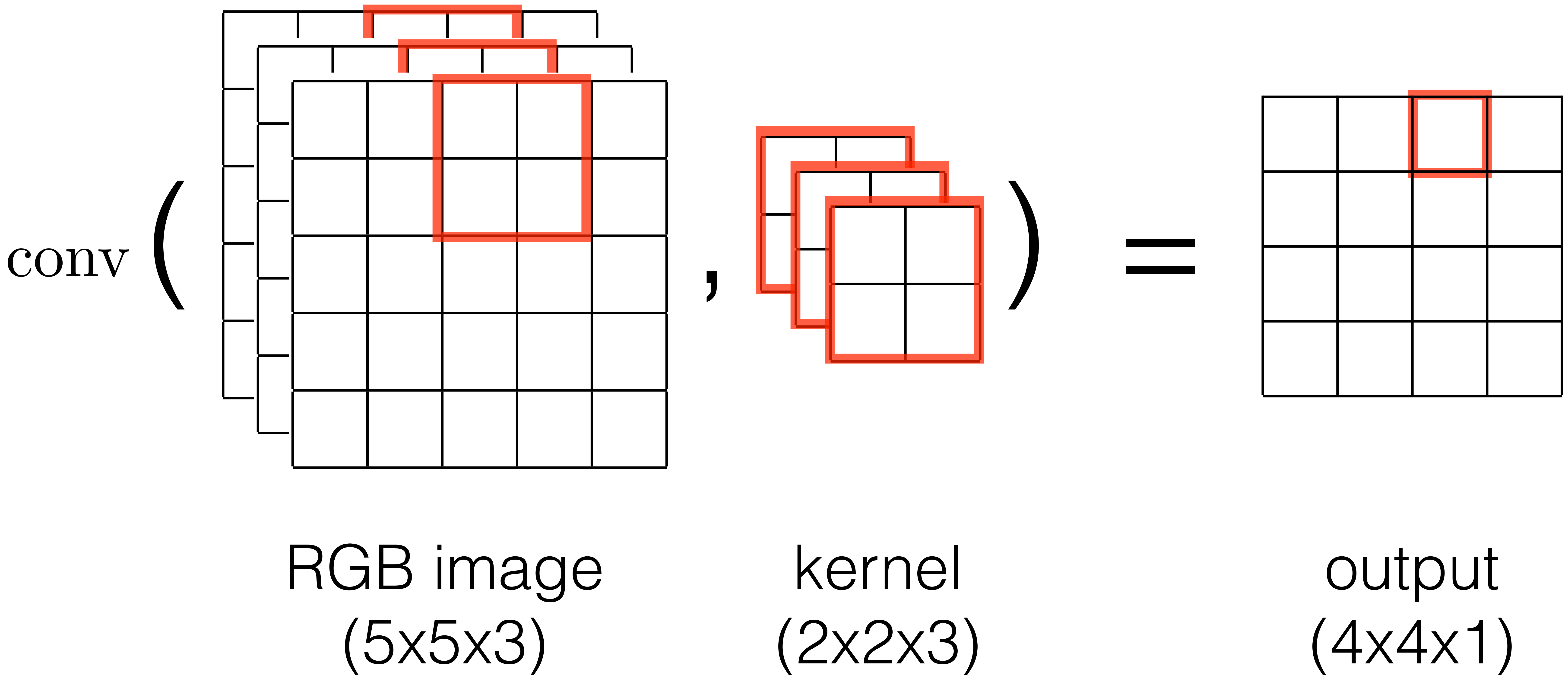
kernel
(2x2x3)

output
(4x4x1)

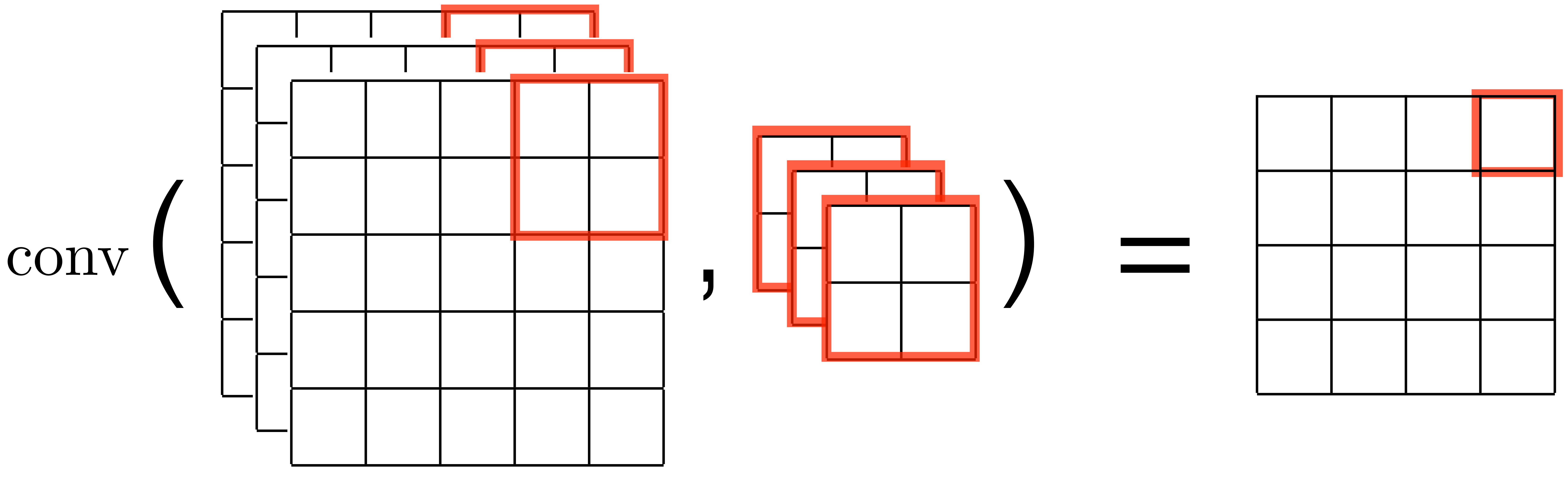
Multi-channel convolution



Multi-channel convolution



Multi-channel convolution

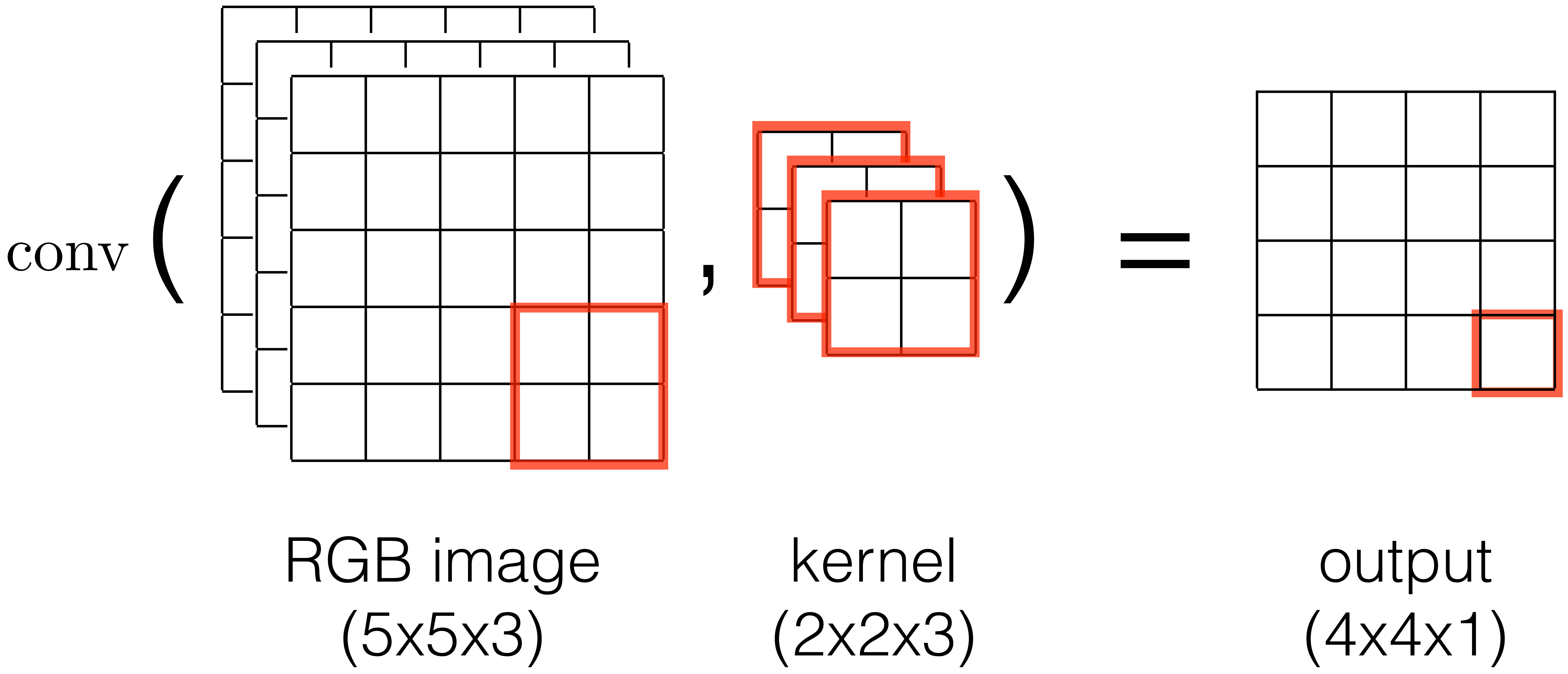


RGB image
(5x5x3)

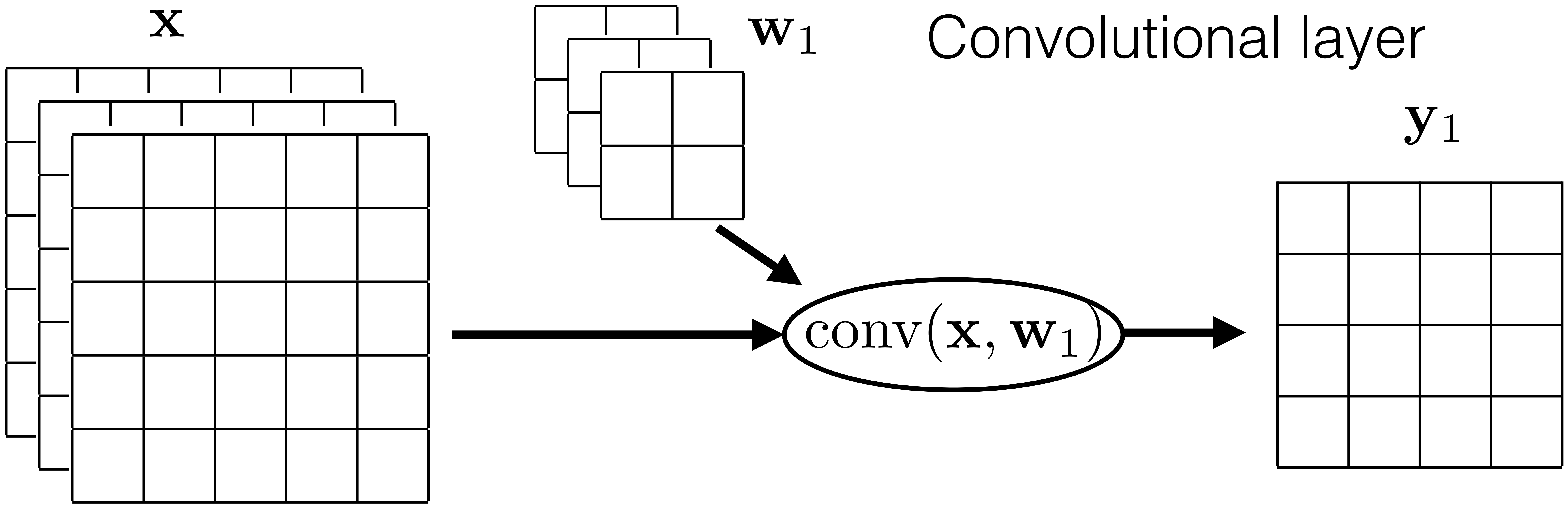
kernel
(2x2x3)

output
(4x4x1)

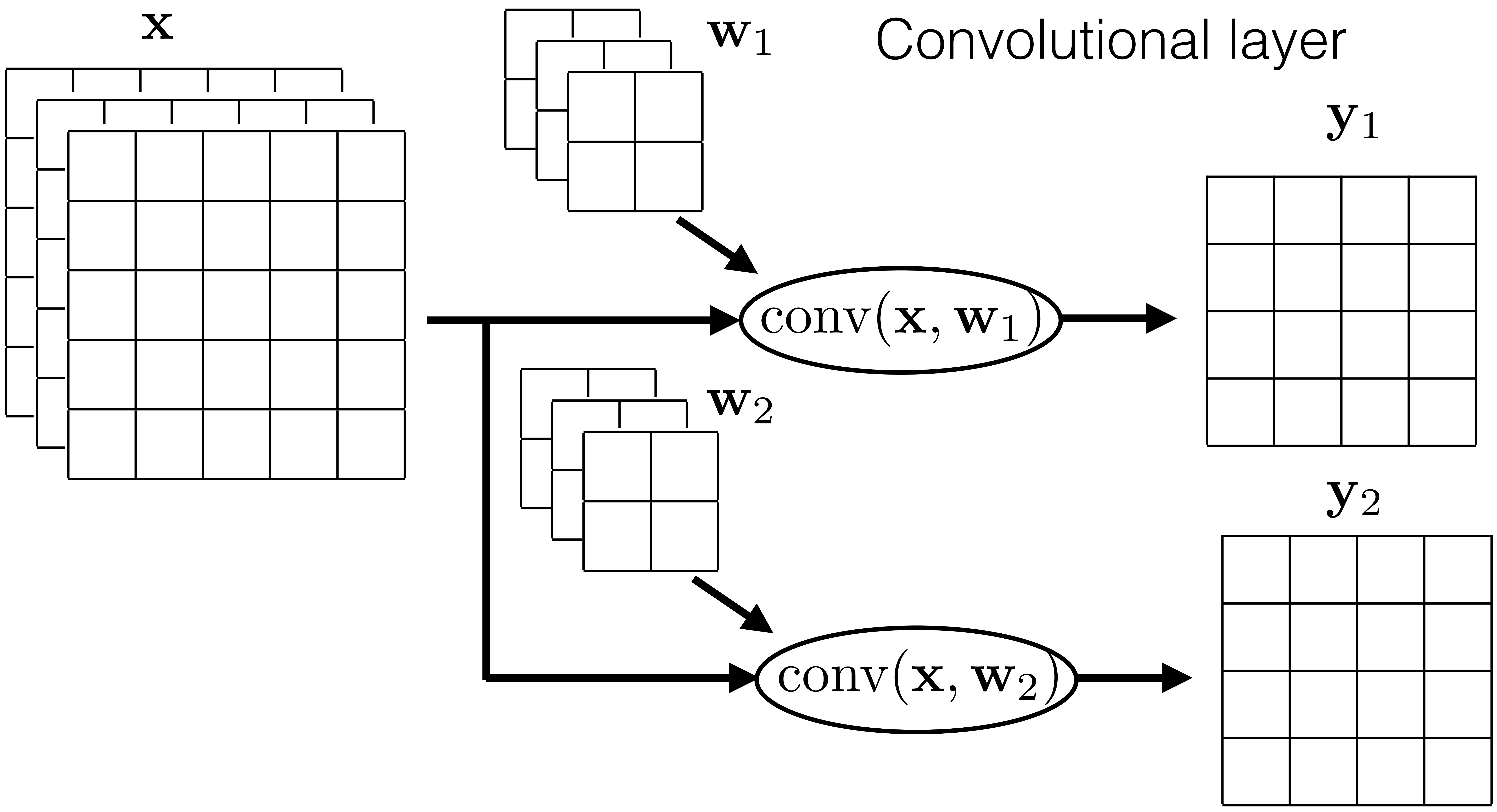
Multi-channel convolution



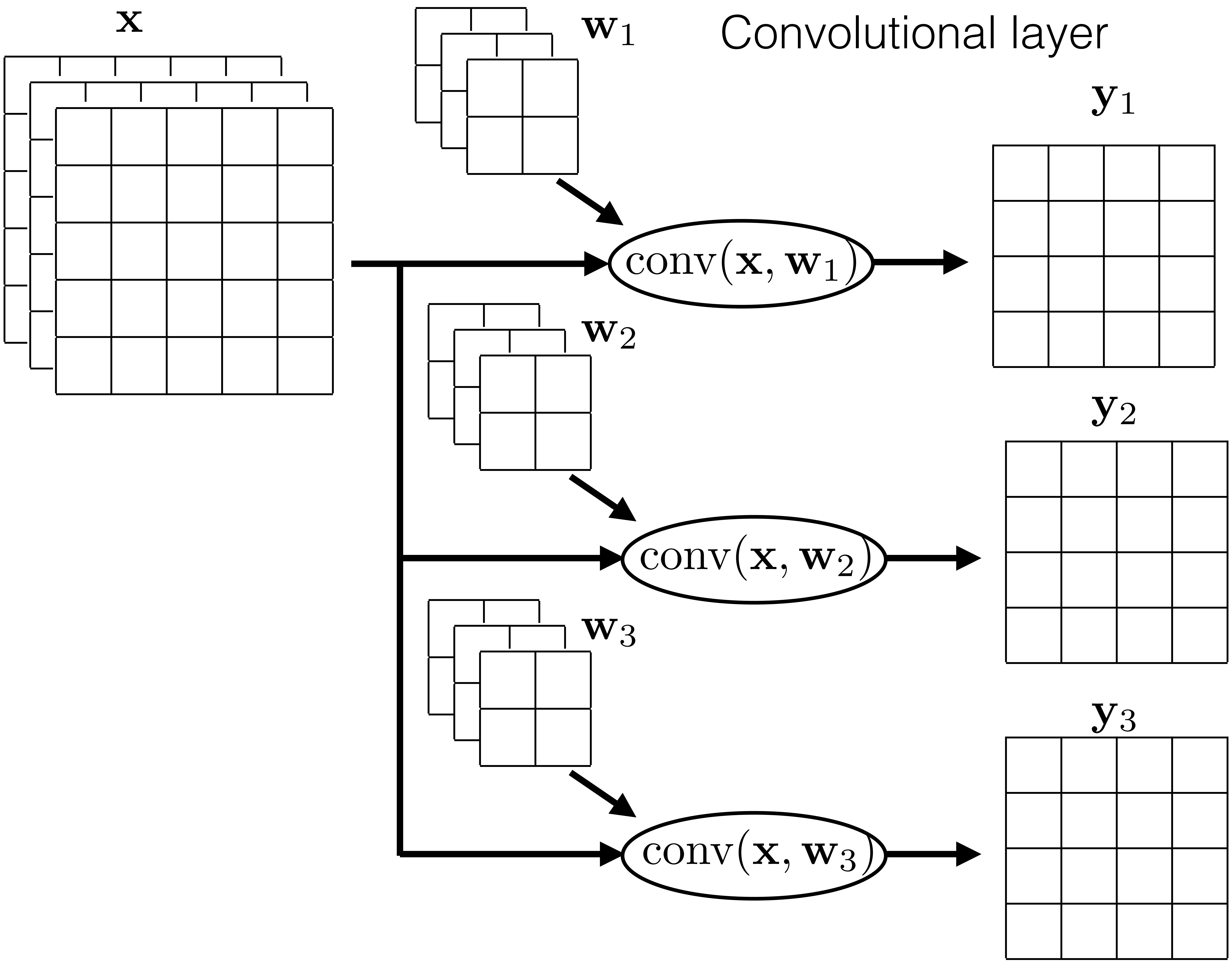
Convolutional layer



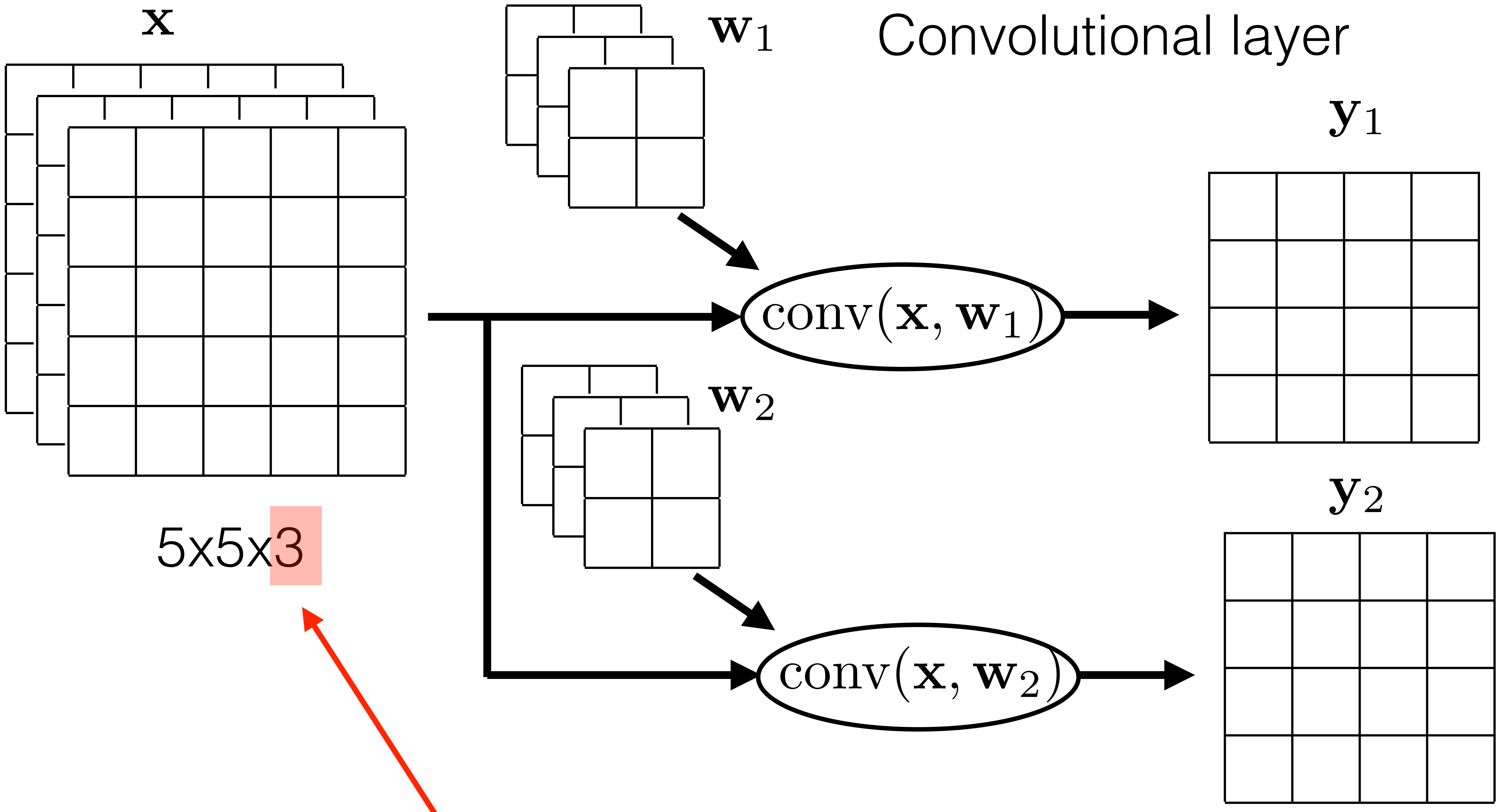
Convolutional layer



Convolutional layer

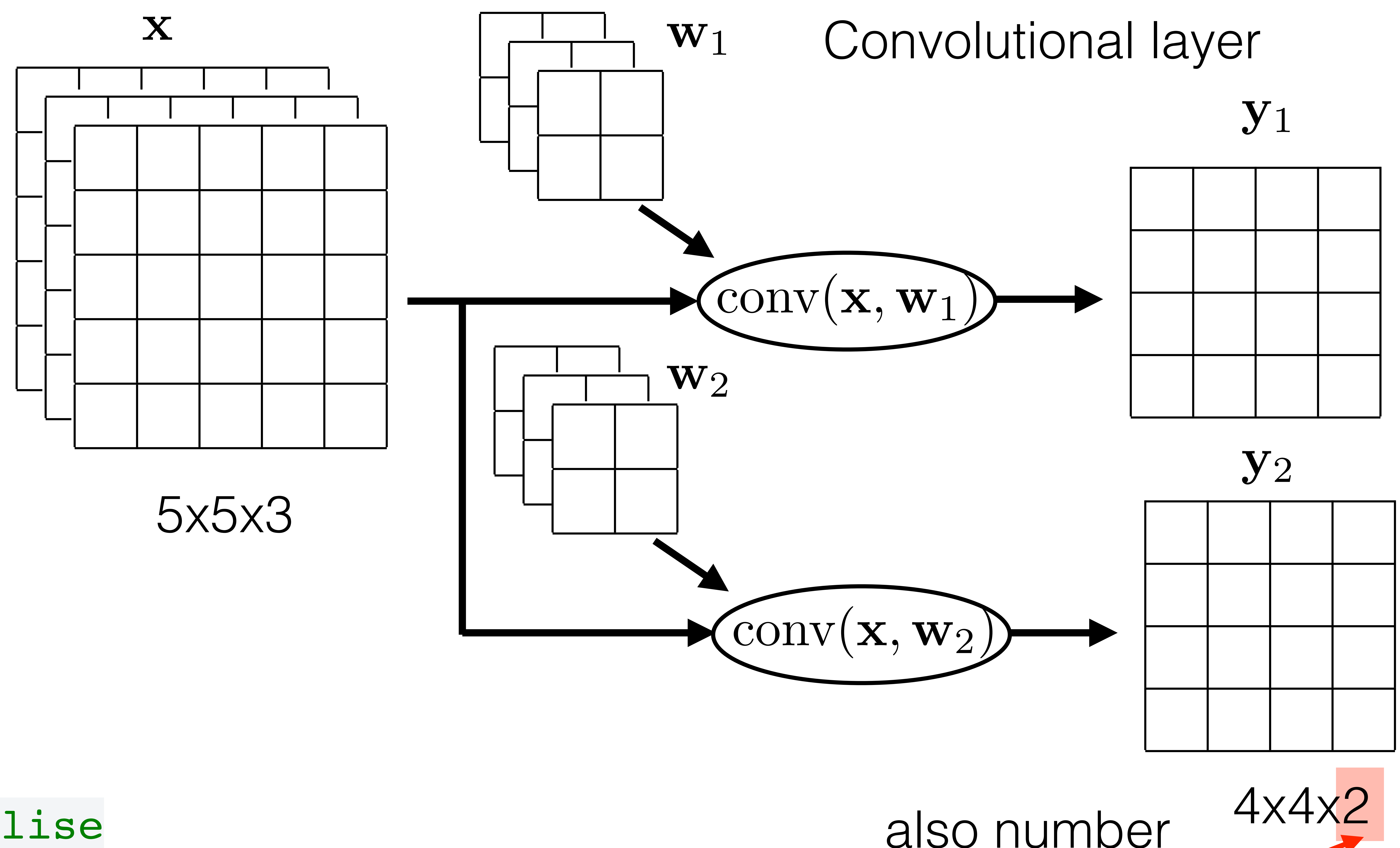


Convolutional layer



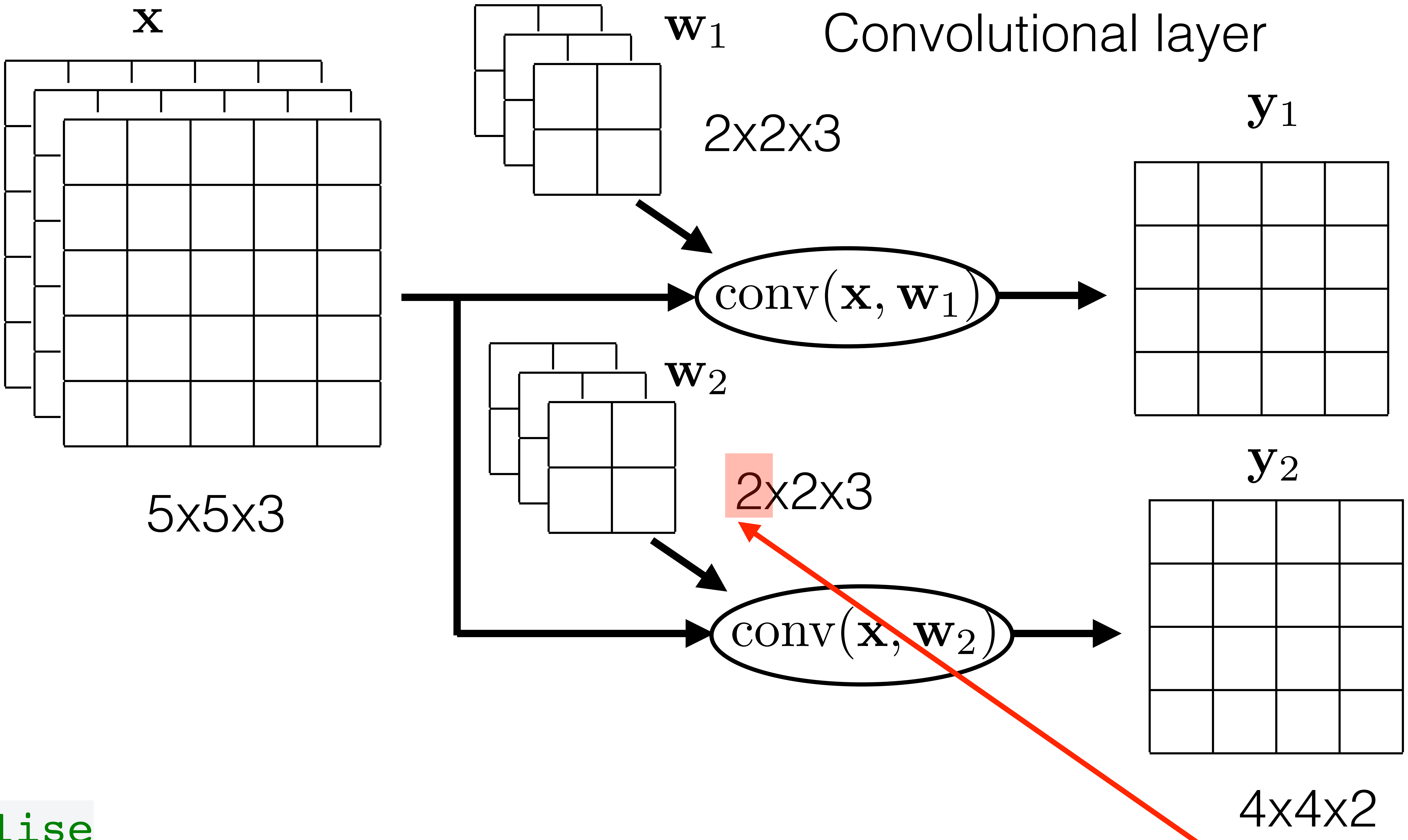

```
# initialise
import torch.nn as nn
# define 2D convolutional layer
first_layer = nn.Conv2d(in_channels=3, out_channels=2, kernel_size=2
                        stride=1, padding=1)
```

Convolutional layer




```python
# initialise
import torch.nn as nn
# define 2D convolutional layer
first_layer = nn.Conv2d(in_channels=3, out_channels=2, kernel_size=2
                        stride=1, padding=1)
```

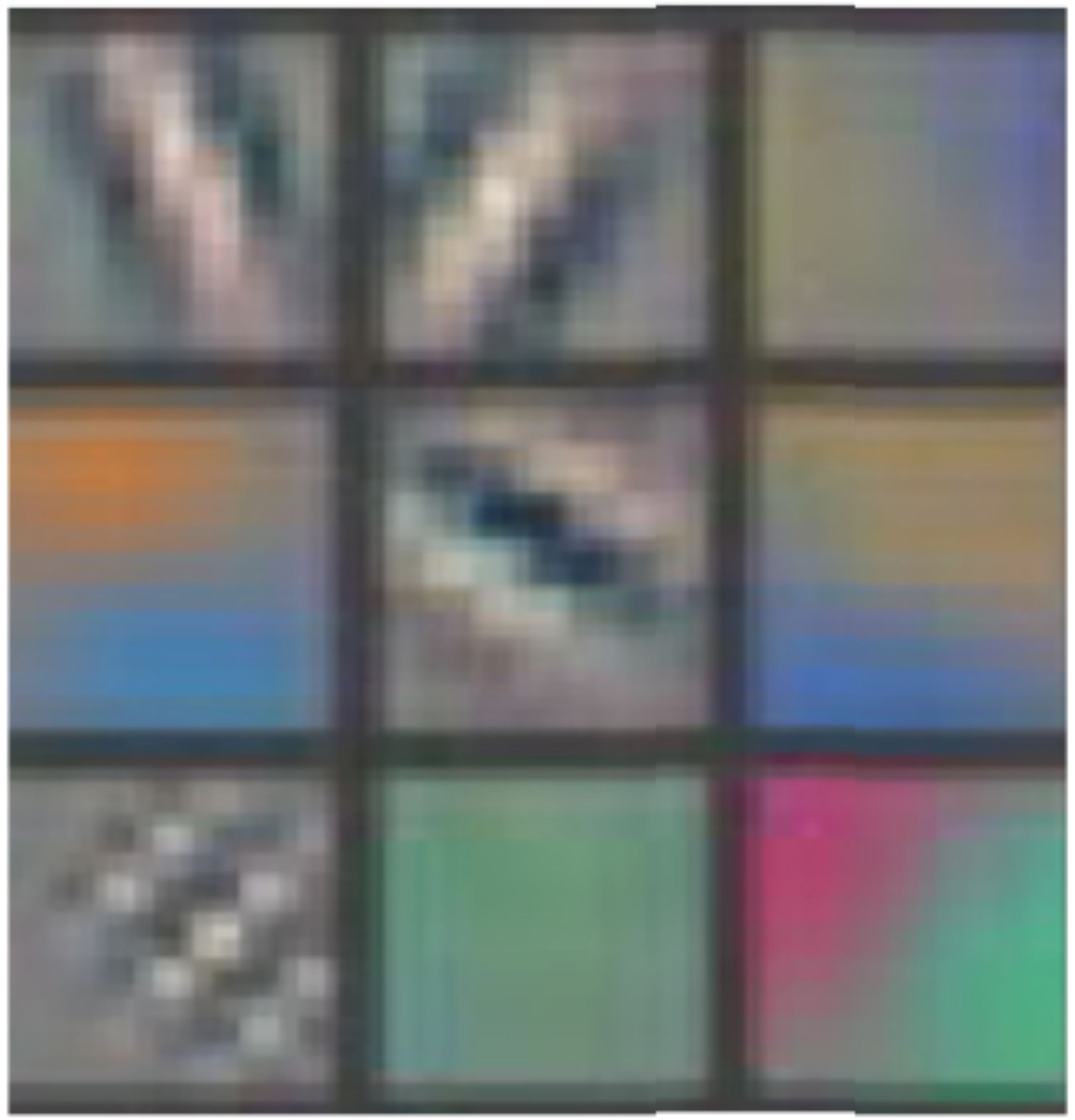
Convolutional layer




```
# initialise
import torch.nn as nn
# define 2D convolutional layer
first_layer = nn.Conv2d(in_channels=3, out_channels=2, kernel_size=2
                        stride=1, padding=1)
```

3. Neurons are sensitive to edges and its orientation

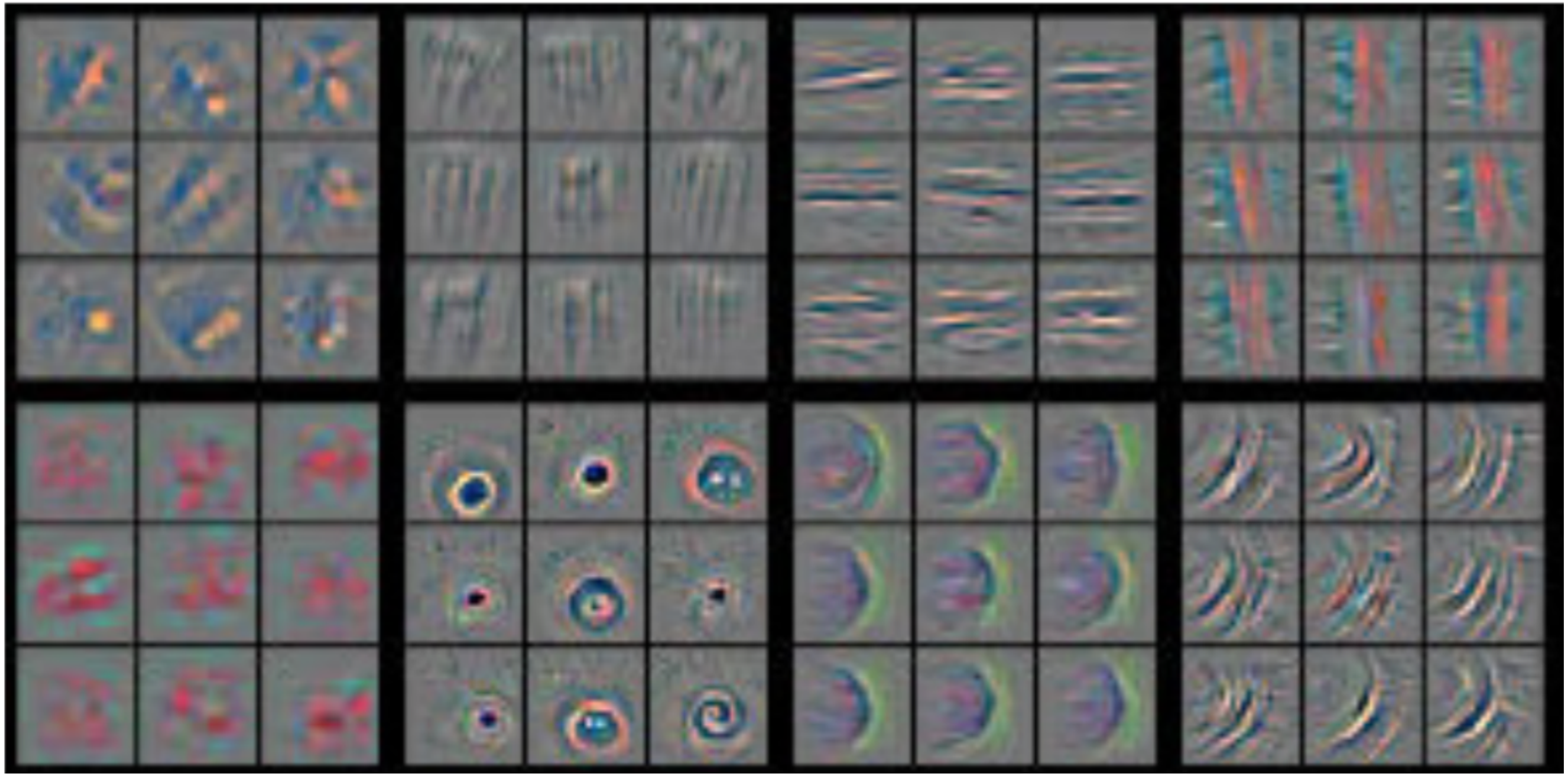
Inputs which maximized output of **layer 1**



[Zeiler and Fergus, ECCV, 2014]

3. Neurons are sensitive to edges and its orientation

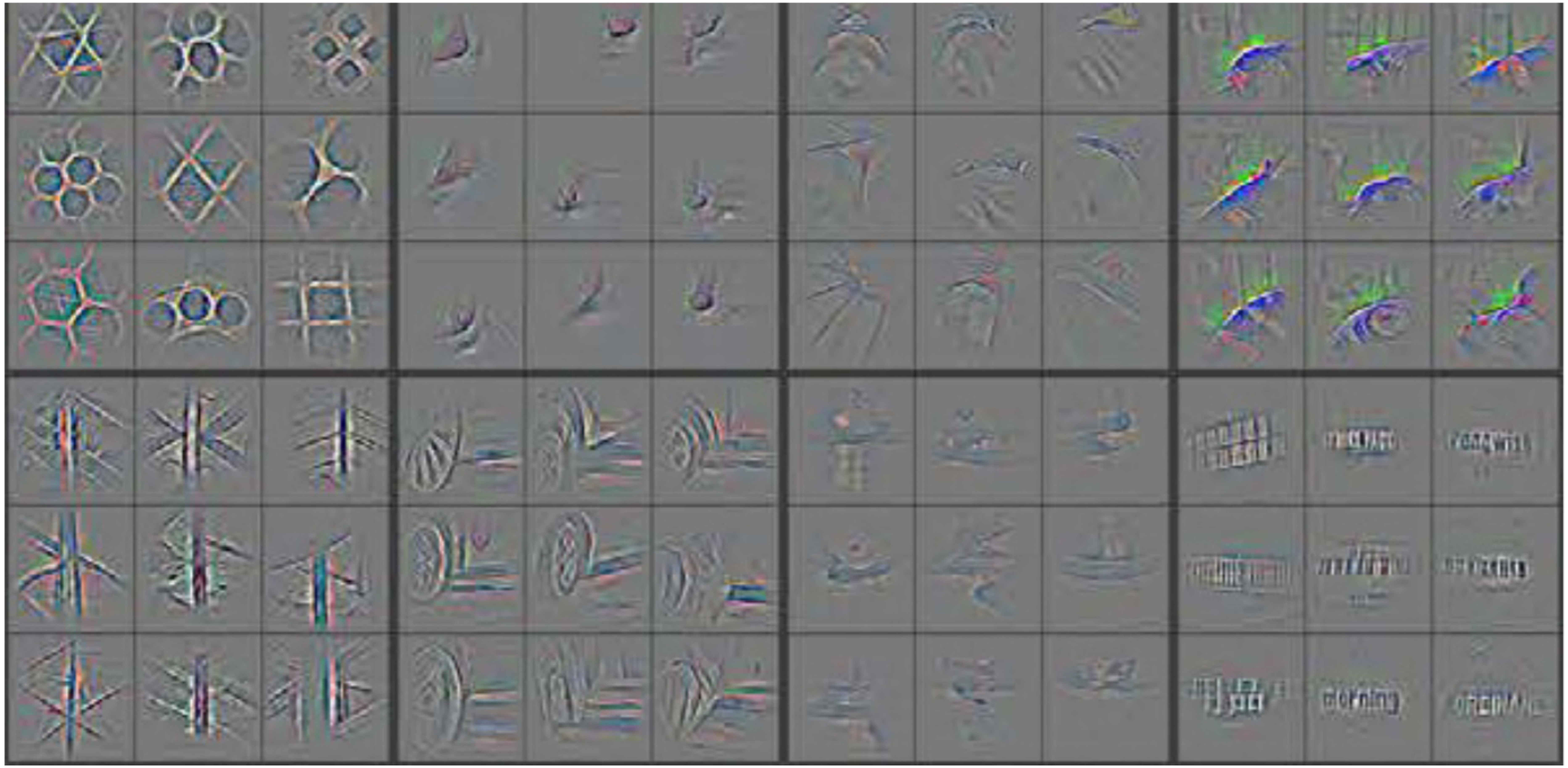
Inputs which maximized output of **layer 2**



[Zeiler and Fergus, ECCV, 2014]

3. Neurons are sensitive to edges and its orientation

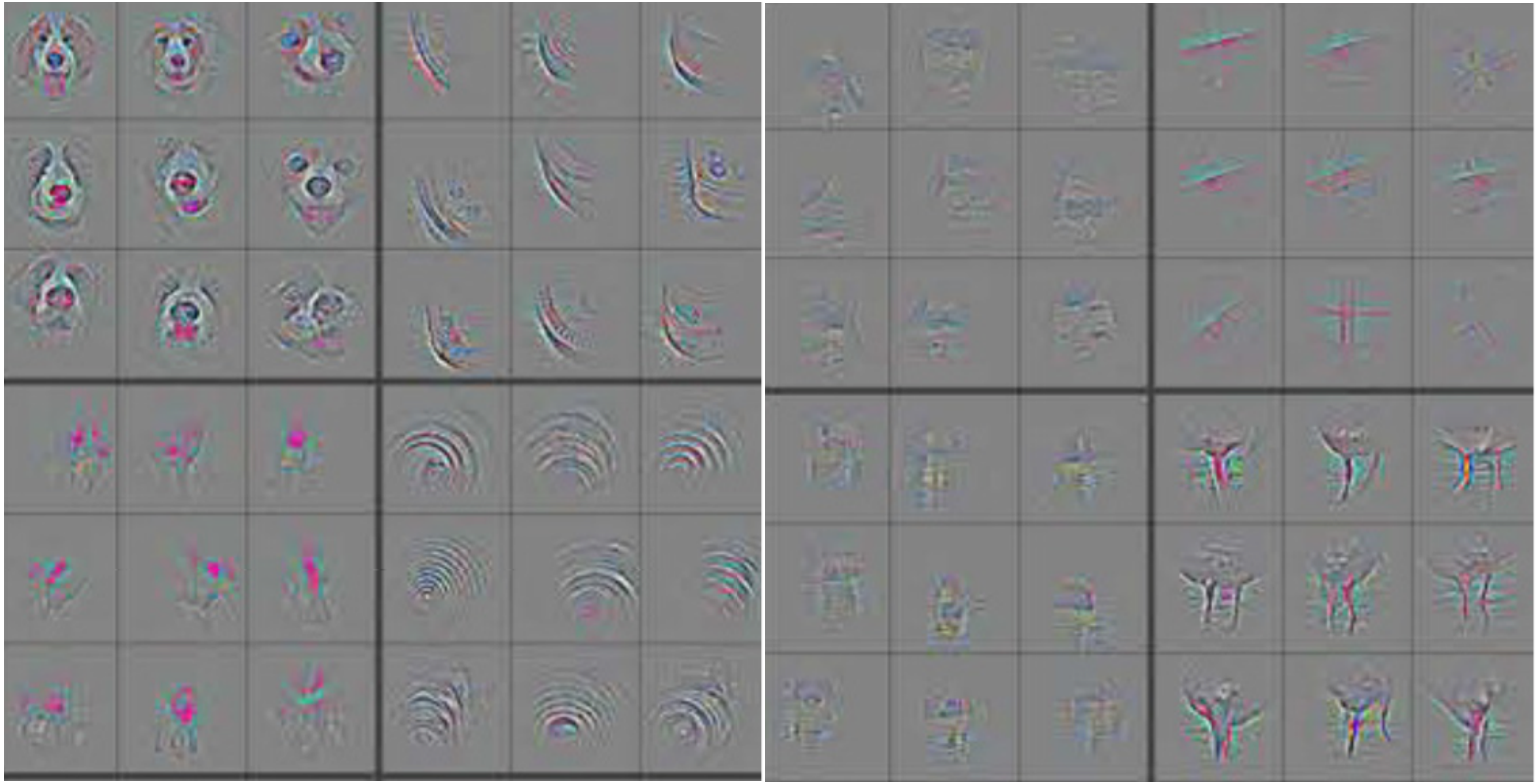
Inputs which maximized output of **layer 3**



[Zeiler and Fergus, ECCV, 2014]

3. Neurons are sensitive to edges and its orientation

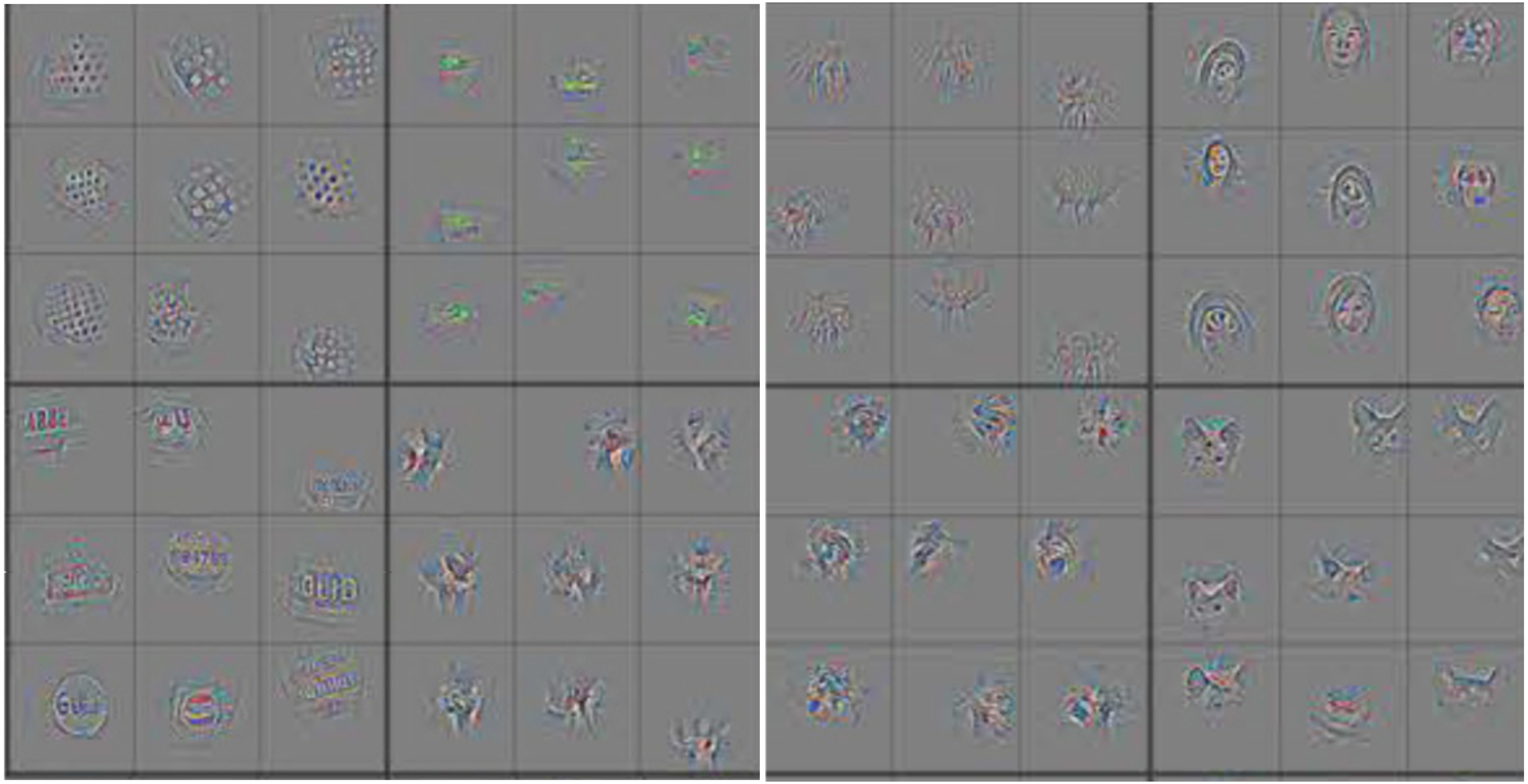
Inputs which maximized output of **layer 4**



[Zeiler and Fergus, ECCV, 2014]

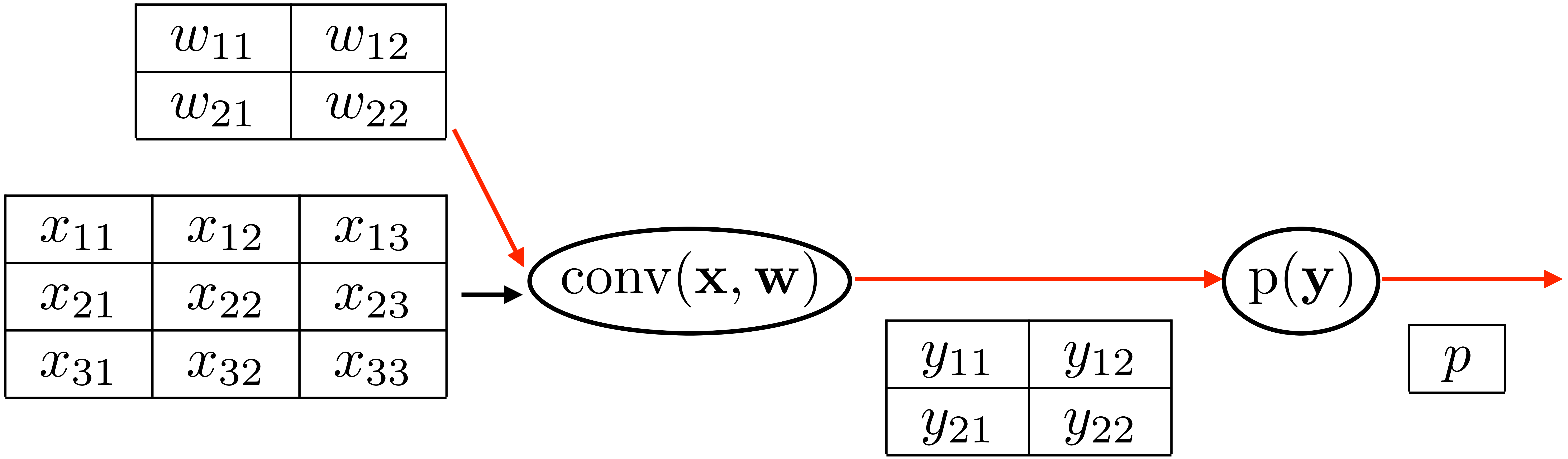
3. Neurons are sensitive to edges and its orientation

Inputs which maximized output of **layer 5**



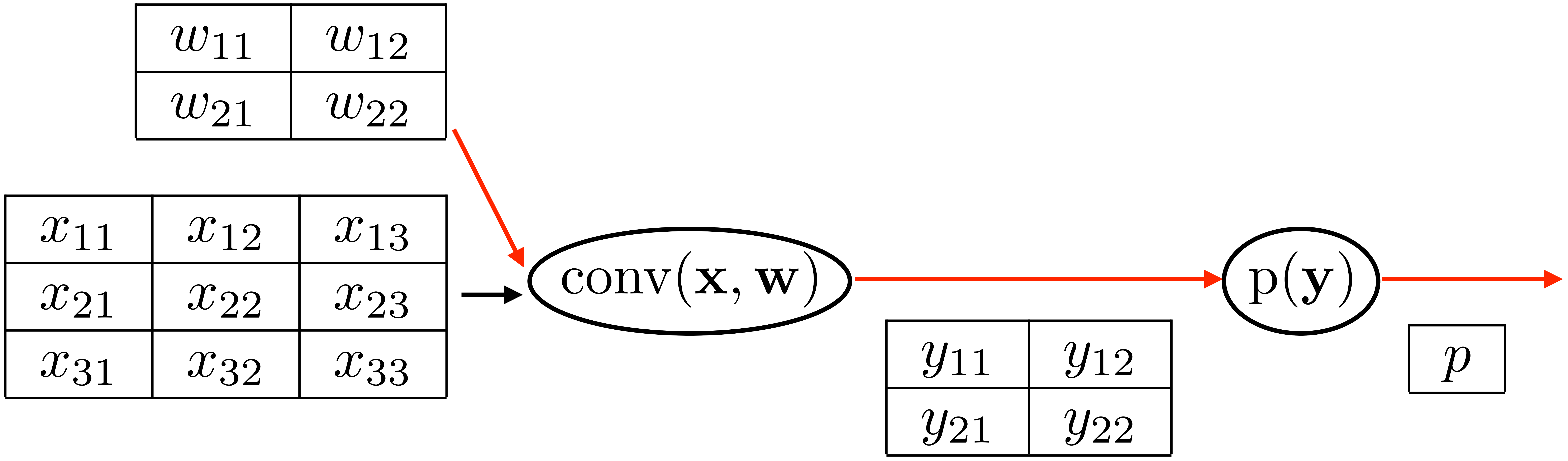
[Zeiler and Fergus, ECCV, 2014]

Convolution backward pass



Convolution backward pass

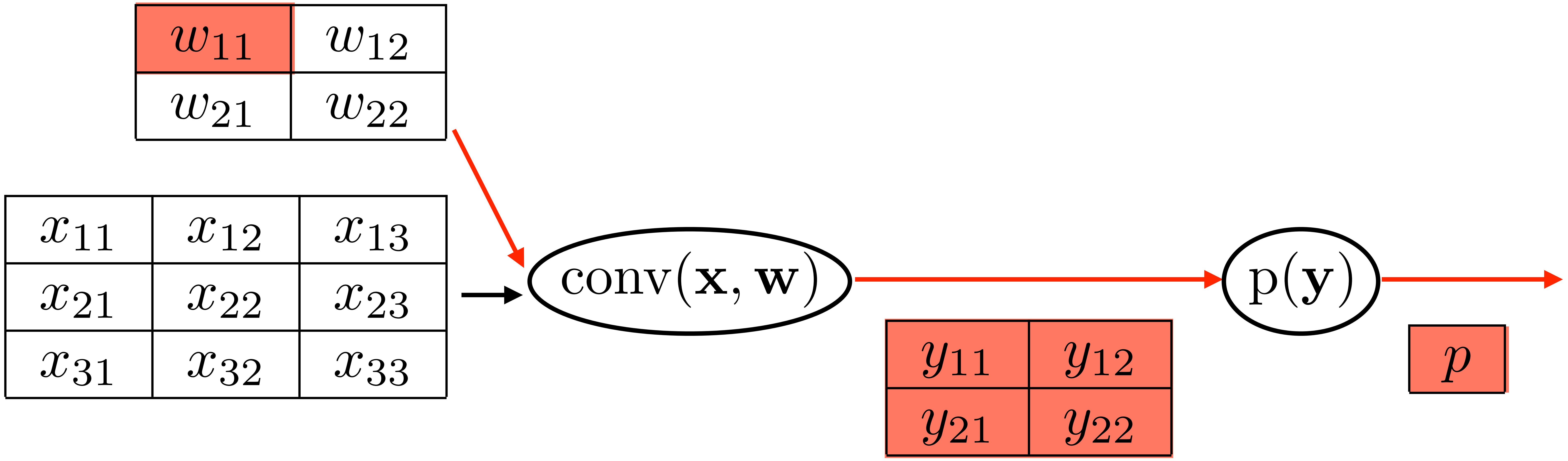
$$\begin{bmatrix} \frac{\partial p}{\partial w_{11}} & \frac{\partial p}{\partial w_{12}} \\ \frac{\partial p}{\partial w_{21}} & \frac{\partial p}{\partial w_{22}} \end{bmatrix} =?$$



Convolution backward pass

$$\begin{bmatrix} \frac{\partial p}{\partial w_{11}} & \frac{\partial p}{\partial w_{12}} \\ \frac{\partial p}{\partial w_{21}} & \frac{\partial p}{\partial w_{22}} \end{bmatrix} = ?$$

$$p(w_{11}) = \mathrm{p}(y_{11}(w_{11}),\ y_{12}(w_{11}),\ y_{21}(w_{11}),\ y_{22}(w_{11}))$$

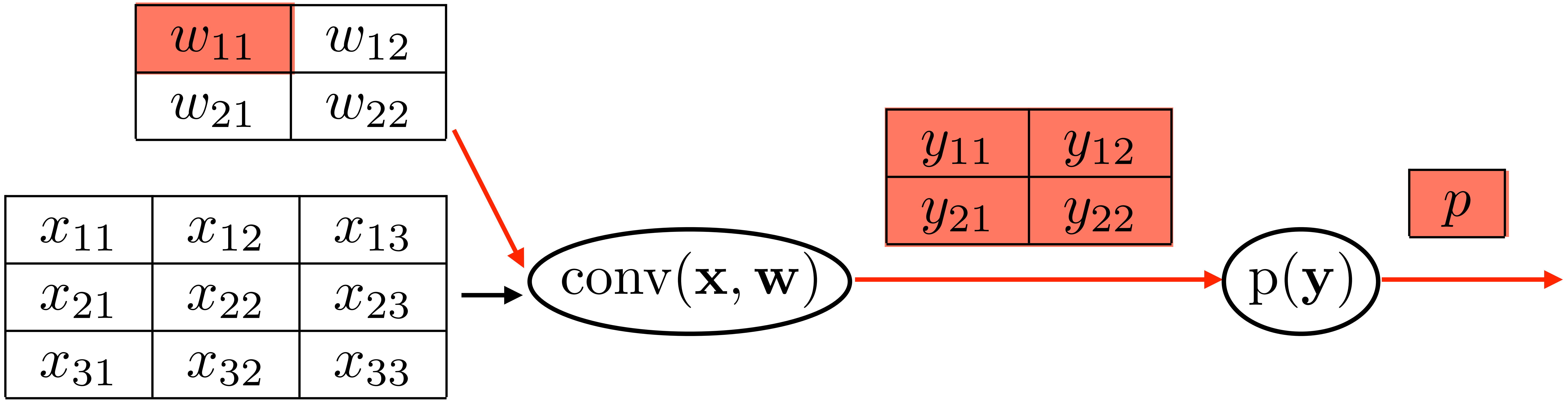


Convolution backward pass

$$\begin{bmatrix} \frac{\partial p}{\partial w_{11}} & \frac{\partial p}{\partial w_{12}} \\ \frac{\partial p}{\partial w_{21}} & \frac{\partial p}{\partial w_{22}} \end{bmatrix} = ?$$

$$p(w_{11}) = \mathrm{p}(y_{11}(w_{11}),\ y_{12}(w_{11}),\ y_{21}(w_{11}),\ y_{22}(w_{11}))$$

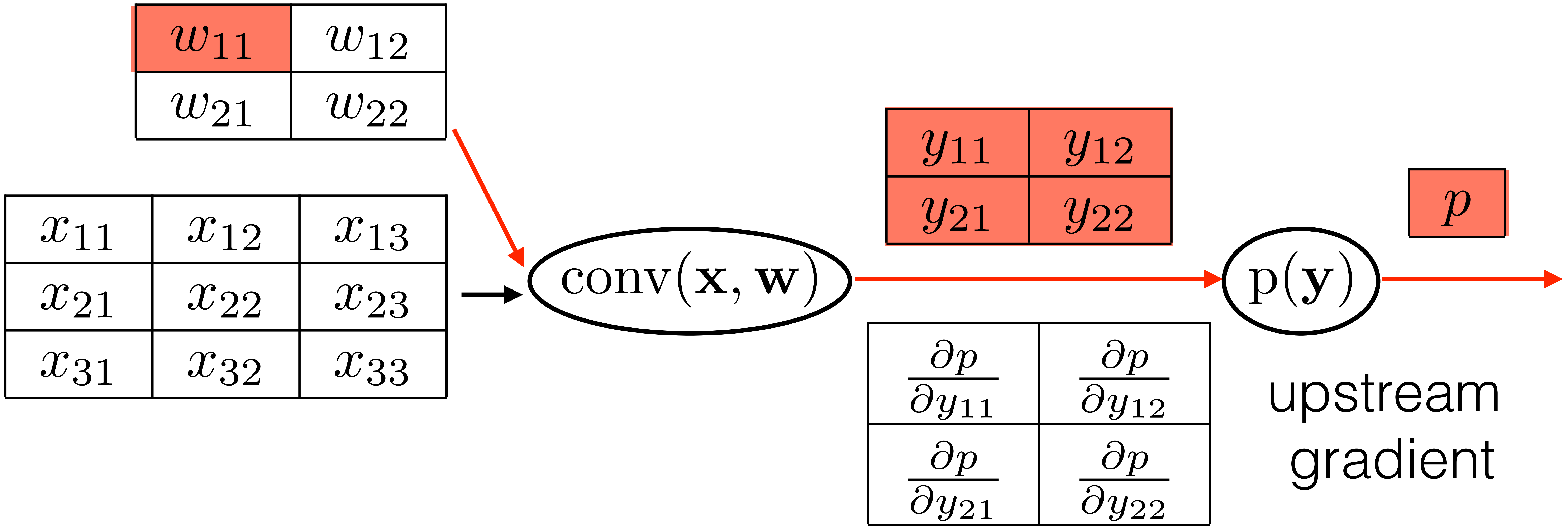
- How does the feed-forward pass work in this setting?
- How is the value of p influenced by w_{11}?



Convolution backward pass

$$\frac{\partial p}{\partial w_{11}} = \ ?$$

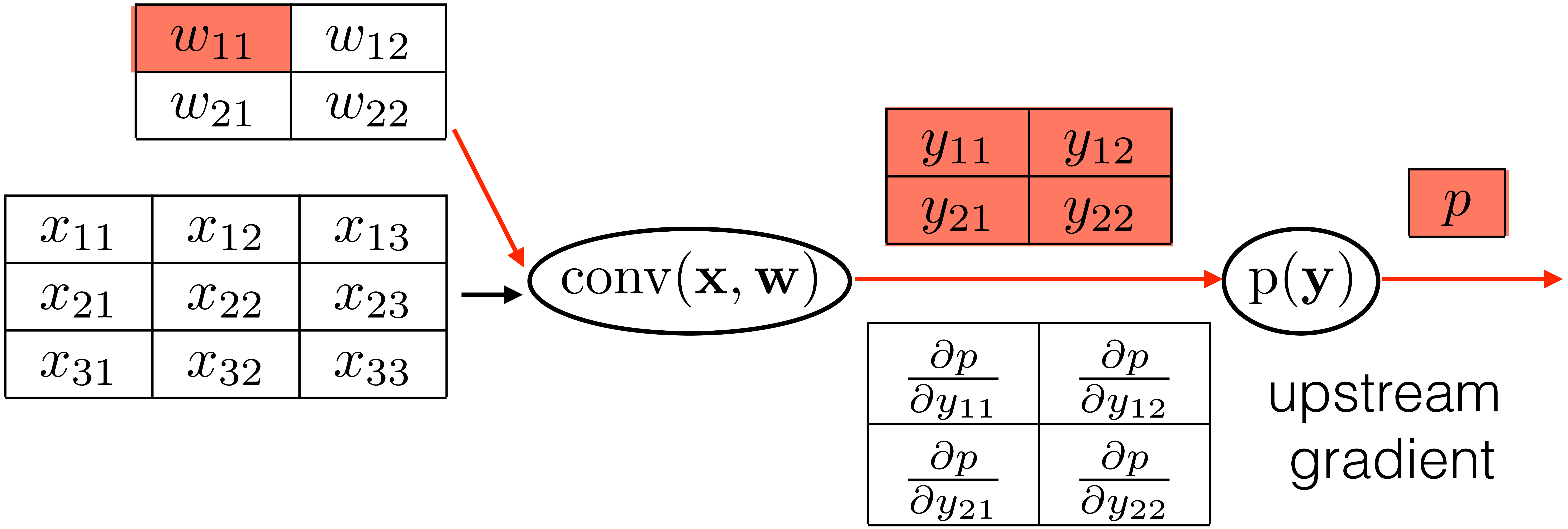
$$p(w_{11}) = \mathrm{p}(y_{11}(w_{11}),\ y_{12}(w_{11}),\ y_{21}(w_{11}),\ y_{22}(w_{11}))$$



Convolution backward pass

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

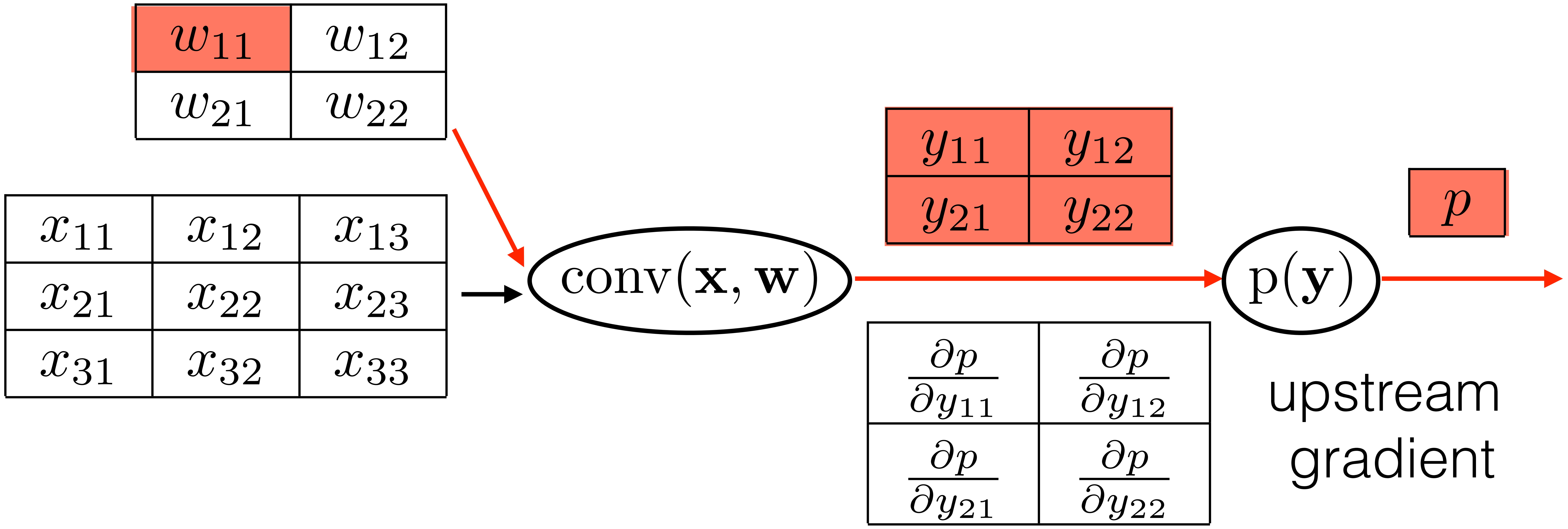
$$p(w_{11}) = \mathrm{p}(y_{11}(w_{11}),\ y_{12}(w_{11}),\ y_{21}(w_{11}),\ y_{22}(w_{11}))$$



Convolution backward pass

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

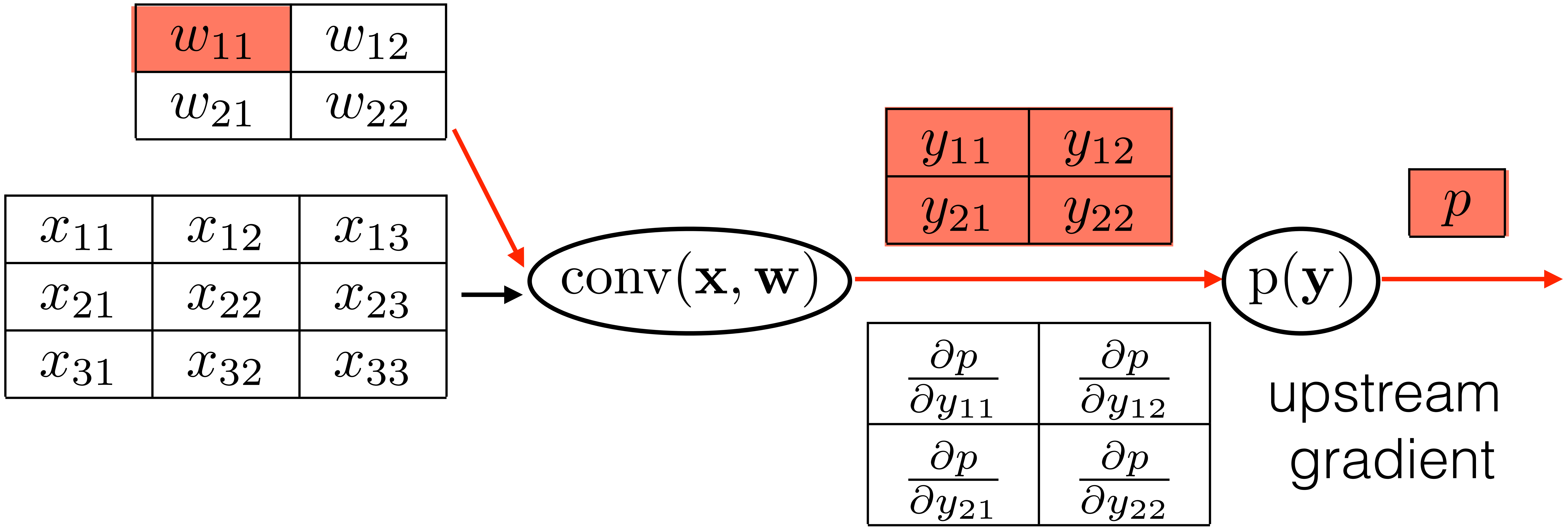
$$\frac{\partial y_{11}}{\partial w_{11}} = \text{?}$$



Convolution backward pass

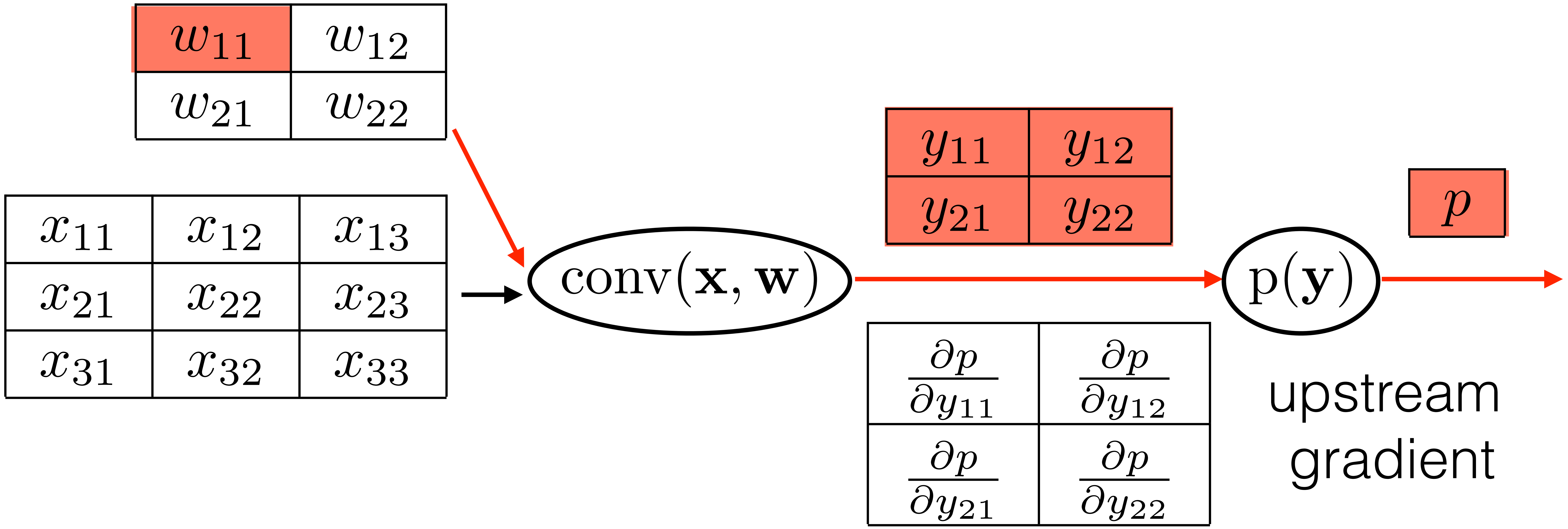
$$\frac{\partial p}{\partial w_{11}} = \frac{\partial p}{\partial y_{11}}\frac{\partial y_{11}}{\partial w_{11}} + \frac{\partial p}{\partial y_{12}}\frac{\partial y_{12}}{\partial w_{11}} + \frac{\partial p}{\partial y_{21}}\frac{\partial y_{21}}{\partial w_{11}} + \frac{\partial p}{\partial y_{22}}\frac{\partial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

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



Convolution backward pass

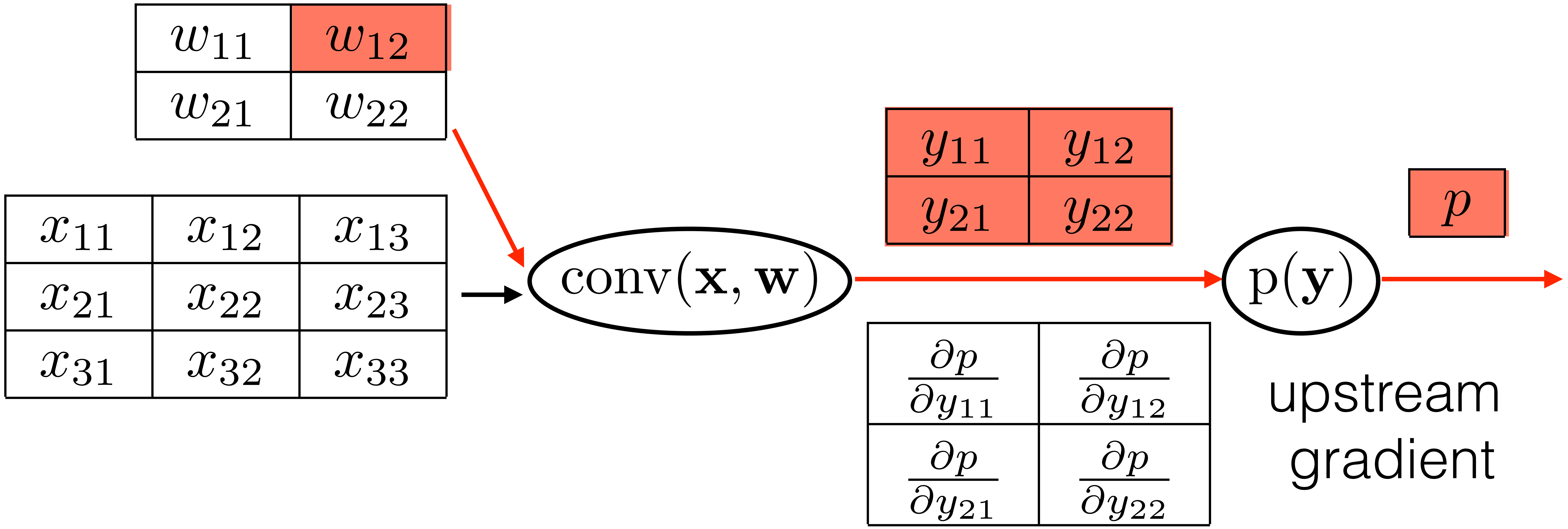
$$\frac{\partial p}{\partial w_{11}} = \frac{\partial p}{\partial y_{11}} x_{11} + \frac{\partial p}{\partial y_{12}} x_{12} + \frac{\partial p}{\partial y_{21}} x_{21} + \frac{\partial p}{\partial y_{22}} x_{22}$$



Convolution backward pass

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

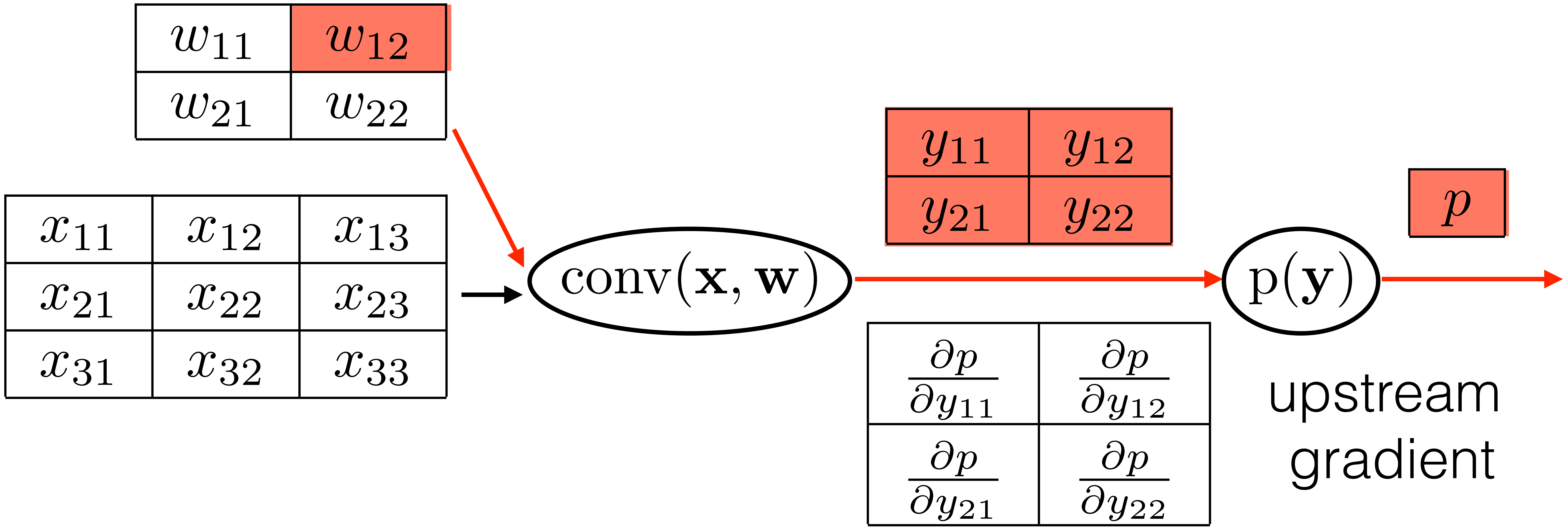
$$\frac{\partial p}{\partial w_{12}} = \text{?}$$



Convolution backward pass

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

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

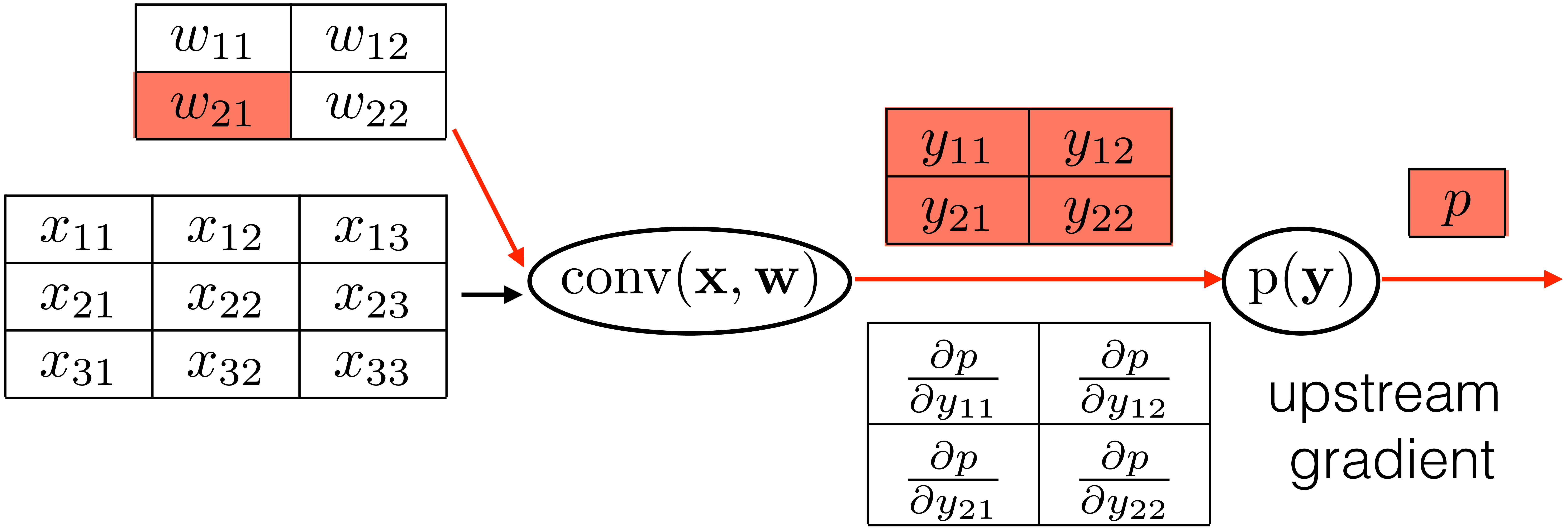


Convolution backward pass

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

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

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



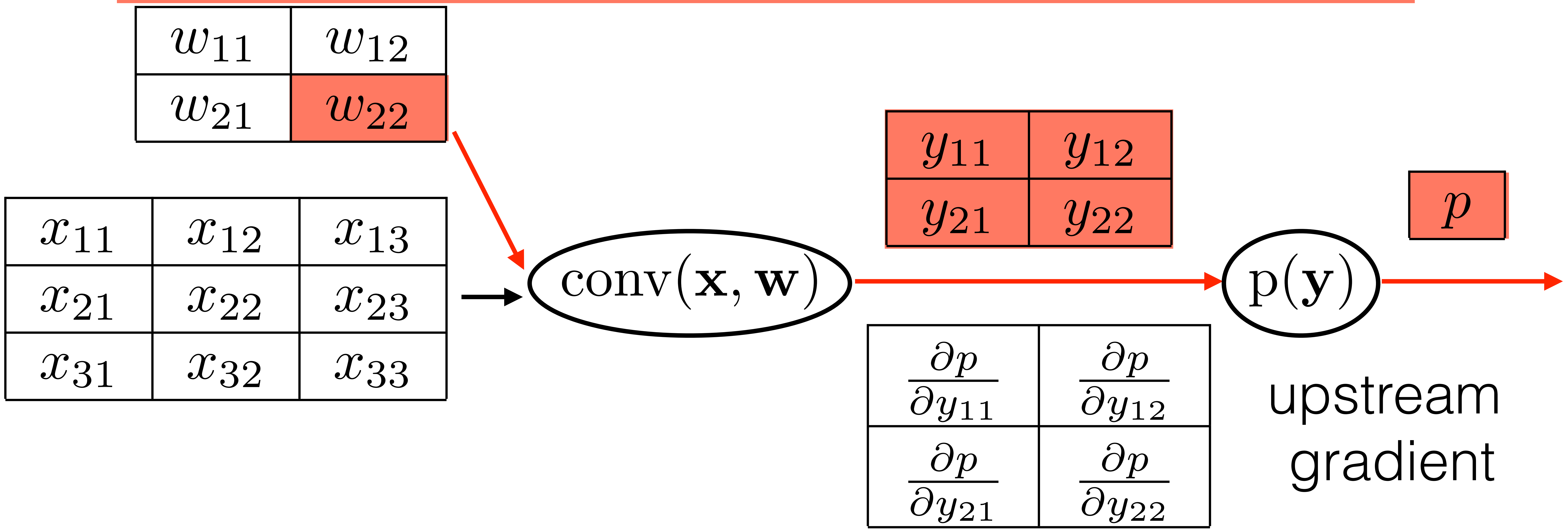
Convolution backward pass

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

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

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

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



upstream gradient

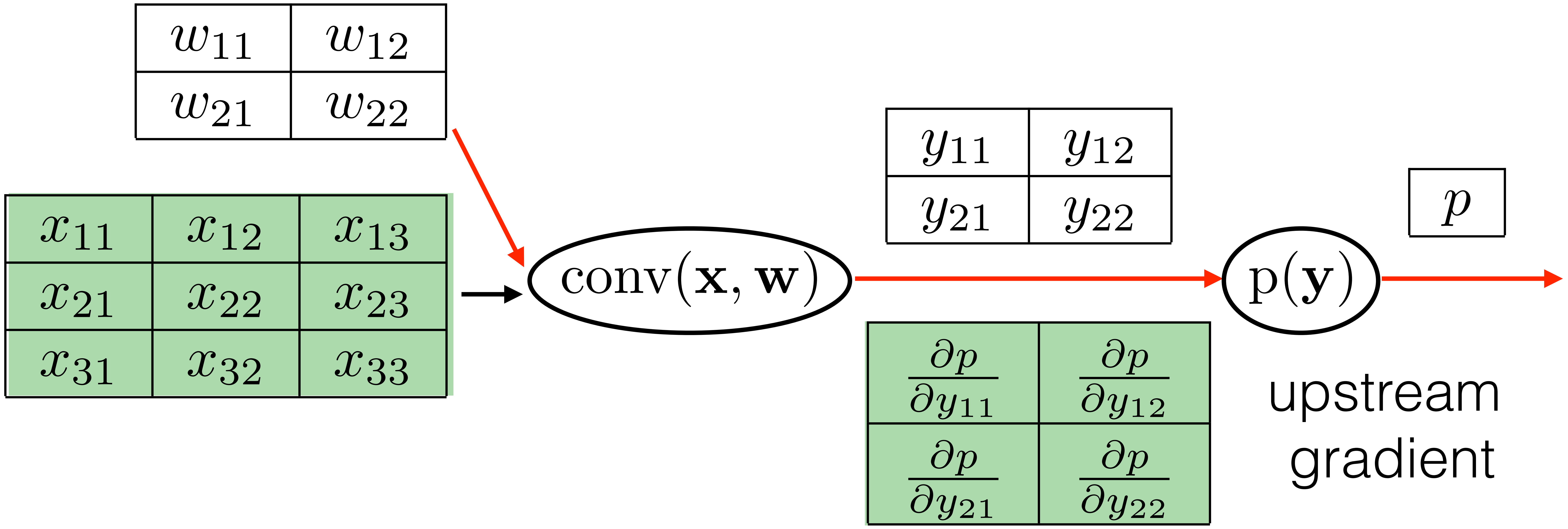
Convolution backward pass

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

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

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

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



Convolution backward pass

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

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

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

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

$$\begin{array}{|c|c|}\hline \frac{\partial p}{\partial w_{11}} & \frac{\partial p}{\partial w_{12}} \\ \hline \frac{\partial p}{\partial w_{21}} & \frac{\partial p}{\partial w_{22}} \\ \hline \end{array} = \mathrm{conv}\left(\begin{array}{|c|c|c|}\hline x_{11} & x_{12} & x_{13} \\ \hline x_{21} & x_{22} & x_{23} \\ \hline x_{31} & x_{32} & x_{33} \\ \hline \end{array}, \begin{array}{|c|c|}\hline \frac{\partial p}{\partial y_{11}} & \frac{\partial p}{\partial y_{12}} \\ \hline \frac{\partial p}{\partial y_{21}} & \frac{\partial p}{\partial y_{22}} \\ \hline \end{array} \right)$$

Convolution backward pass

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

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

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

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

$$\begin{array}{|c|c|}\hline \frac{\partial p}{\partial w_{11}} & \frac{\partial p}{\partial w_{12}} \\ \hline \frac{\partial p}{\partial w_{21}} & \frac{\partial p}{\partial w_{22}} \\ \hline \end{array} = \mathrm{conv}\left(\begin{array}{|c|c|c|}\hline x_{11} & x_{12} & x_{13} \\ \hline x_{21} & x_{22} & x_{23} \\ \hline x_{31} & x_{32} & x_{33} \\ \hline \end{array}, \begin{array}{|c|c|}\hline \frac{\partial p}{\partial y_{11}} & \frac{\partial p}{\partial y_{12}} \\ \hline \frac{\partial p}{\partial y_{21}} & \frac{\partial p}{\partial y_{22}} \\ \hline \end{array} \right)$$

Convolution backward pass

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

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

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

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

$$\begin{array}{|c|c|}\hline \frac{\partial p}{\partial w_{11}} & \frac{\partial p}{\partial w_{12}} \\ \hline \frac{\partial p}{\partial w_{21}} & \frac{\partial p}{\partial w_{22}} \\ \hline \end{array} = \text{conv}\left(\begin{array}{|c|c|c|}\hline x_{11} & x_{12} & x_{13} \\ \hline x_{21} & x_{22} & x_{23} \\ \hline x_{31} & x_{32} & x_{33} \\ \hline \end{array}, \begin{array}{|c|c|}\hline \frac{\partial p}{\partial y_{11}} & \frac{\partial p}{\partial y_{12}} \\ \hline \frac{\partial p}{\partial y_{21}} & \frac{\partial p}{\partial y_{22}} \\ \hline \end{array} \right)$$

Convolution backward pass

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

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

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

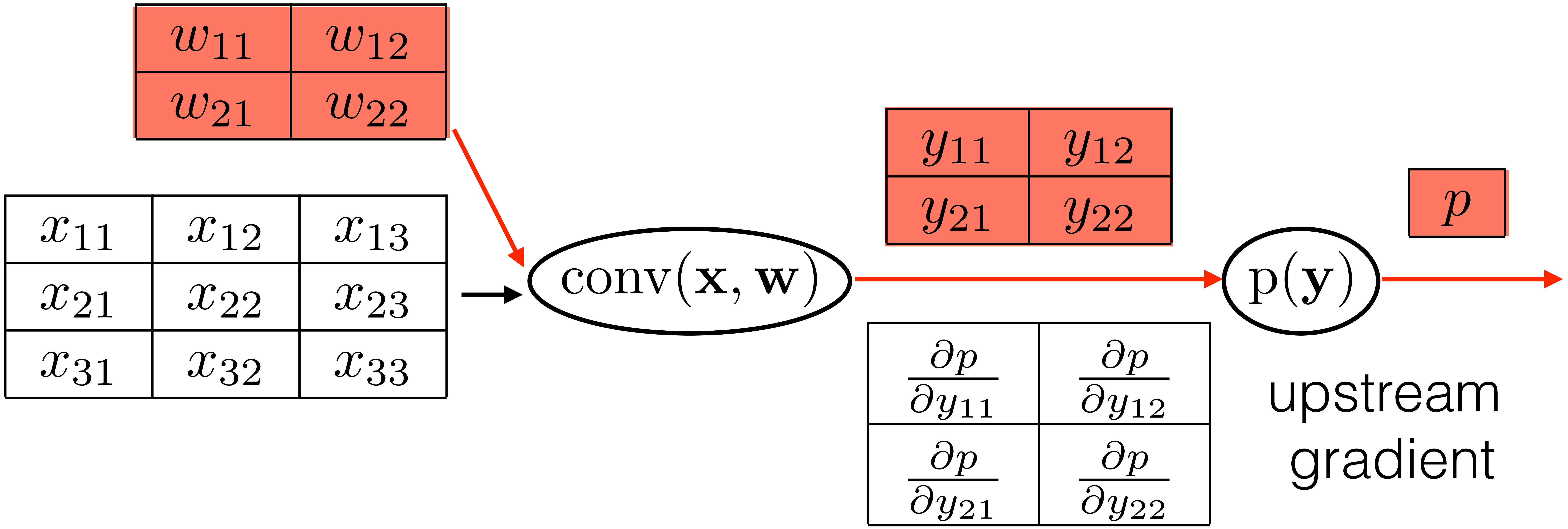
$$\frac{\partial p}{\partial w_{22}} = \frac{\partial p}{\partial y_{11}} x_{22} + \frac{\partial p}{\partial y_{12}} x_{23} + \frac{\partial p}{\partial y_{21}} x_{32} + \frac{\partial p}{\partial y_{22}} x_{33}$$

$$\begin{array}{|c|c|} \hline \frac{\partial p}{\partial w_{11}} & \frac{\partial p}{\partial w_{12}} \\ \hline \frac{\partial p}{\partial w_{21}} & \frac{\partial p}{\partial w_{22}} \\ \hline \end{array} = \text{conv}\left(\begin{array}{|c|c|c|} \hline x_{11} & x_{12} & x_{13} \\ \hline x_{21} & x_{22} & x_{23} \\ \hline x_{31} & x_{32} & x_{33} \\ \hline \end{array}, \begin{array}{|c|c|} \hline \frac{\partial p}{\partial y_{11}} & \frac{\partial p}{\partial y_{12}} \\ \hline \frac{\partial p}{\partial y_{21}} & \frac{\partial p}{\partial y_{22}} \\ \hline \end{array} \right)$$

Convolution backward pass wrt weights

- Backpropagation in convolutional layer wrt weights is:
  "**convolution of input feature map with upstream gradient**"

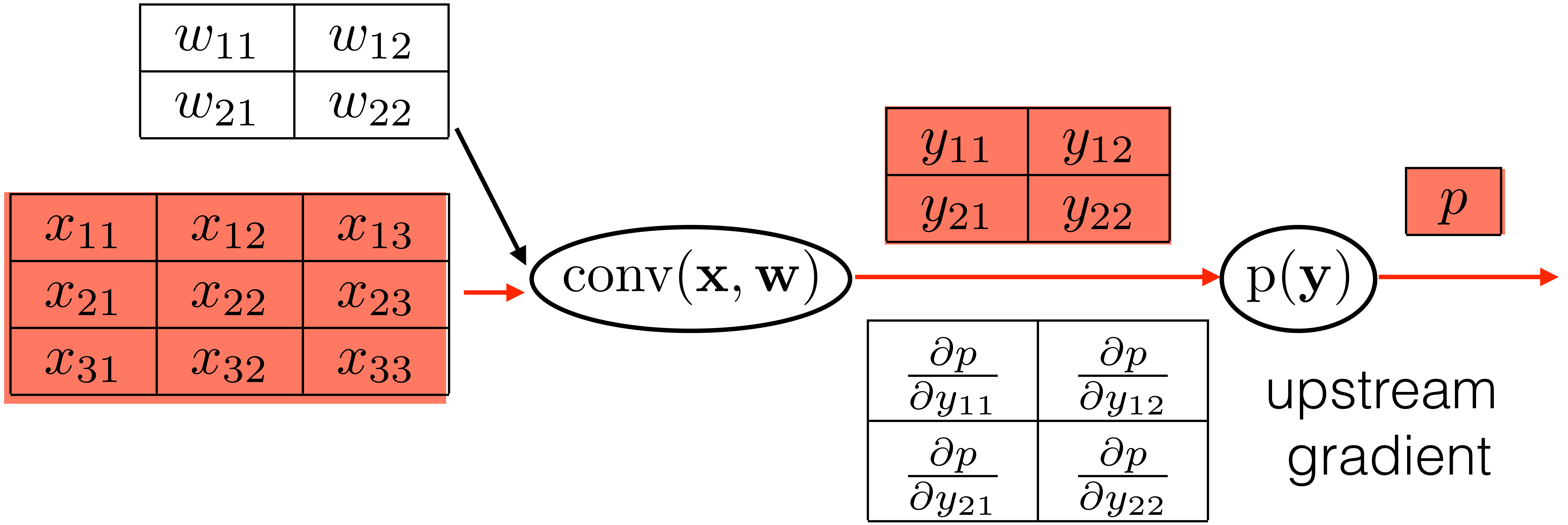
$$\begin{bmatrix} \frac{\partial p}{\partial w_{11}} & \frac{\partial p}{\partial w_{12}} \\ \frac{\partial p}{\partial w_{21}} & \frac{\partial p}{\partial w_{22}} \end{bmatrix} = \mathrm{conv}\left(\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}, \begin{bmatrix} \frac{\partial p}{\partial y_{11}} & \frac{\partial p}{\partial y_{12}} \\ \frac{\partial p}{\partial y_{21}} & \frac{\partial p}{\partial y_{22}} \end{bmatrix} \right)$$



Convolution backward pass wrt input feature map

- Backpropagation in convolutional layer is:
  "**convolution of padded upstream gradient with mirrored weights**"

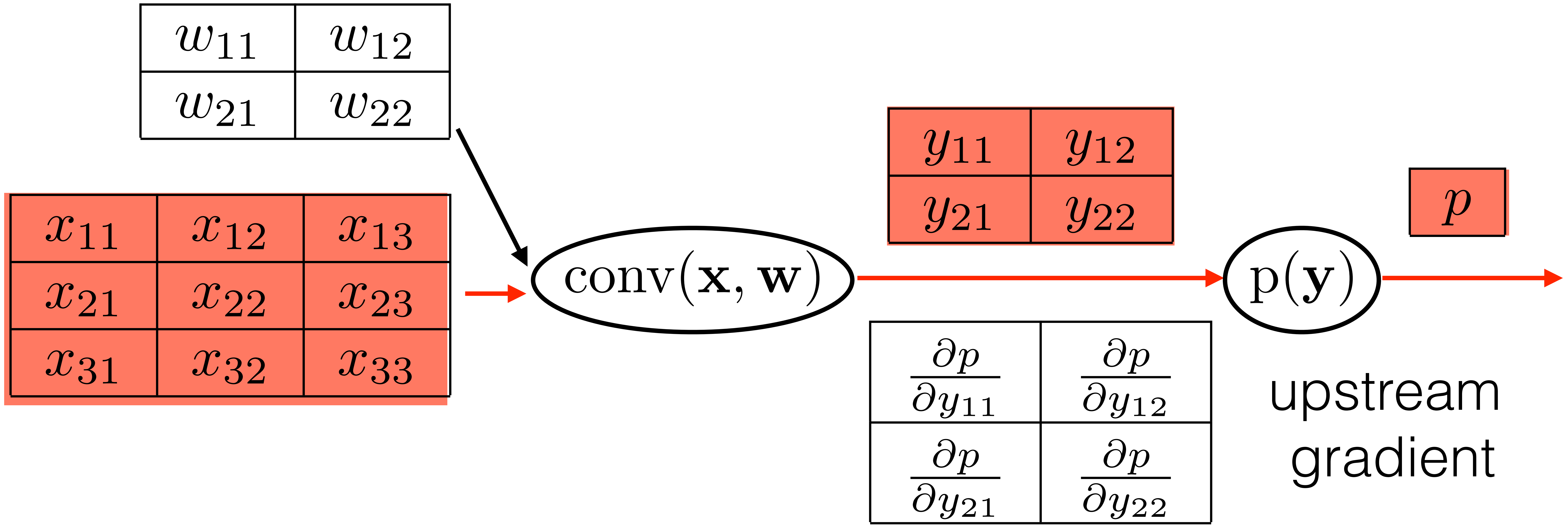
$$\begin{bmatrix} \frac{\partial p}{\partial x_{11}} & \frac{\partial p}{\partial x_{12}} & \frac{\partial p}{\partial x_{13}} \\ \frac{\partial p}{\partial x_{21}} & \frac{\partial p}{\partial x_{22}} & \frac{\partial p}{\partial x_{23}} \\ \frac{\partial p}{\partial x_{31}} & \frac{\partial p}{\partial x_{32}} & \frac{\partial p}{\partial x_{33}} \end{bmatrix} = \mathrm{conv}\left(\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial p}{\partial y_{11}} & \frac{\partial p}{\partial y_{12}} & 0 \\ 0 & \frac{\partial p}{\partial y_{21}} & \frac{\partial p}{\partial y_{22}} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} \right)$$



Convolution backward pass wrt input feature map

Very important property of convolutional layer is:

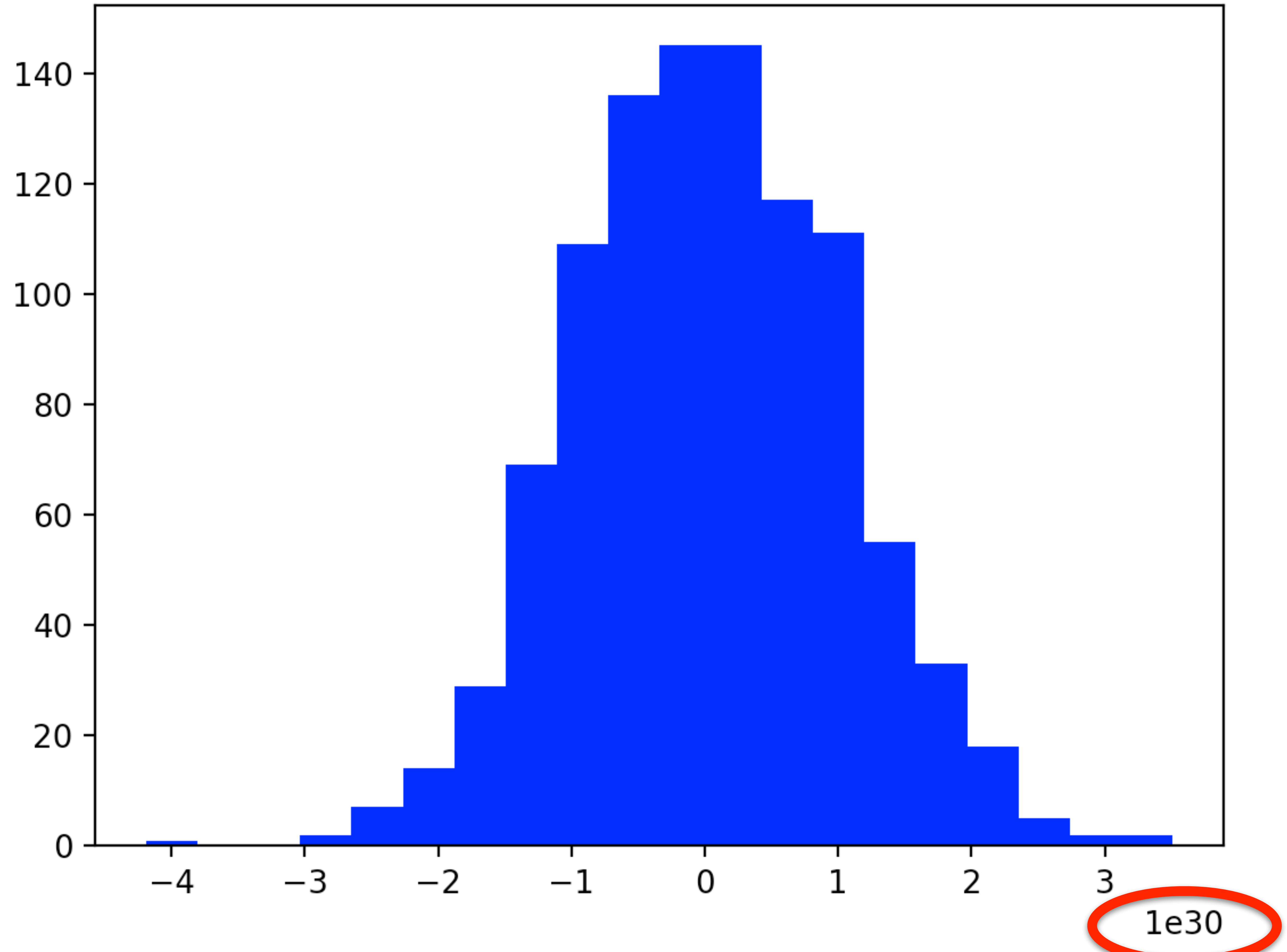
Local gradient is also convolution !!!



Learning

What happens to deep **conv outputs** when weights are **huge**?

```
y = torch.randn(1000,1)
for i in range(20):
    weights = torch.randn(1000,1000)
    y =  weights @ y
```



=>Gradient clipping

Value-based

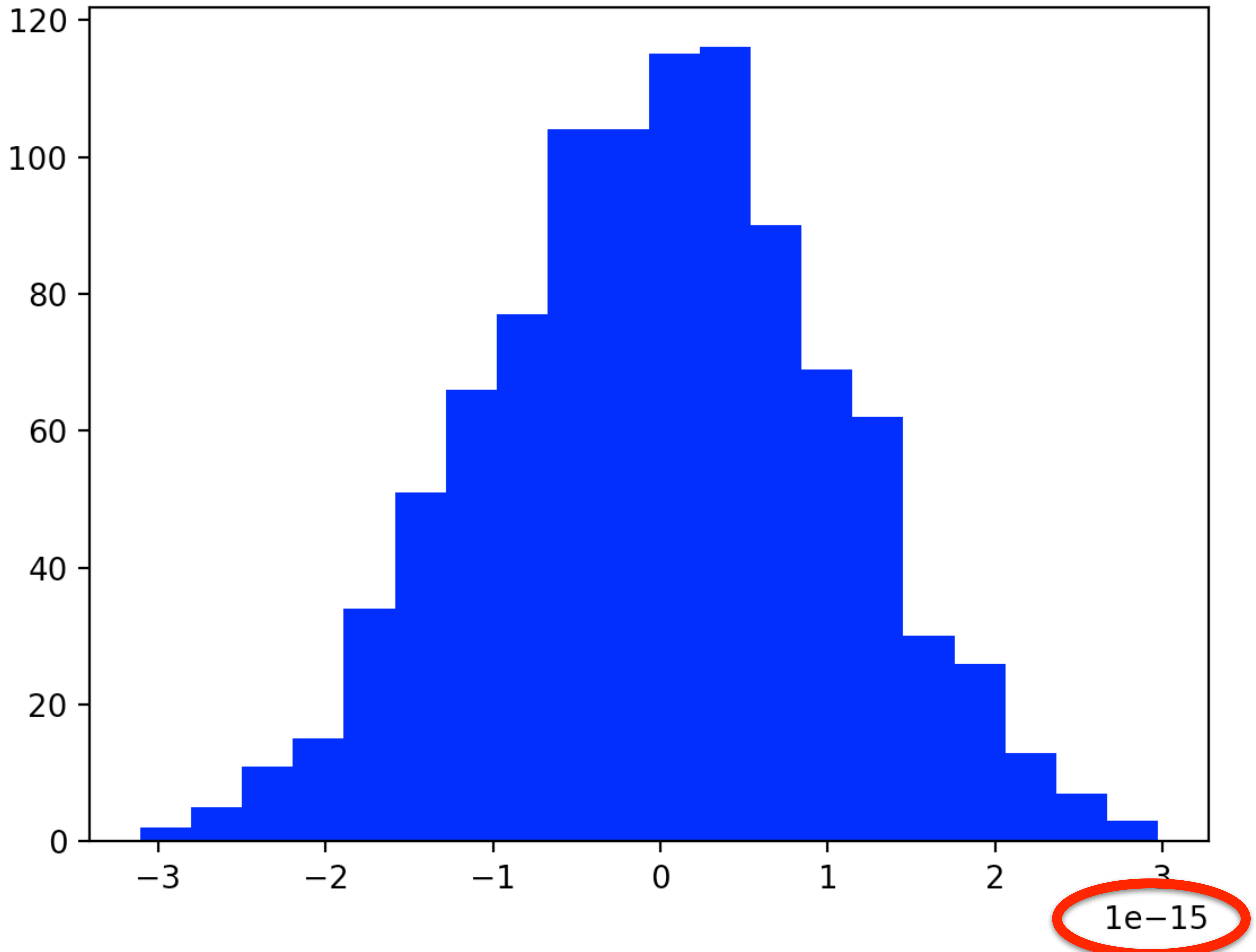
vs

Norm-based

Learning

What happens to deep **conv outputs** when weights are **small**?

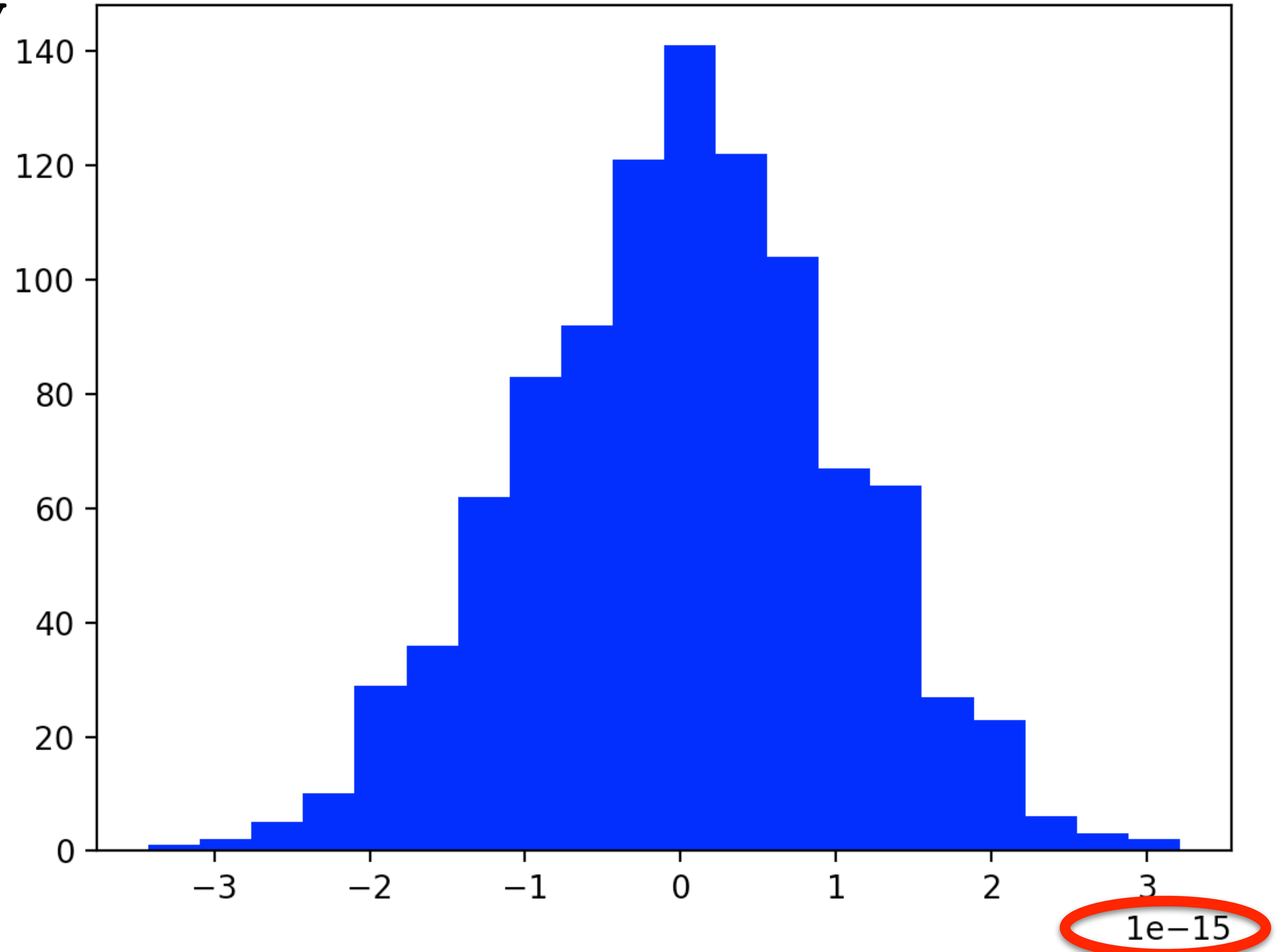
```
y = torch.randn(1000,1)
for i in range(30):
    weights = torch.randn(1000,1000)/100
    y =  weights @ y
```



Learning

What happens to deep **conv gradient** when weights are **small**?

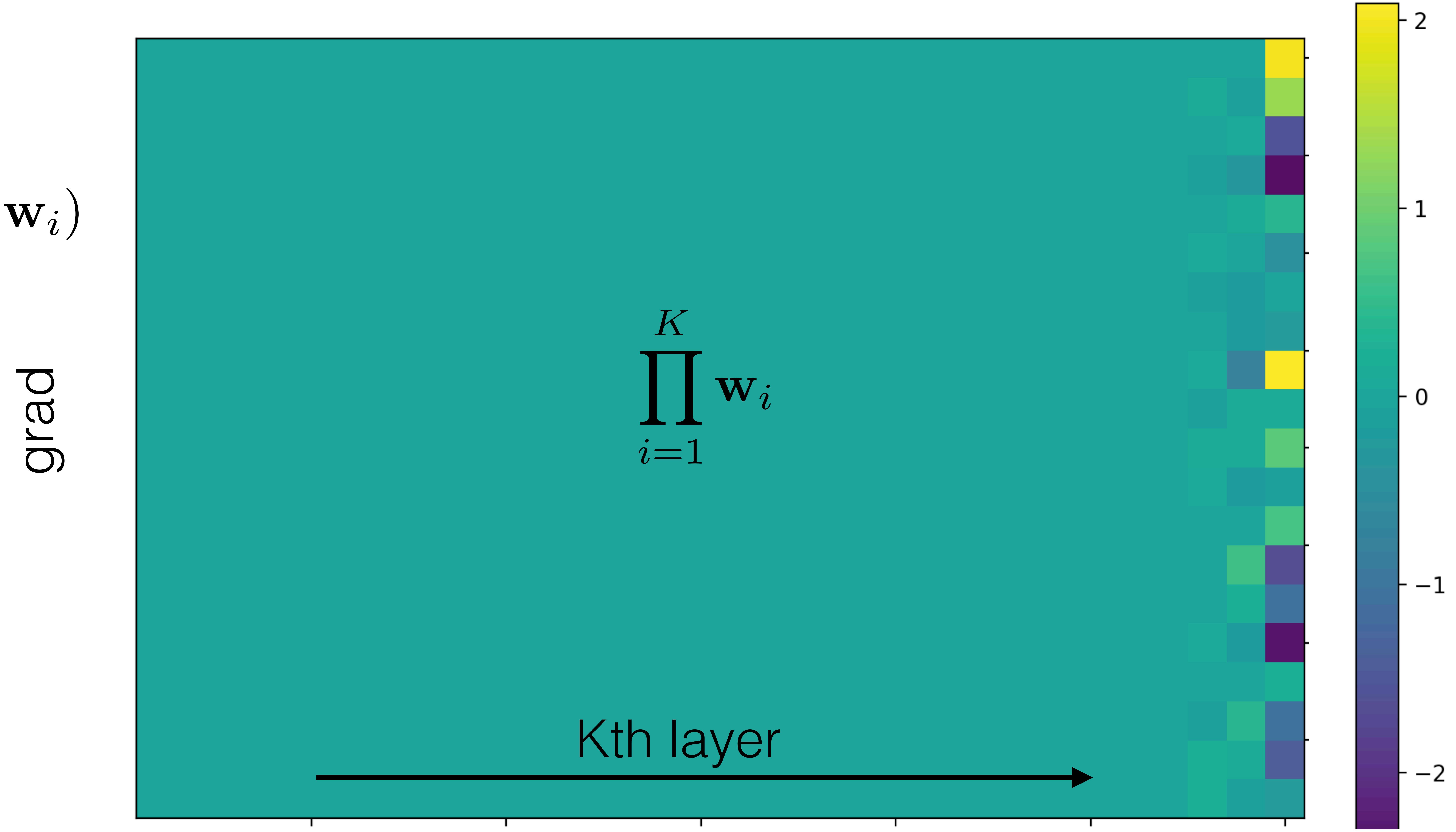
```
x = torch.randn(1000,1)
x.requires_grad_()
y=x
for i in range(30):
    weights = torch.randn(1000,1000)/100
    y =  weights @ y
y.sum().backward()
x.grad
```



Learning

What happens to deep **conv gradient** when weights are **small**?

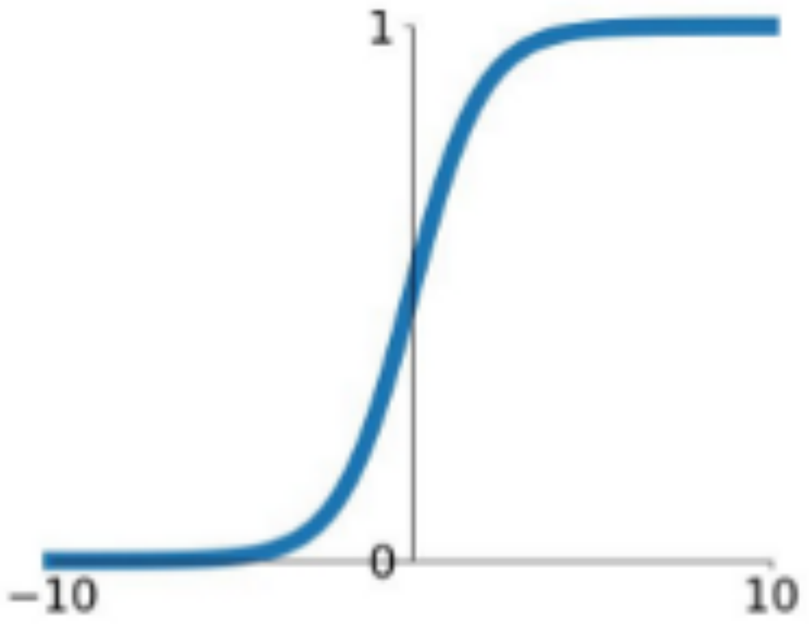
$$\prod_{i=1}^{K} \mathrm{grad}(\mathbf{w}_i)$$



Activation functions

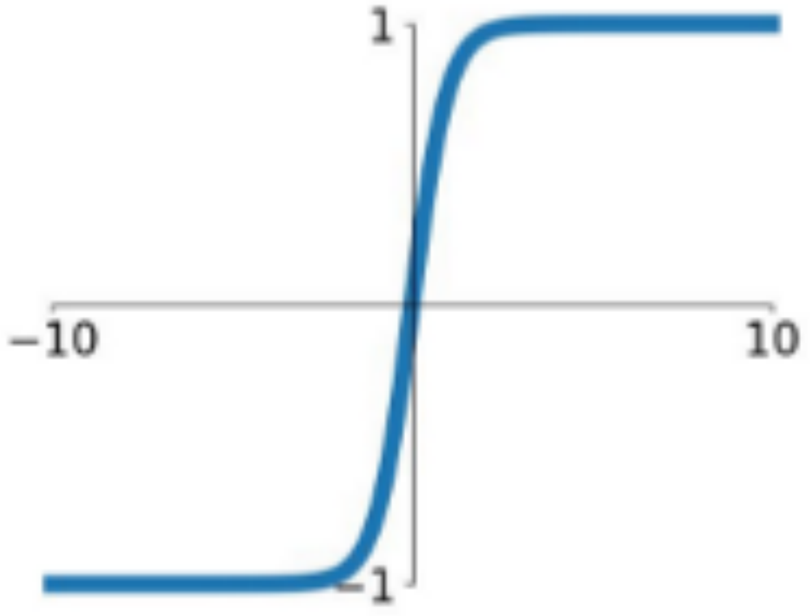
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



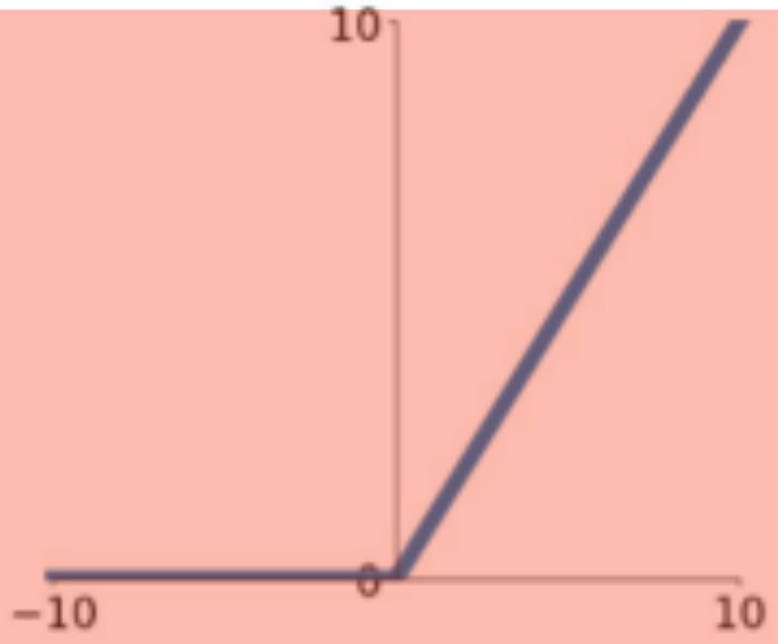
tanh

$$\tanh(x)$$



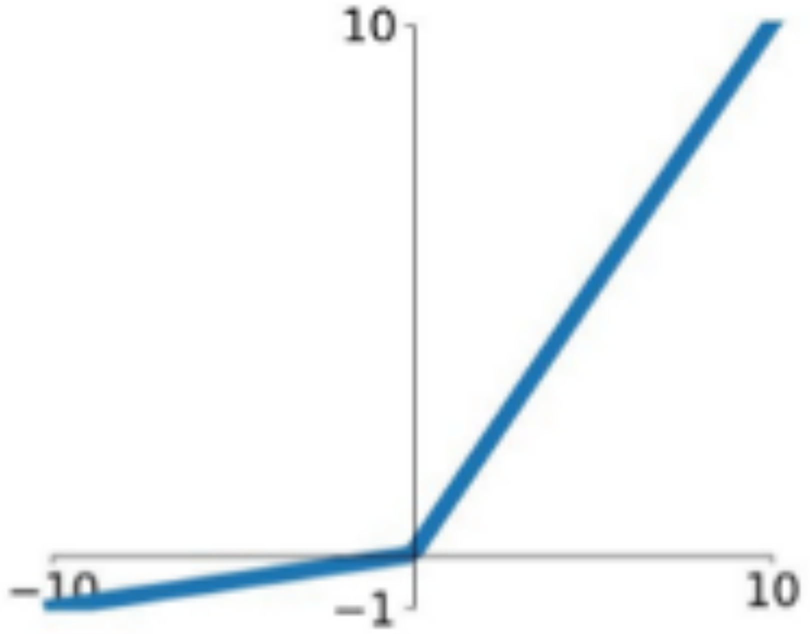
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

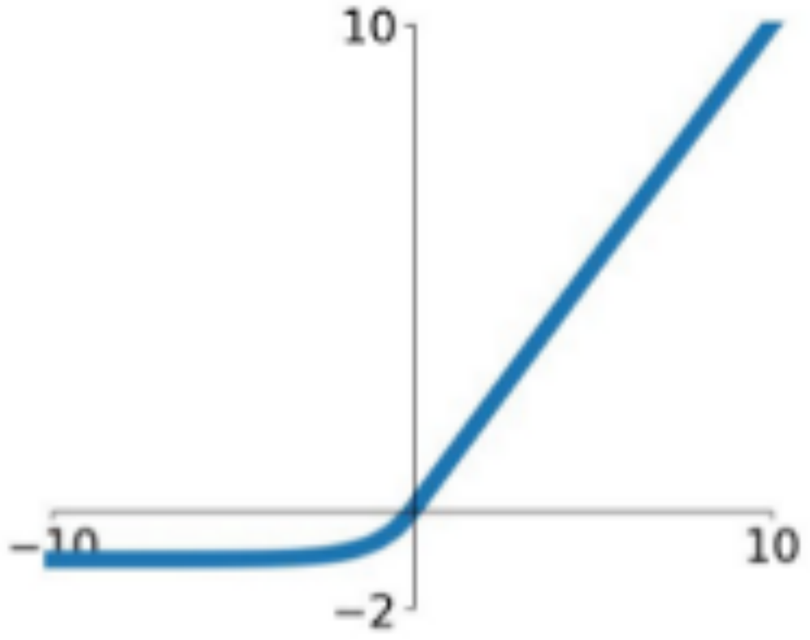


Maxout

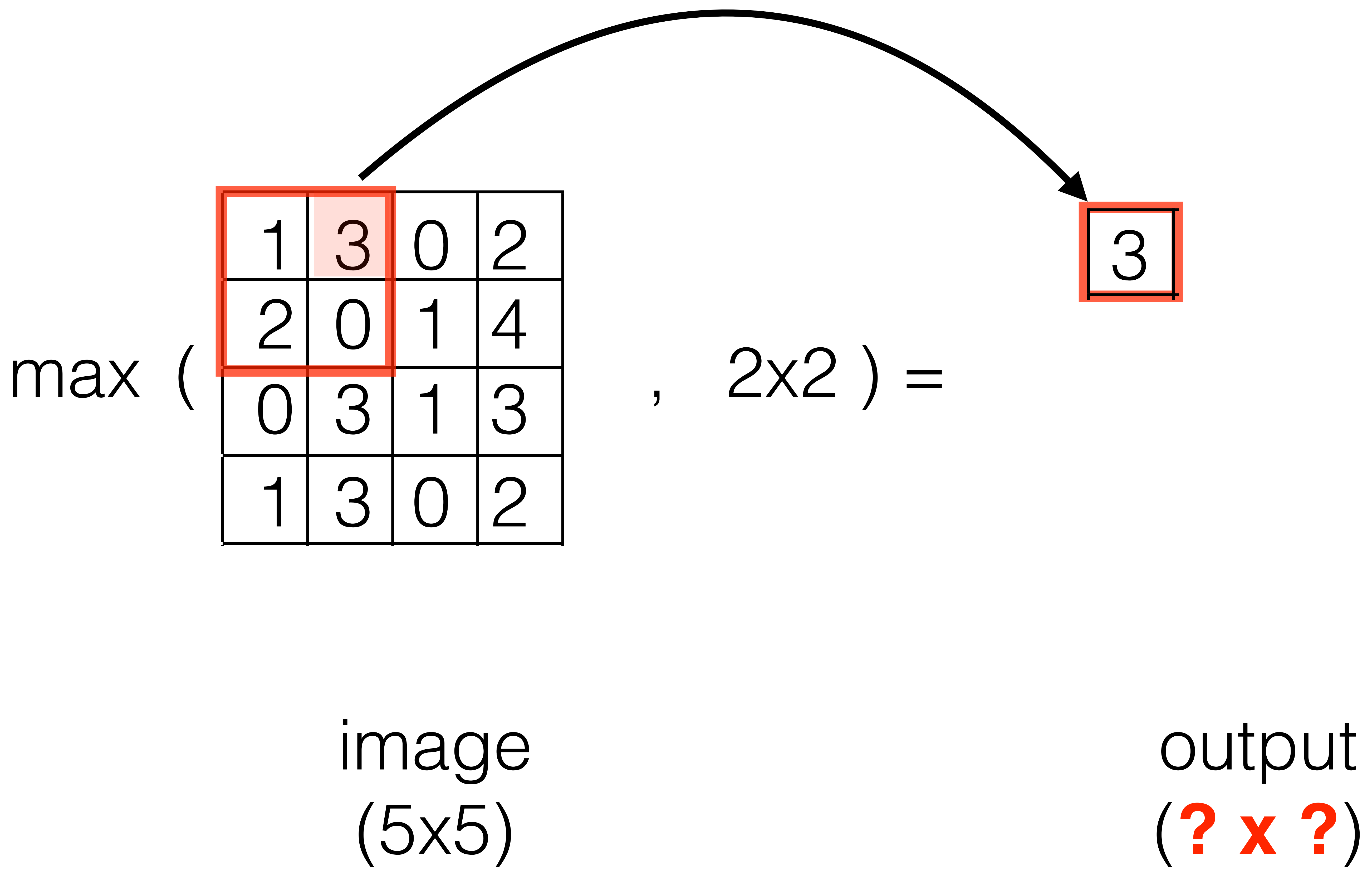
$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Max-pooling



Max-pooling

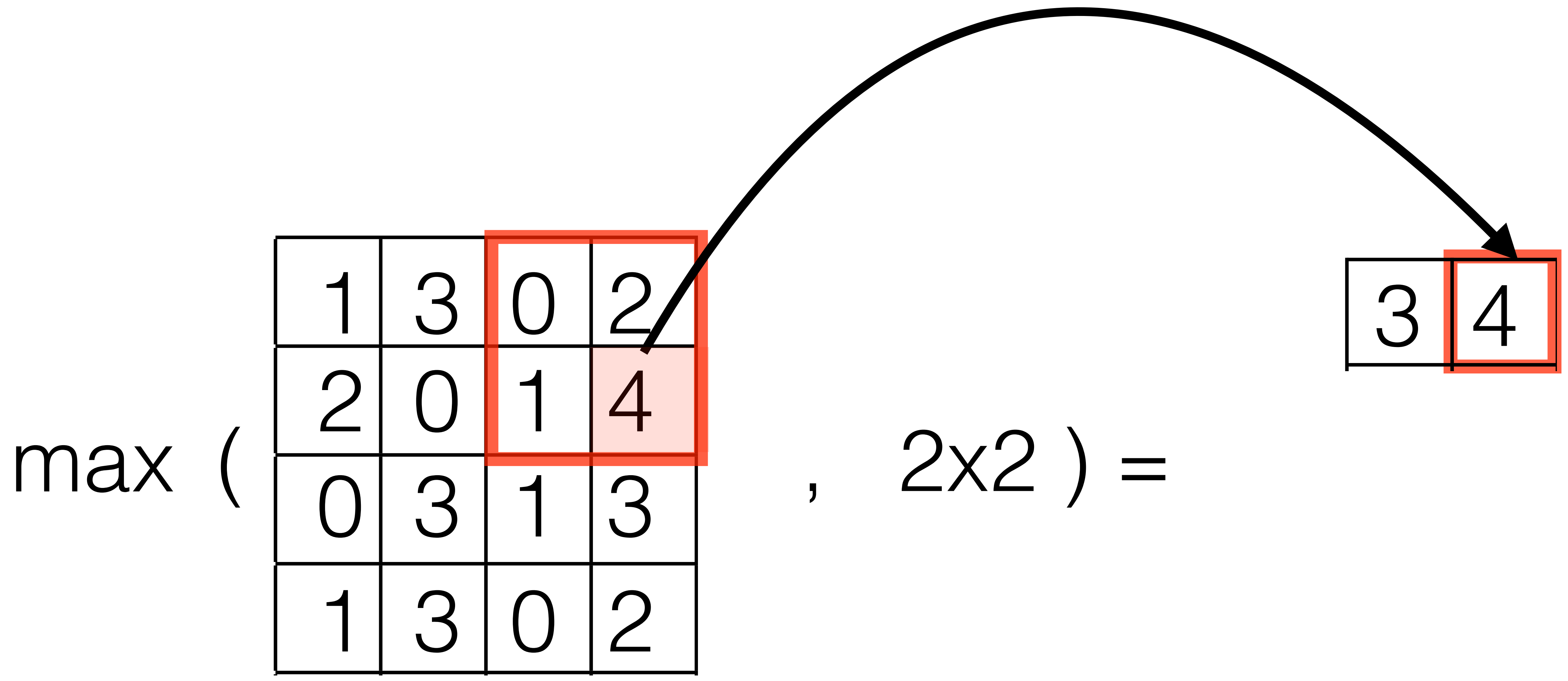


image
(5x5)

output
(**? x ?**)

Max-pooling

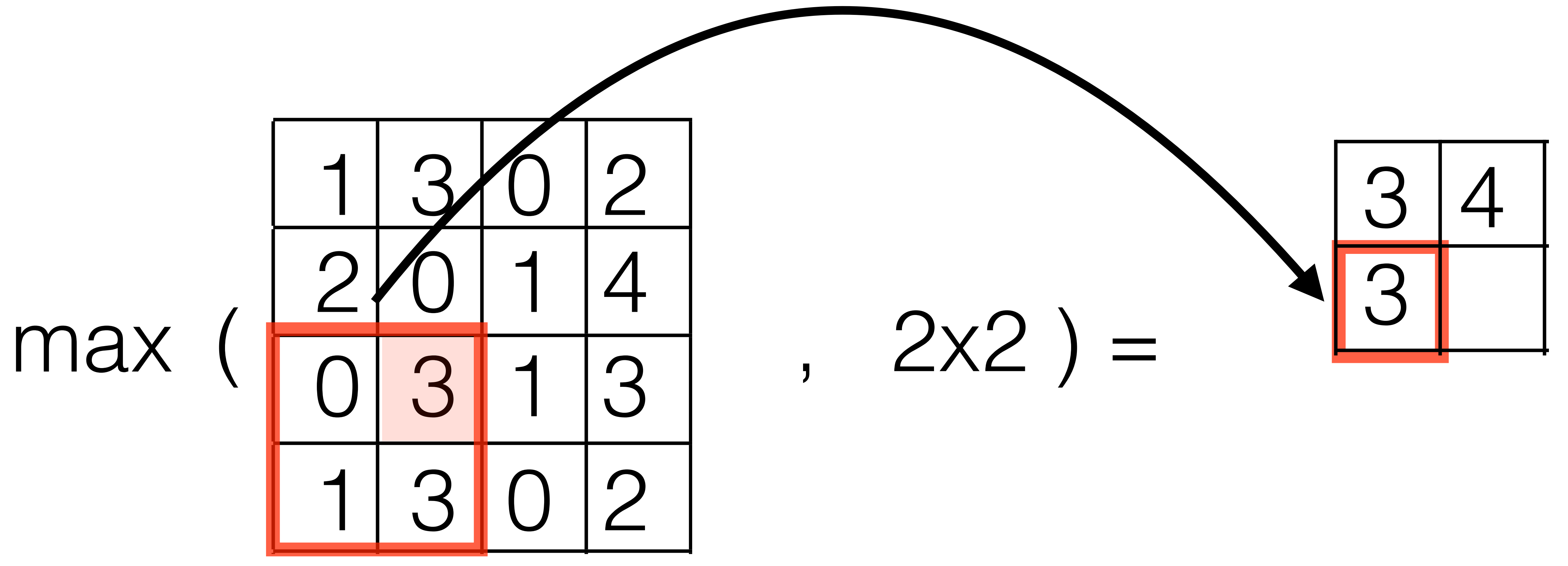


image
(5x5)

output
(**? x ?**)

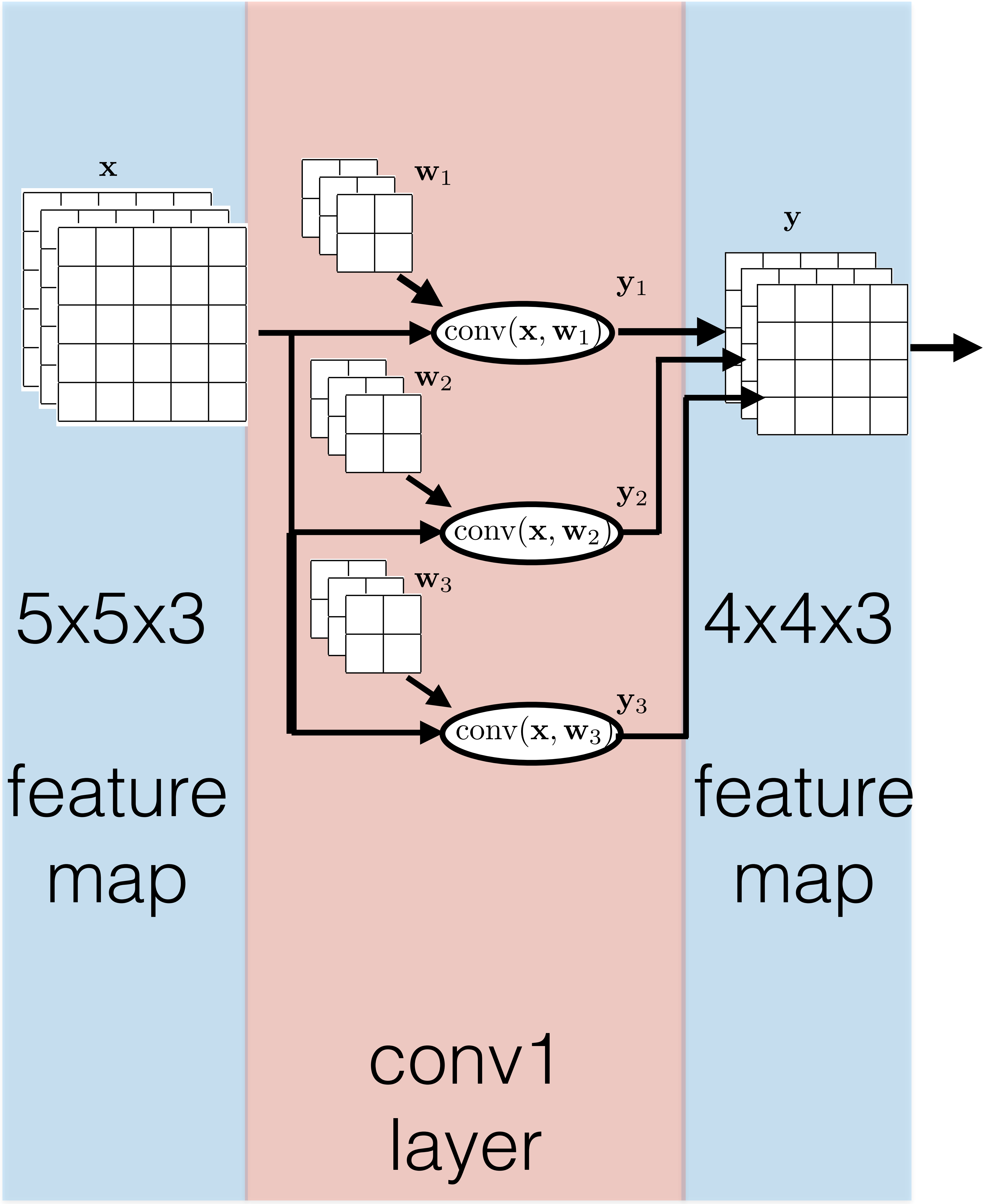
Max-pooling

$$\max\left(\begin{array}{|c|c|c|c|}\hline 1 & 3 & 0 & 2 \\ \hline 2 & 0 & 1 & 4 \\ \hline 0 & 3 & 1 & 3 \\ \hline 1 & 3 & 0 & 2 \\ \hline \end{array}\;,\quad 2x2\right) = \begin{array}{|c|c|}\hline 3 & 4 \\ \hline 3 & \boxed{3} \\ \hline \end{array}$$

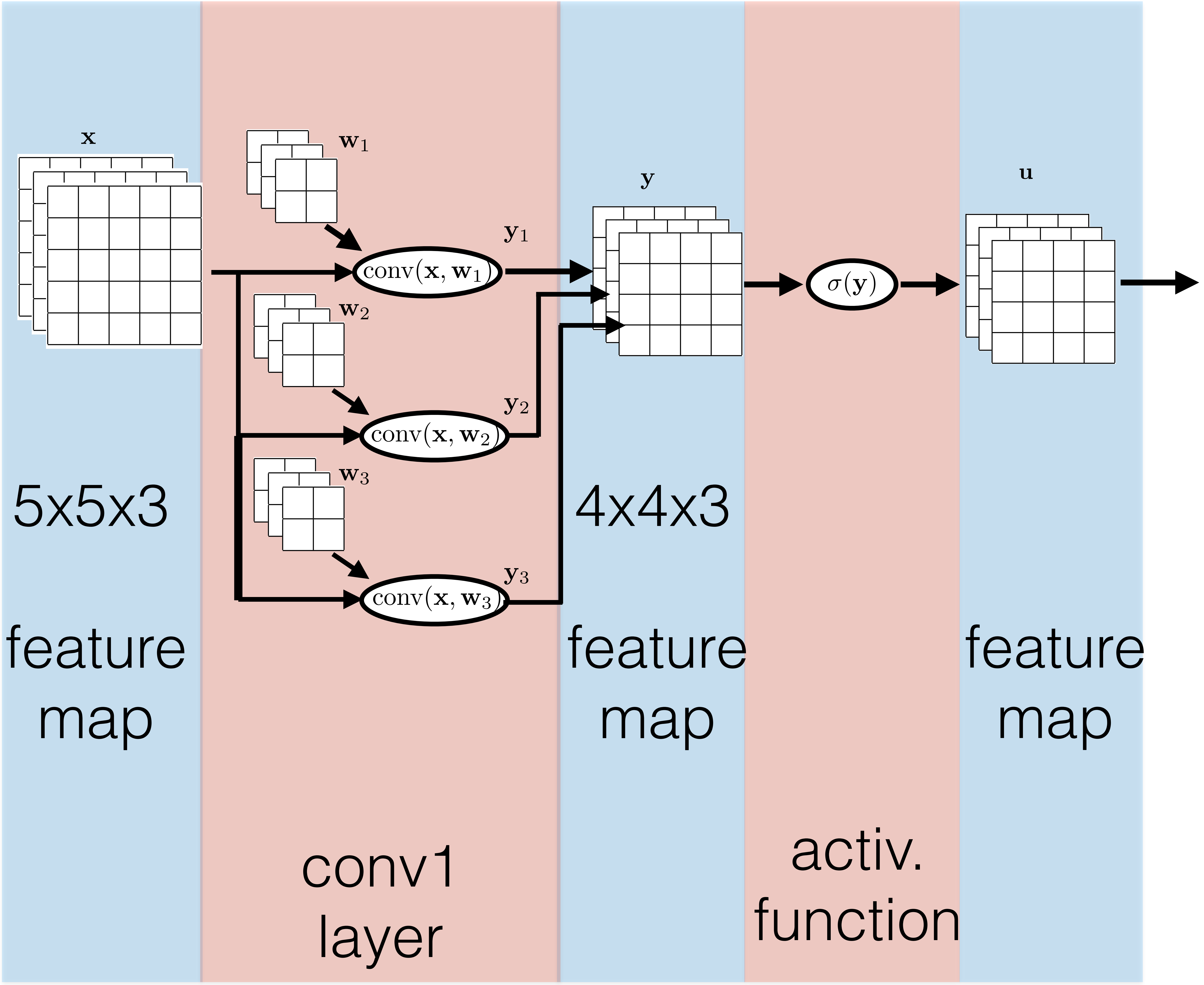
image
(5x5)

output
(**? x ?**)

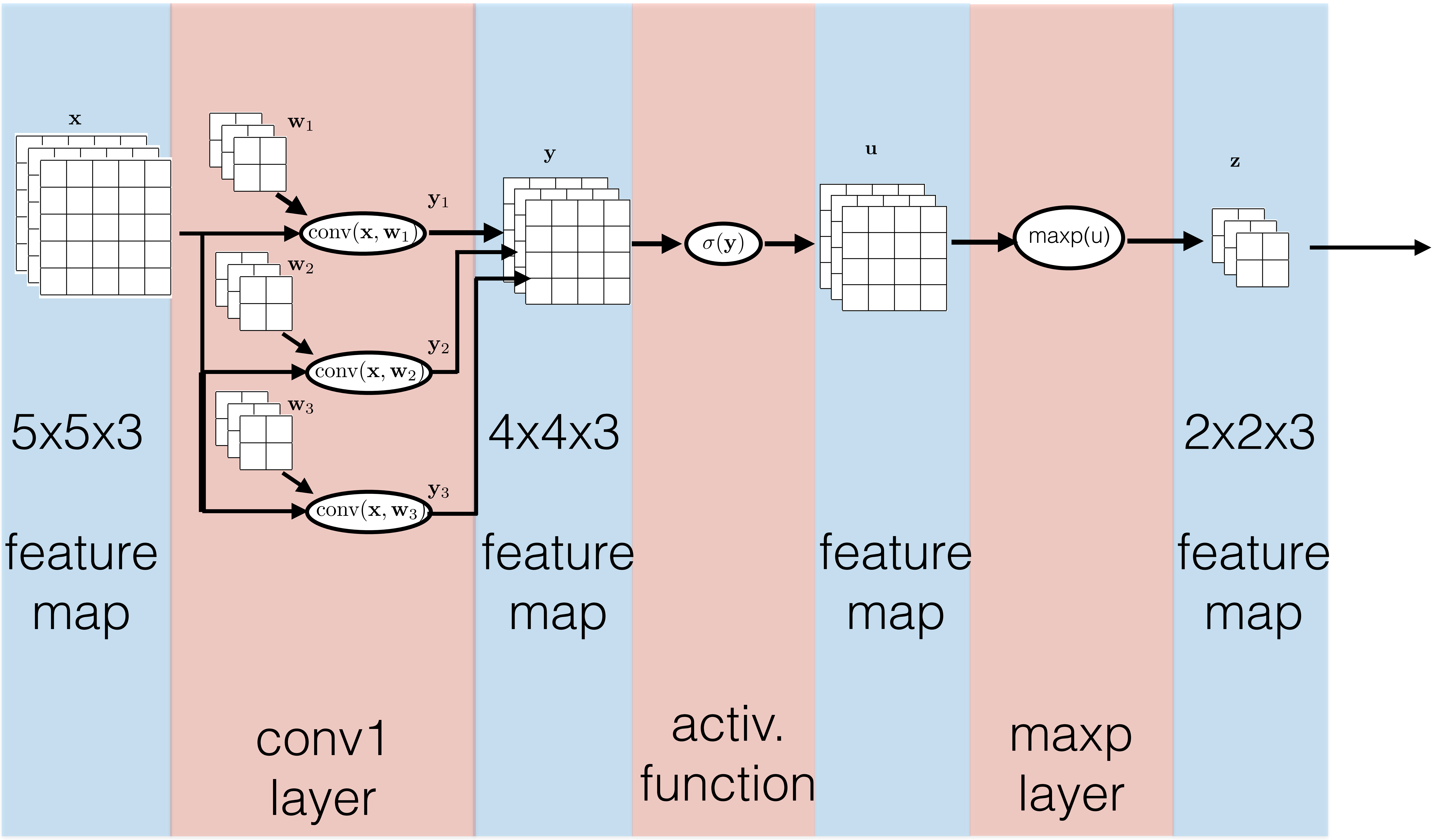
Simple convolutional network



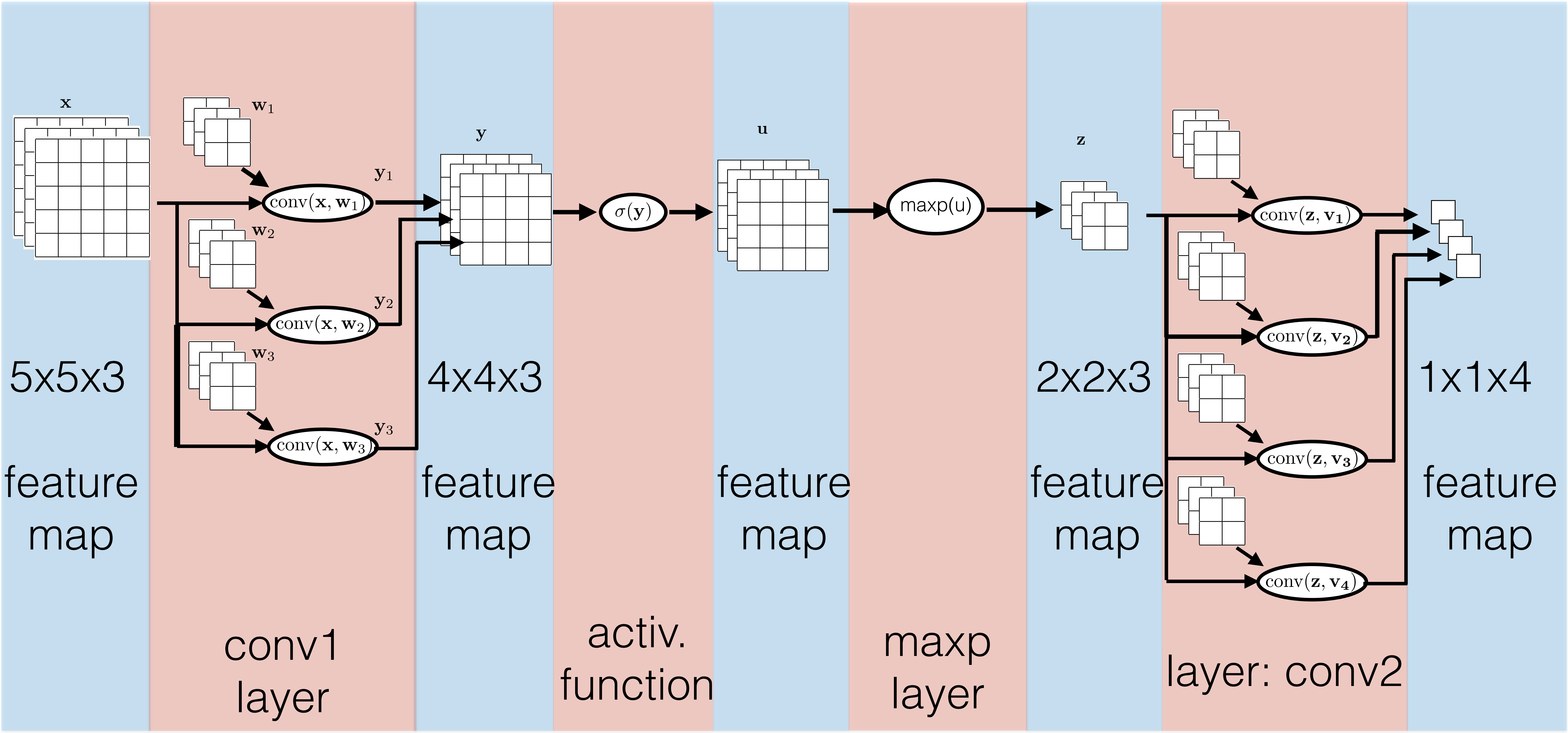
Simple convolutional network



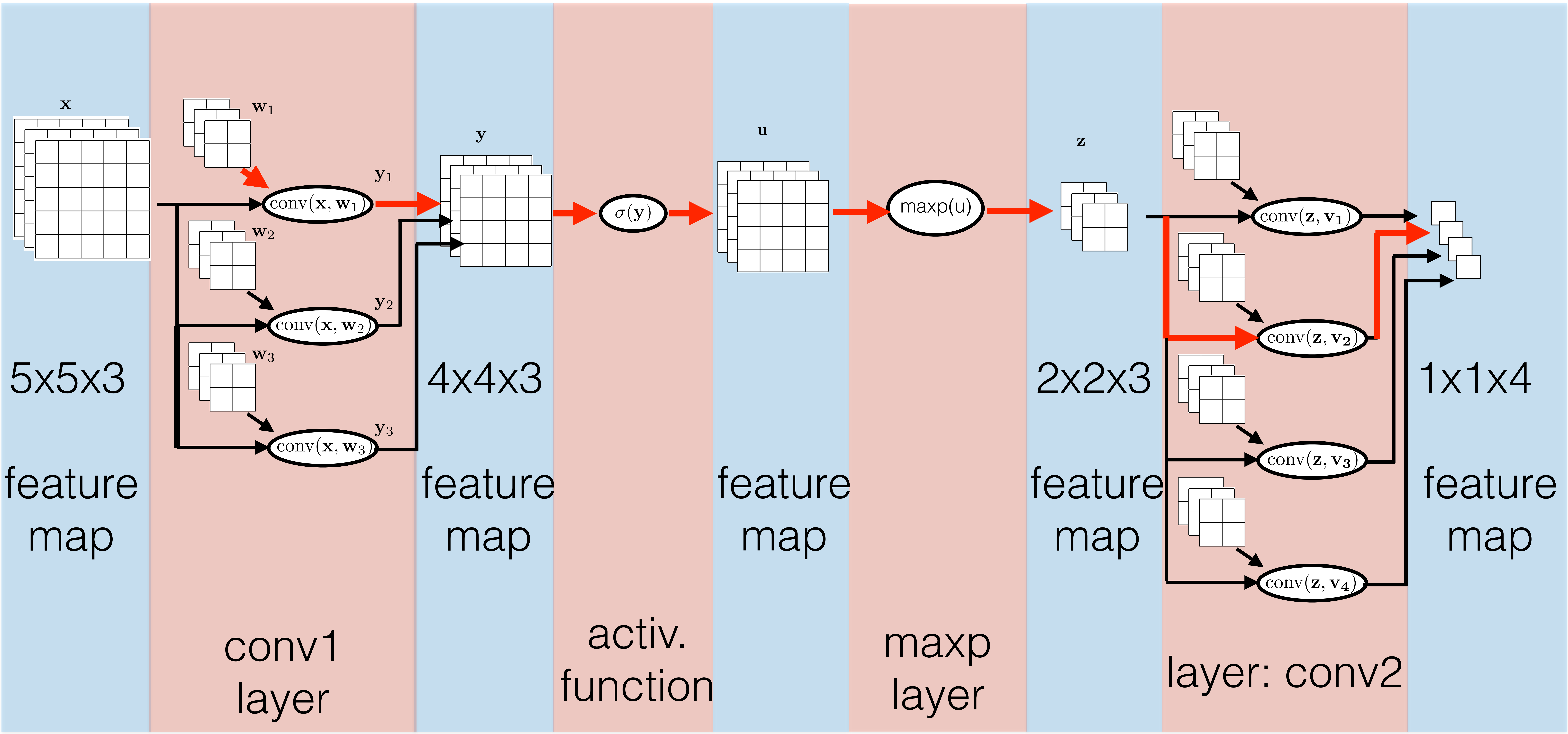
Simple convolutional network



Simple convolutional network



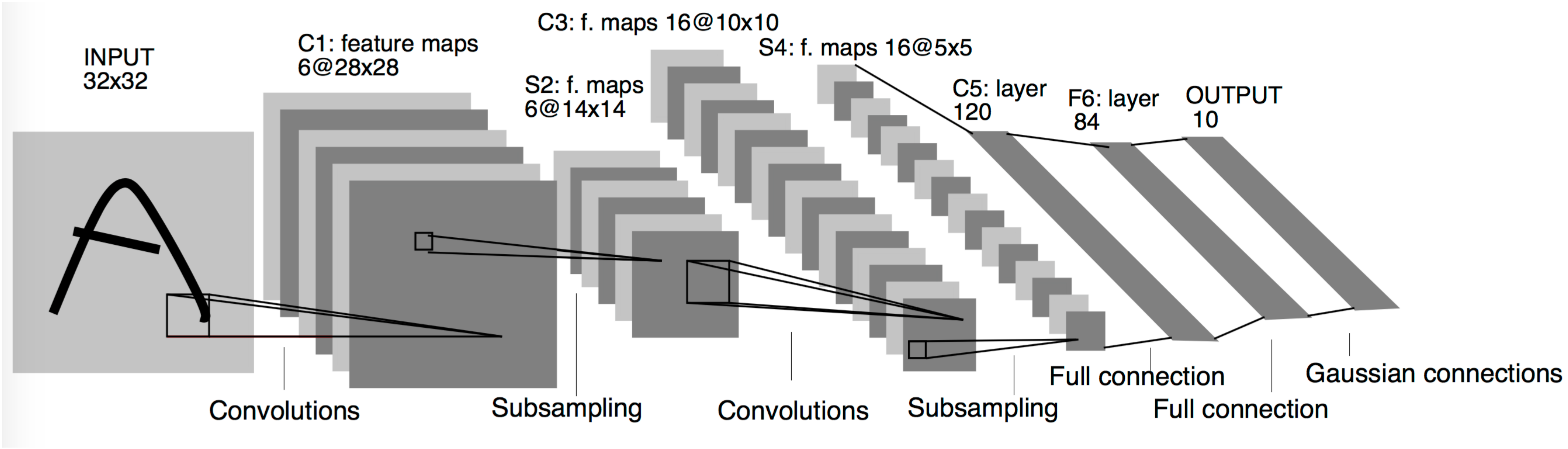
Learning of a simple convolutional network



Convolutional net

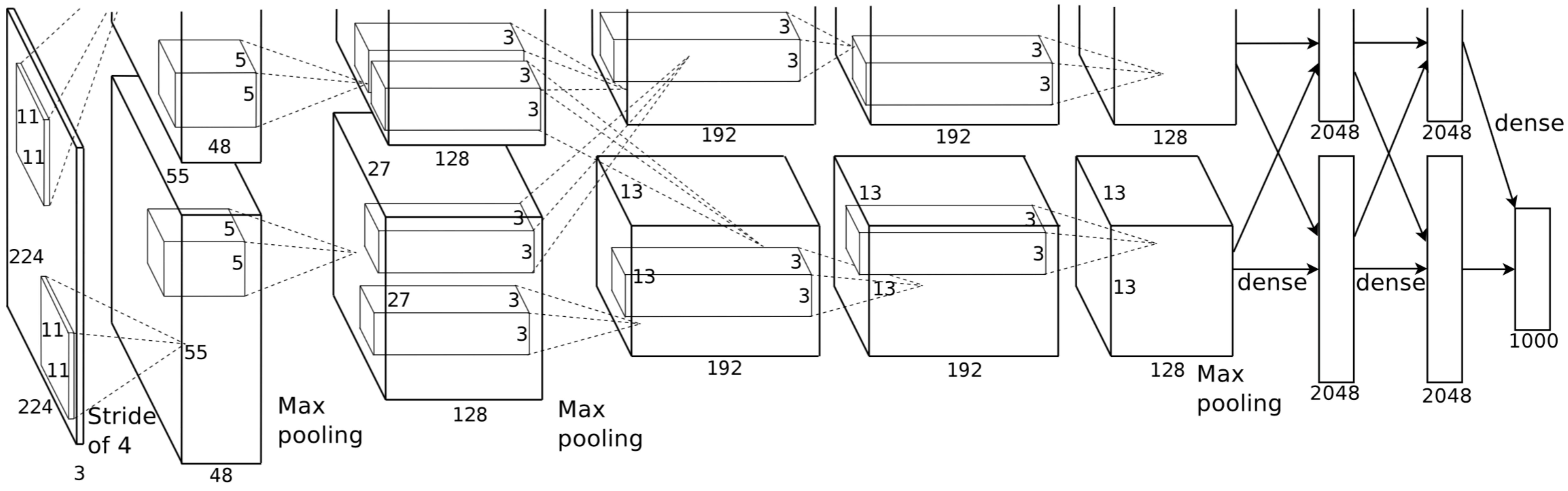
- Convolutional network (ConvNet) is concatenation of convolutional layers, acitvation function and optionally max-pooling functions.
- Backprop in convolutiona layer is convolution of feature maps or kernels or feature maps with the upstream gradient.
- Feed-forward and backprop are convolutions => efficient implementation on GPU

LeCun's letter recognition 1998 (over 13k citations !!!)



LeCun et al, Gradient based learning applied to document recognition, IEEE, 1998
http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf

AlexNet on ImageNet 2012 (**over 27k citations !!!**)



Alex Krizhevsky et al, Imagenet classification with deep convolutional neural networks, NIPS, 2012
https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

IMAGENET

Classification results

http://image-net.org/challenges/LSVRC/2017/index

Steel drum

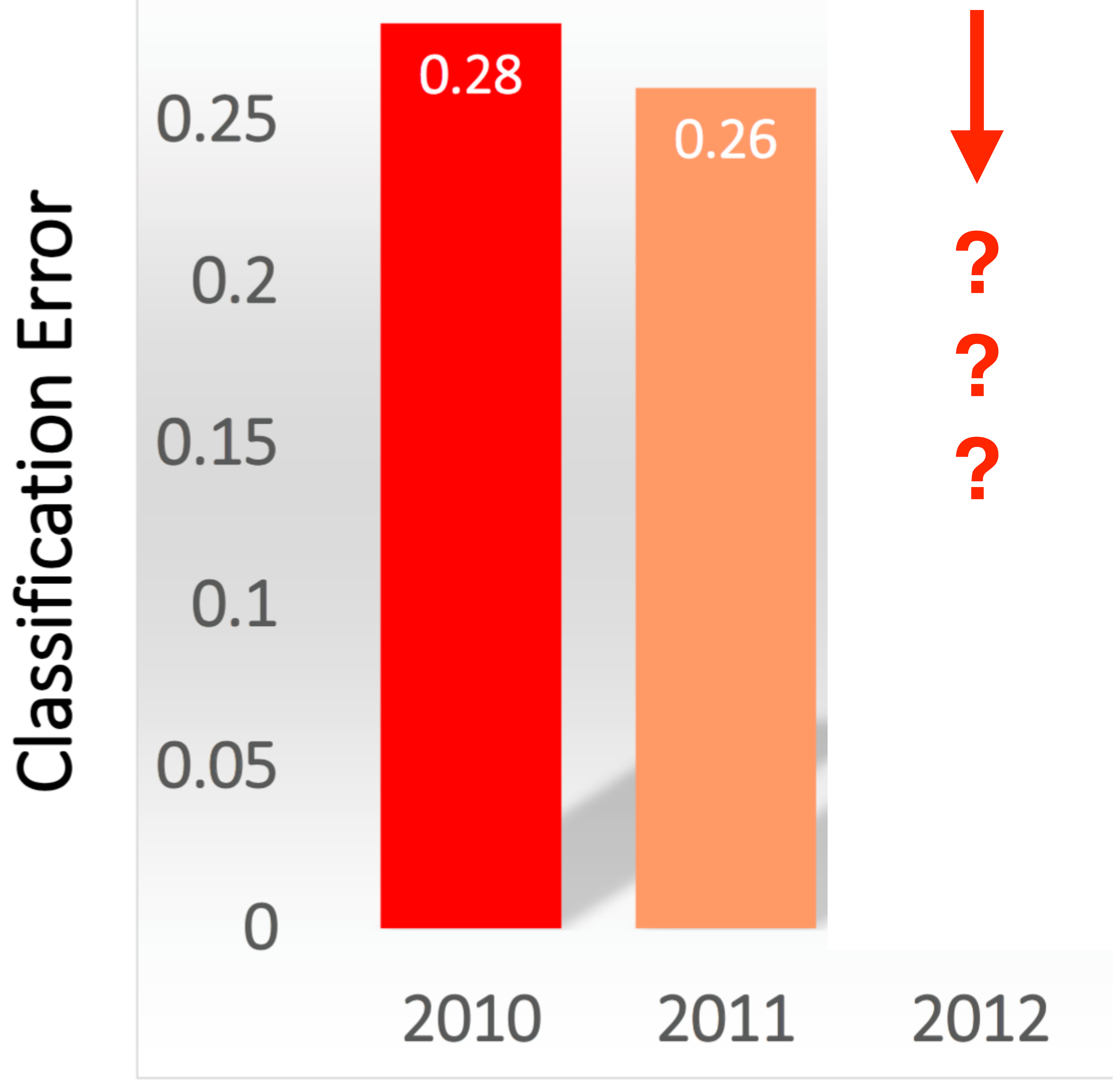
Output:	Output:
Scale	Scale
T-shirt	T-shirt
Steel drum	Giant panda
Drumstick	Drumstick
Mud turtle	Mud turtle

$$\text{Error} = \frac{1}{100{,}000} \sum_{\substack{100{,}000 \\ \text{images}}} 1[\text{incorrect on image } i]$$

IMAGENET

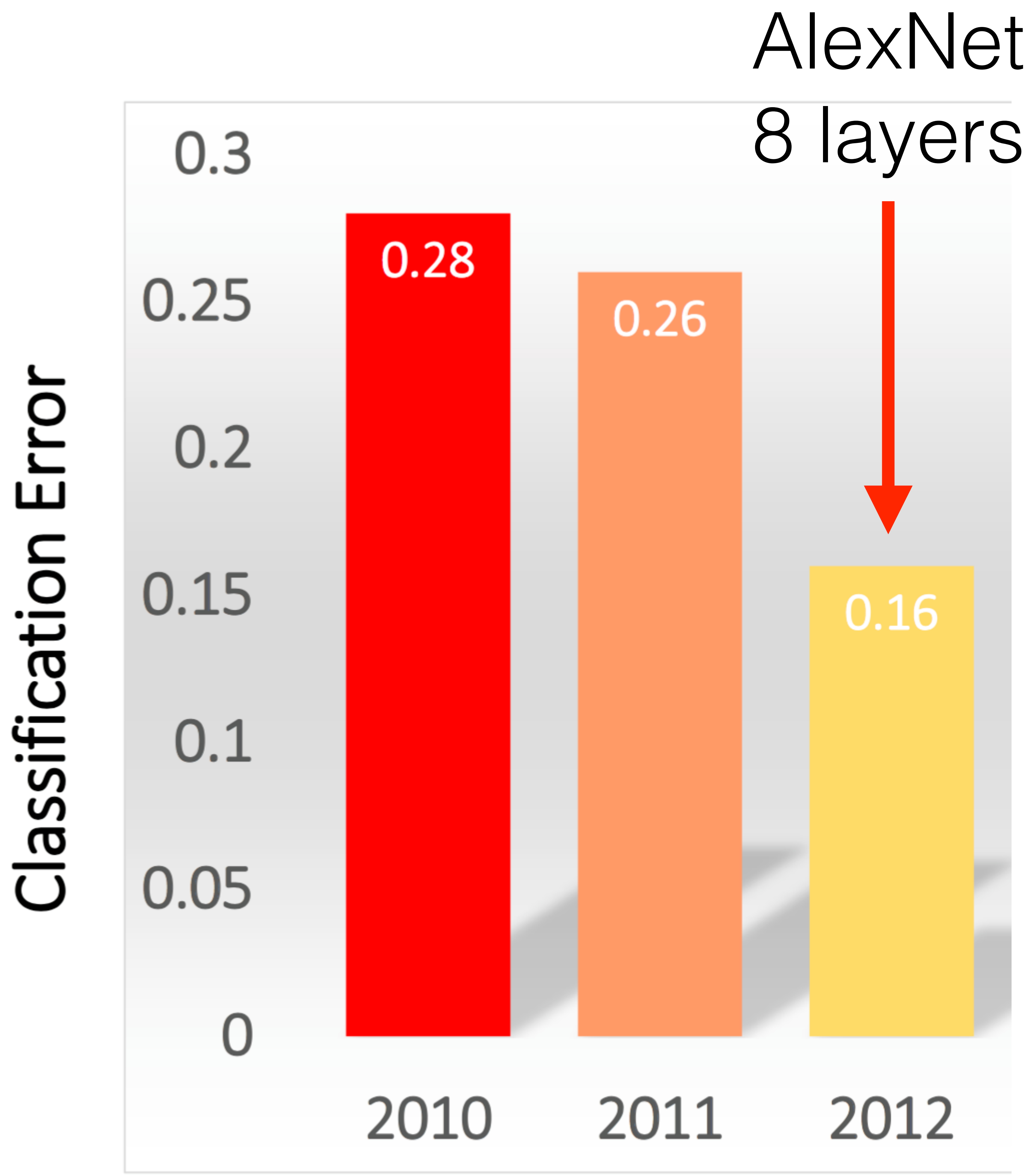
Classification results

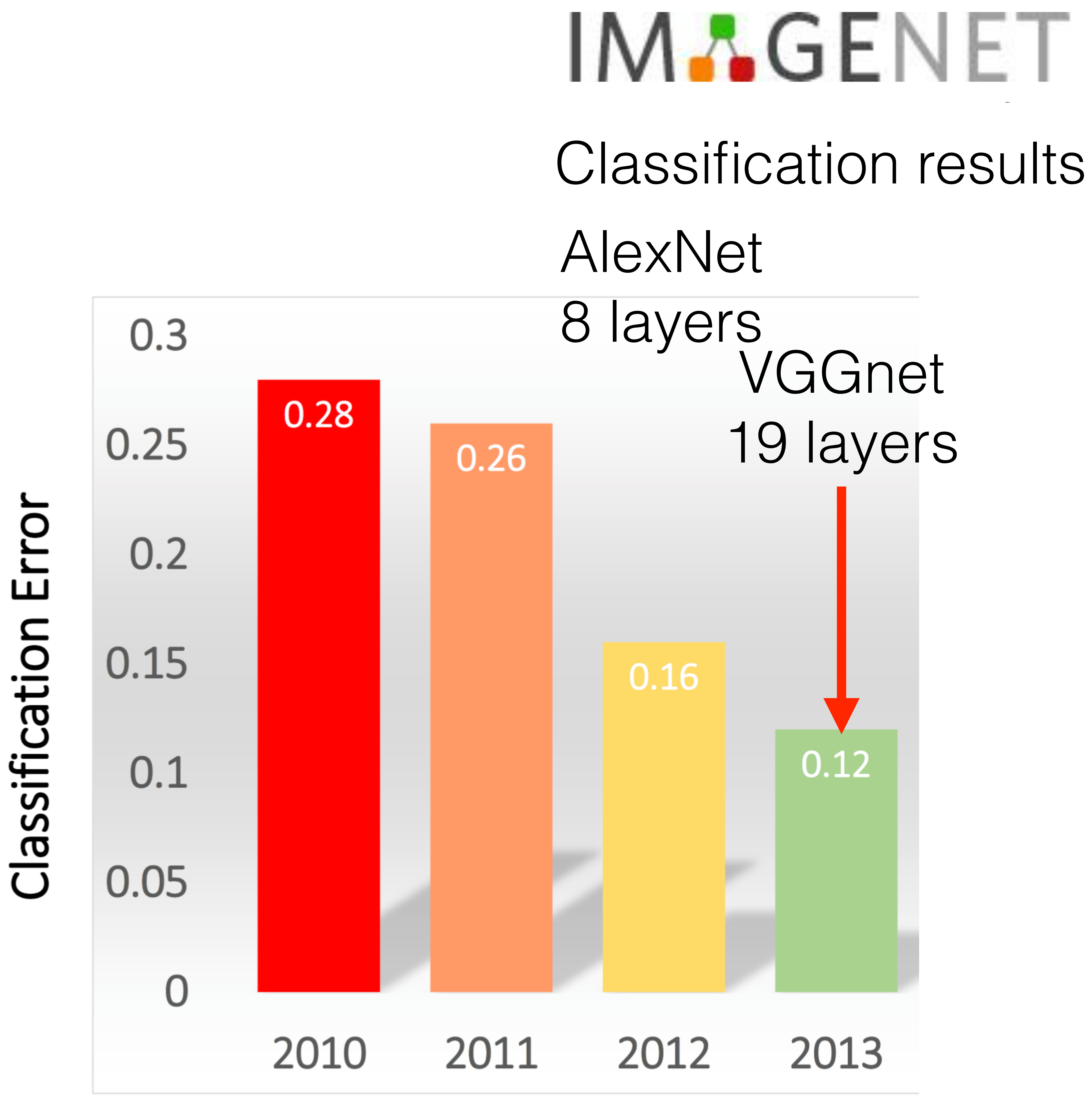
AlexNet
8 layers



IMAGENET

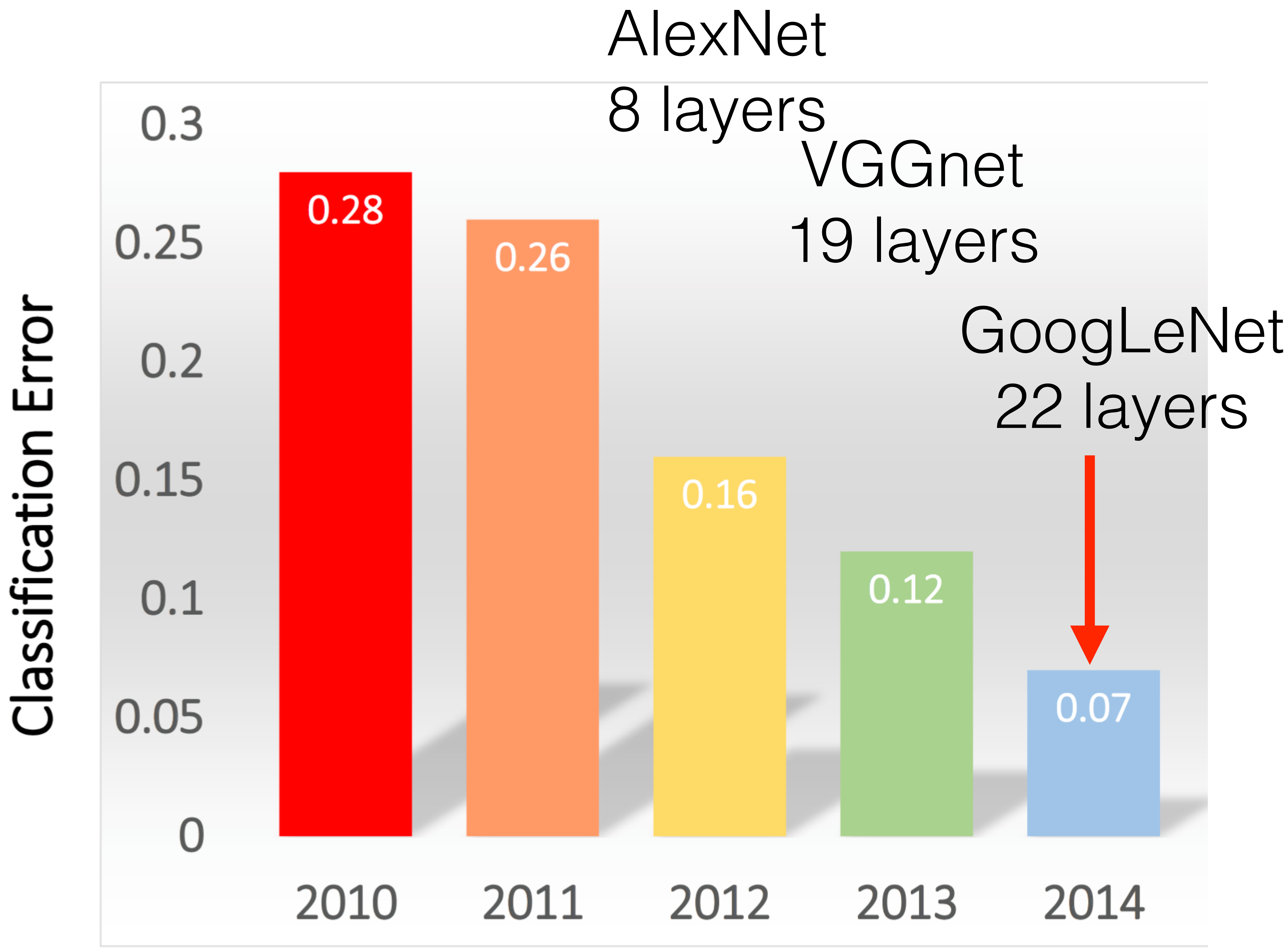
Classification results



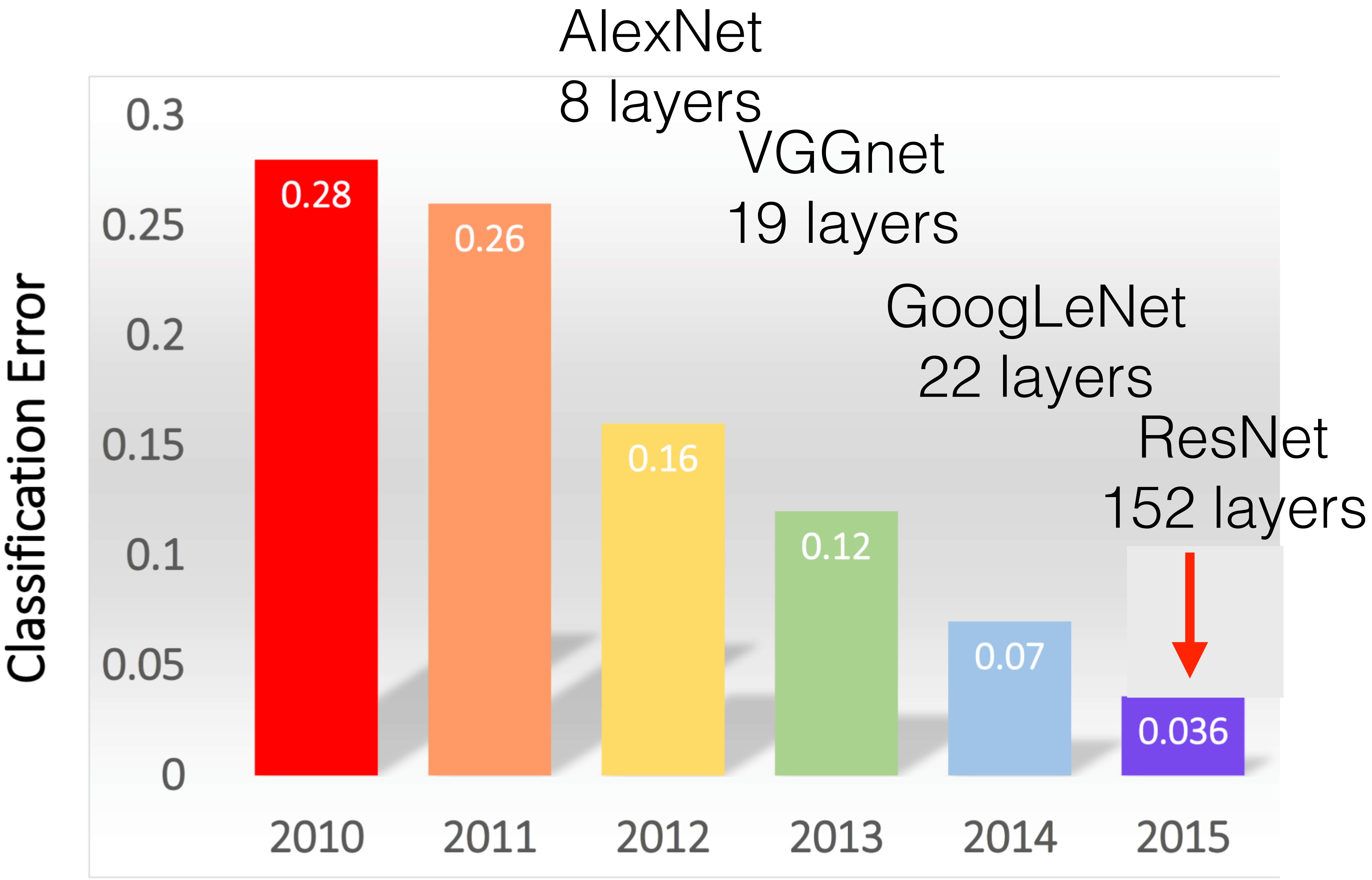

IMAGENET
Classification results
AlexNet
8 layers
VGGnet
19 layers
Classification Error
0.3
0.25
0.2
0.15
0.1
0.05
0
0.28
0.26
0.16
0.12
2010
2011
2012
2013

IMAGENET

Classification results

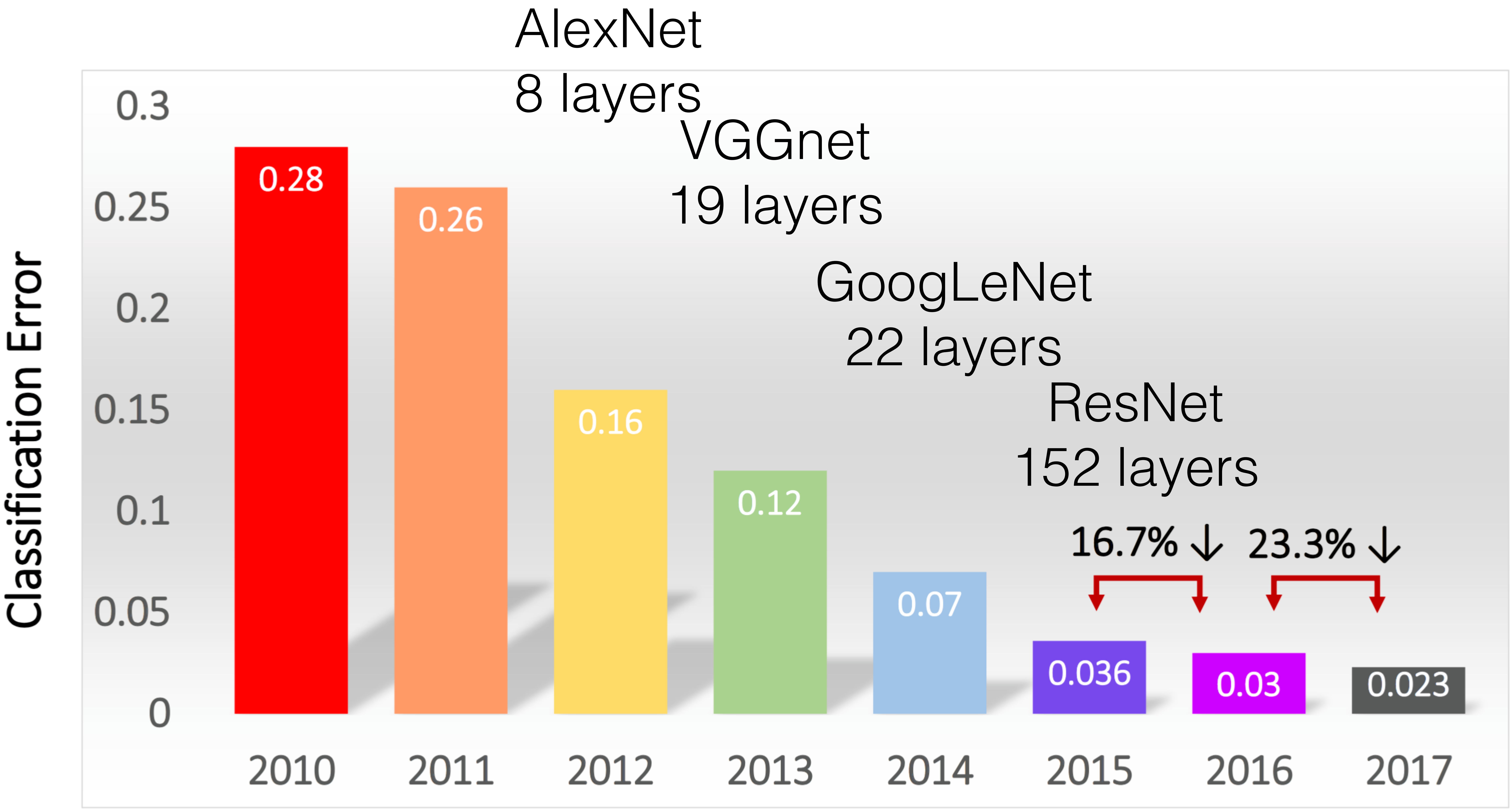


Classification results

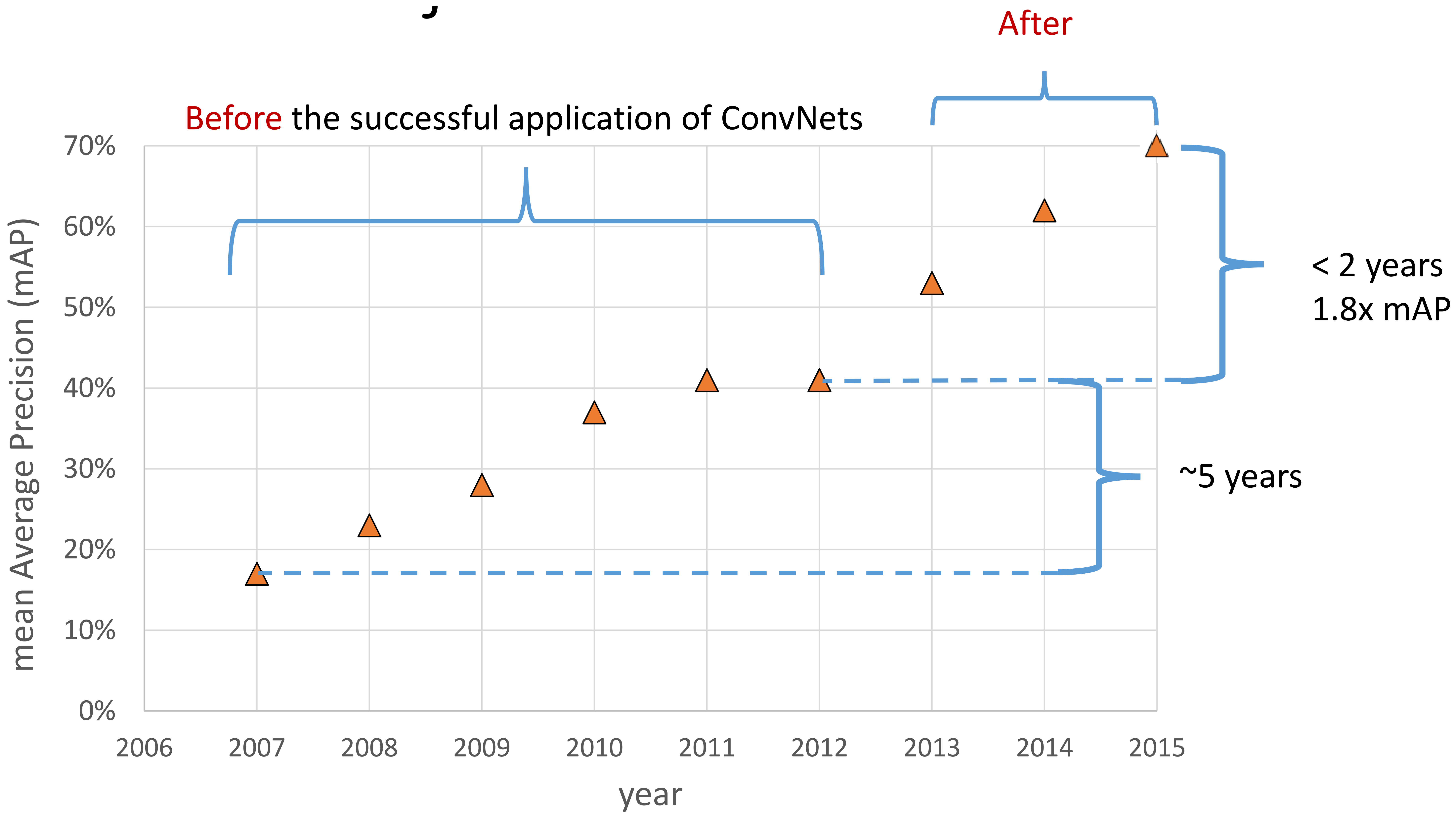


IMAGENET

Classification results



Pascal VOC object detection challenge



Test competencies

- ConvNet/Layer feed-forward pass
- ConvNet/Layer backpropagation
- Diminishing and exploding values/gradient in deep ConvNets

Next lecture

- gradient learning (what makes it tough)
- other layers:
  - deconvolution layer
  - activation function, maxpool,
  - batch normalization,
  - drop out,
  - loss layers