Learning a neural net amounts to curve fitting

We’re just estimating a function

Neural Net as Function Approximation

Given `input`, learn a computer program that computes `output`

this is a **function**

Multinomial logistic regression:

```
def f(input):
    output = softmax(np.dot(input, W) + b)
    return output
```

the only things that we are learning
(we fix their dimensions in advance)

We are fixing what the function `f` looks like in code
and are only adjusting `W` and `b`!!!

Neural Net as Function Approximation

Given `input`, learn a computer program that computes `output`

Multinomial logistic regression:

```
output = softmax(np.dot(input, W) + b)
```

Multilayer perceptron:

```
intermediate = relu(np.dot(input, W1) + b1)

output = softmax(np.dot(intermediate, W2) + b2)
```

Learning a neural net: learning a simple computer program that maps inputs (raw feature vectors) to outputs (predictions)

Complexity of a Neural Net?

- Increasing number of layers (depth) makes neural net more "complex"
  - Learn computer program that has more lines of code
  - Some times, more parameters may be needed
    - If so, more training data may be needed

Earlier: multinomial logistic regression had fewer parameters than multilayer perceptron example

Upcoming: we'll see examples of deep nets with *fewer* parameters than "shallower" nets

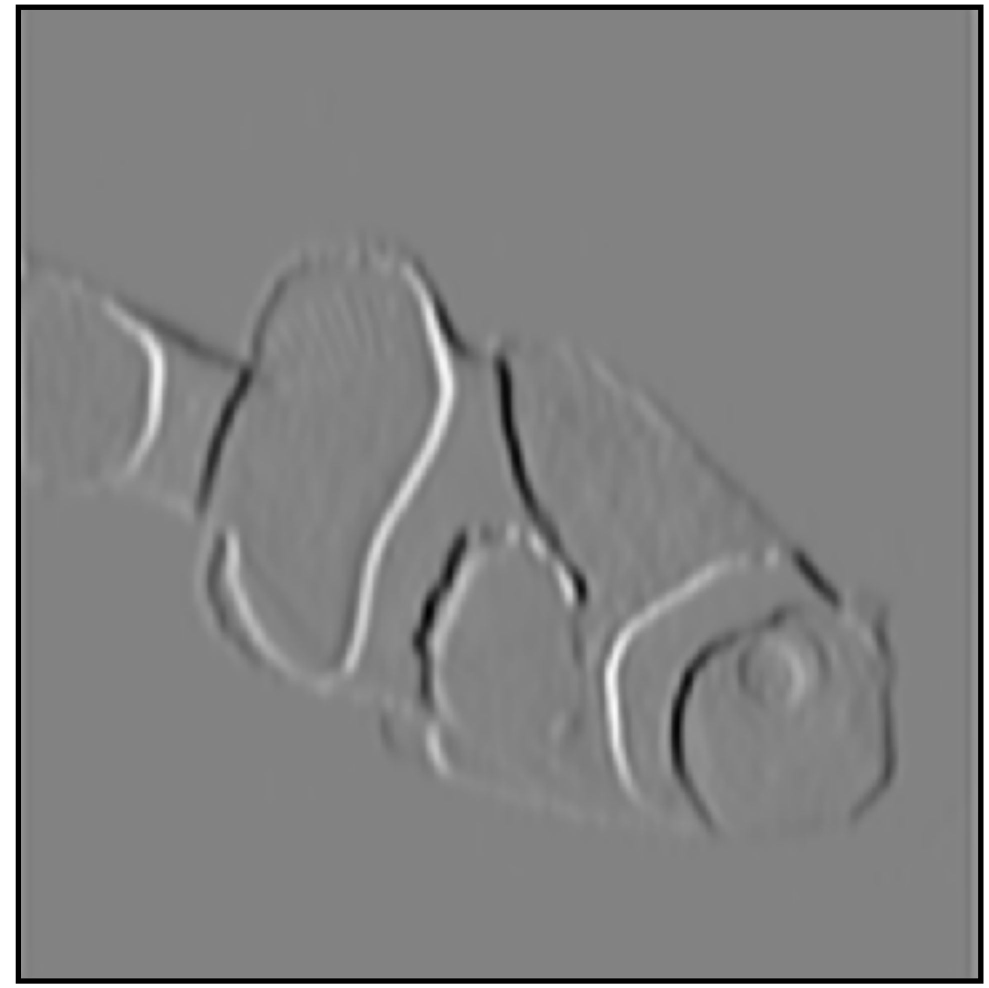
Carnegie Mellon University
Heinz College

Unstructured Data Analysis

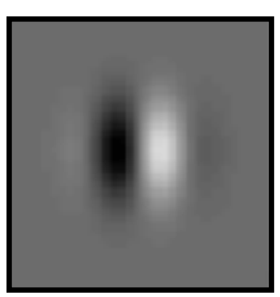
Lecture 13: Image analysis with convolutional neural nets (also called CNNs or convnets)

George Chen

Convolution



filter



Slide by Phillip Isola

Convolution

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

Input image

0	0	0
0	1	0
0	0	0

Filter
(also called “kernel”)

Convolution

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

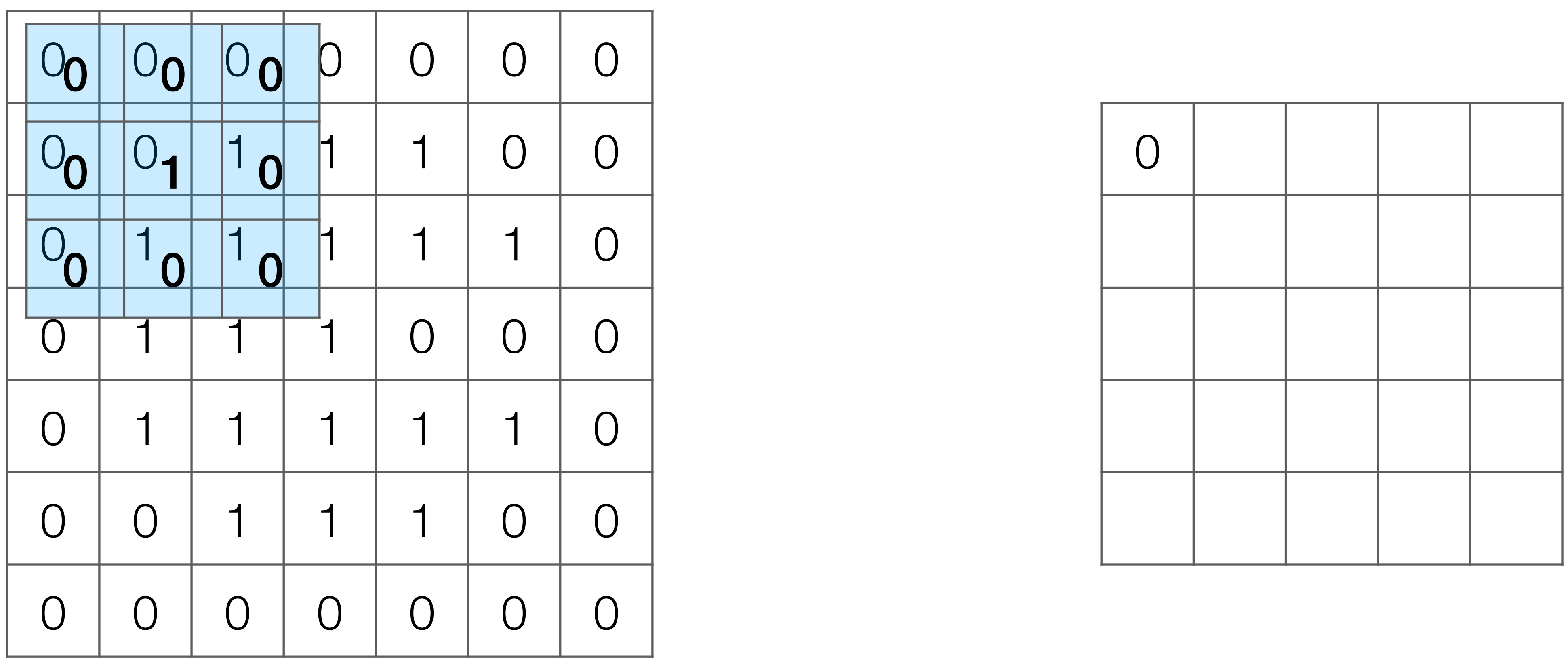
Input image

0	0	0
0	1	0
0	0	0

Filter
(also called “kernel”)

Convolution

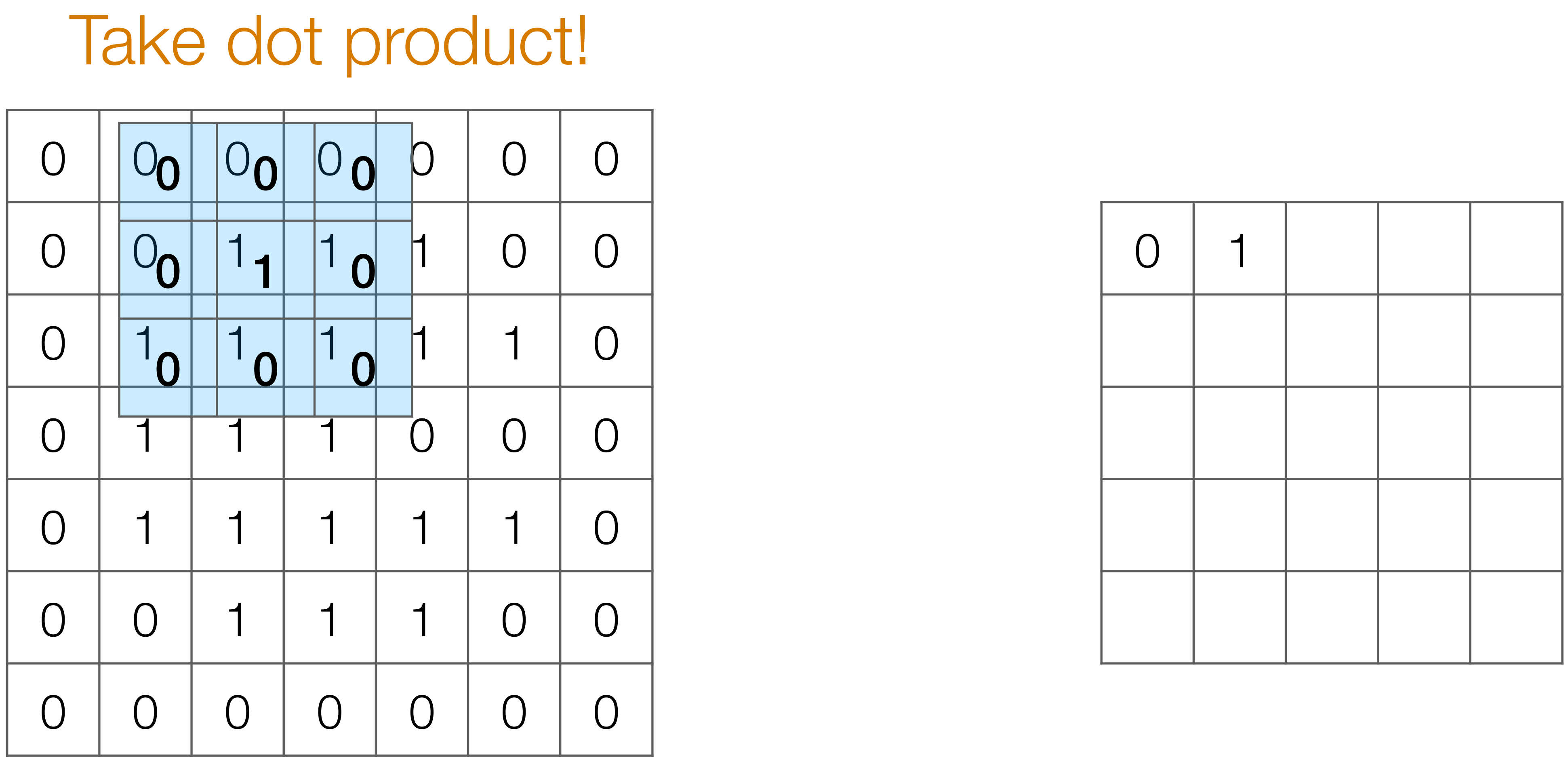
Take dot product!



Input image

Output image

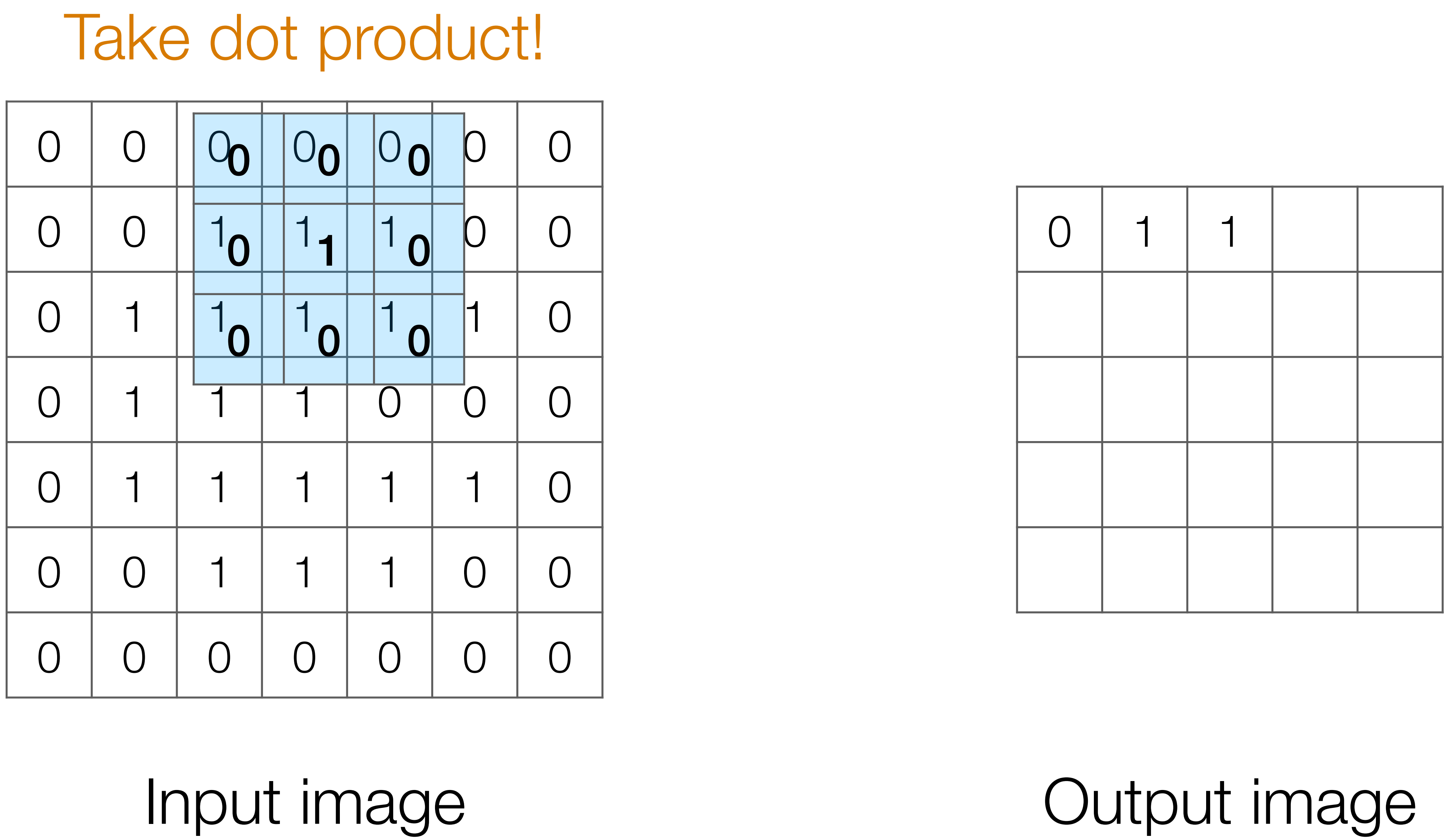
Convolution



Input image

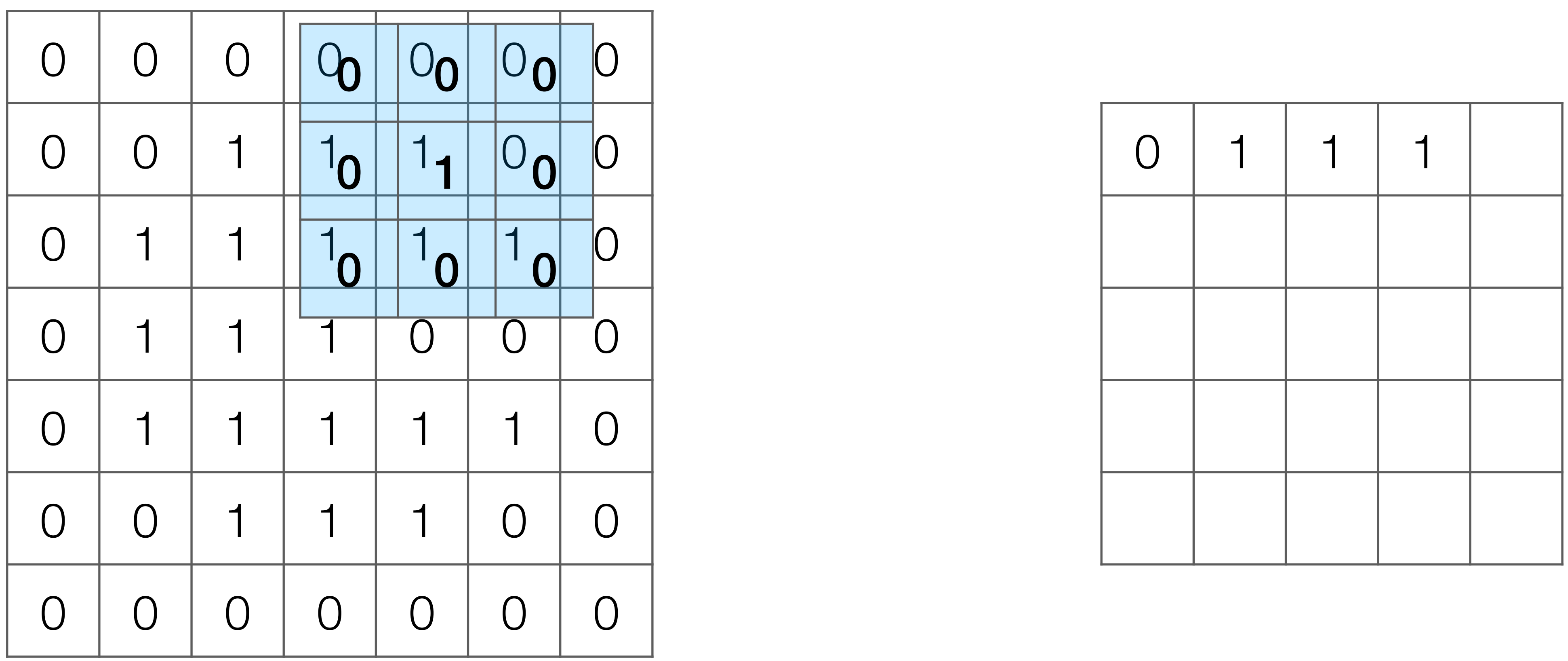
Output image

Convolution



Convolution

Take dot product!

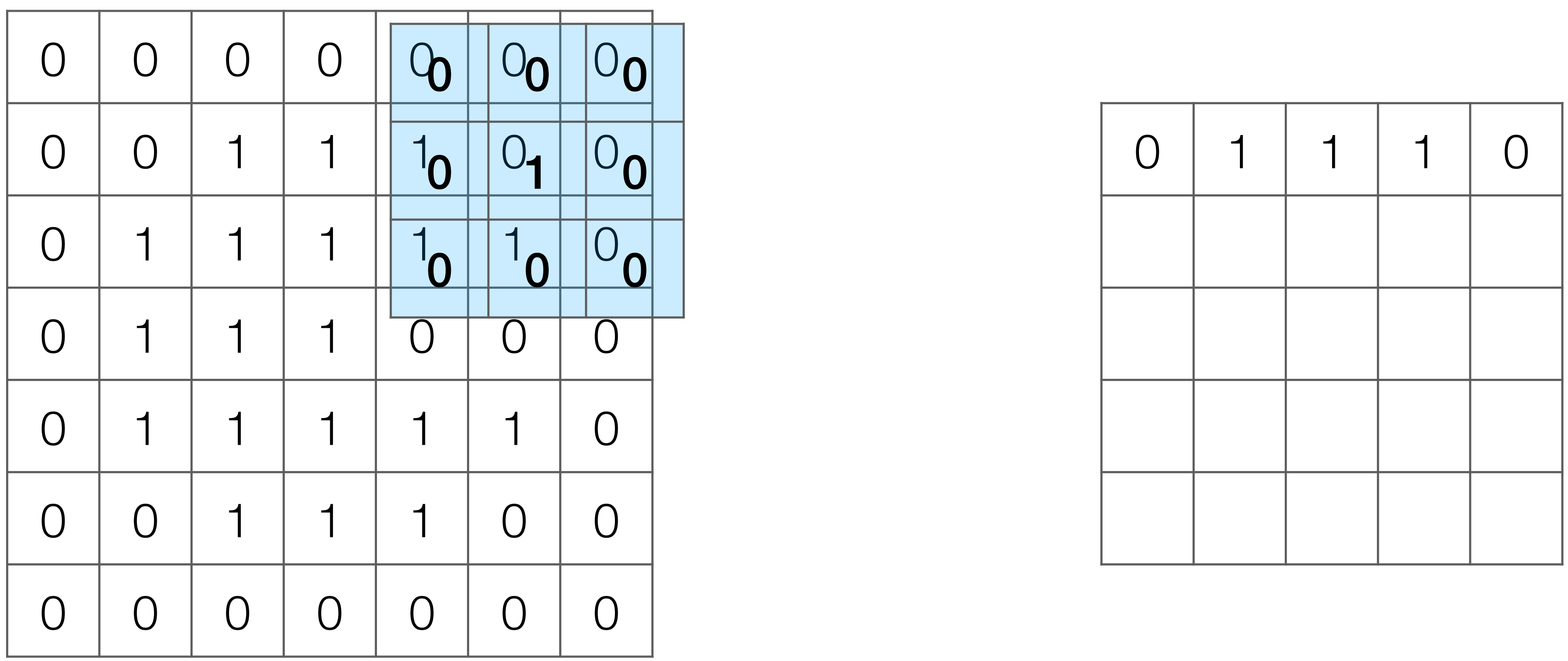


Input image

Output image

Convolution

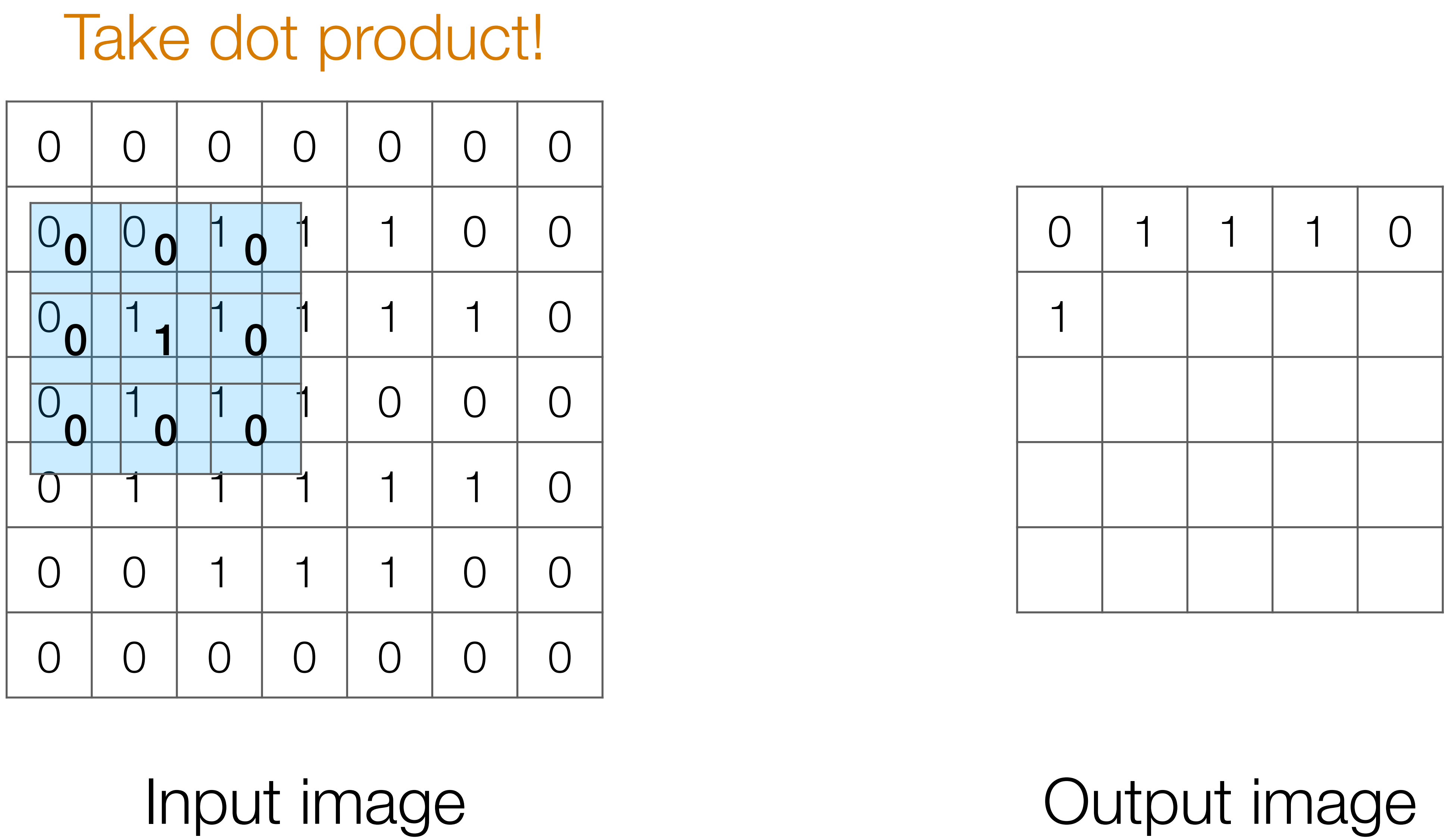
Take dot product!



Input image

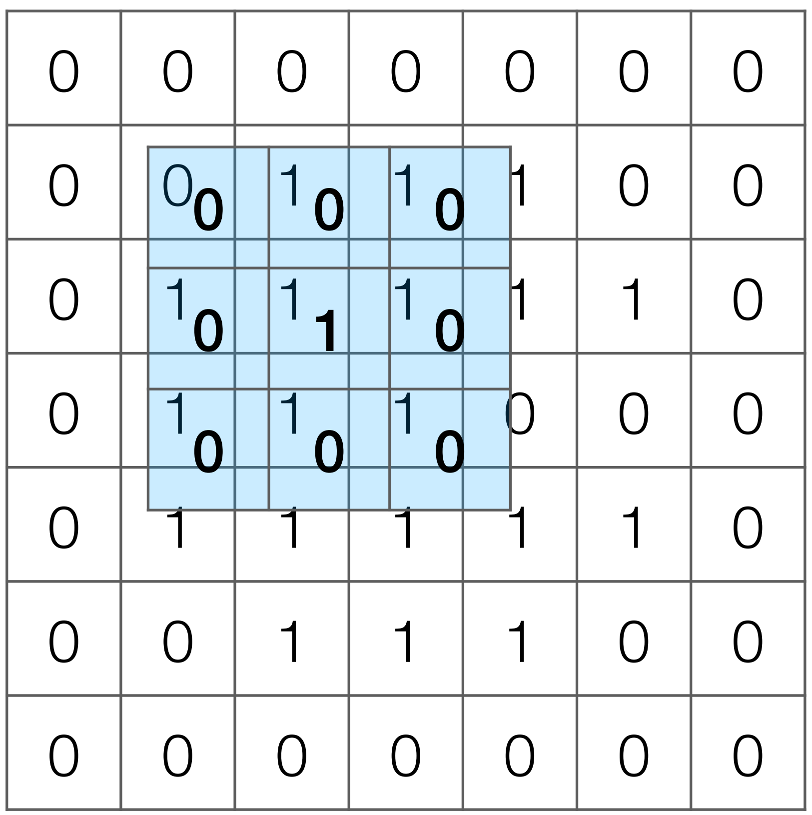
Output image

Convolution



Convolution

Take dot product!



Input image

0	1	1	1	0
1	1			

Output image

Convolution

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Input image

Output image

Note: output image is smaller than input image

If you want output size to be same as input, pad 0's to input

Convolution

$$
\begin{bmatrix}
0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0\\
0&0&1&1&1&1&1&0&0\\
0&0&1&1&1&0&0&0&0\\
0&0&1&1&1&1&1&0&0\\
0&0&0&1&1&1&0&0&0\\
0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0
\end{bmatrix}
*
\begin{bmatrix}
0&0&0\\
0&1&0\\
0&0&0
\end{bmatrix}
=
\begin{bmatrix}
0&0&0&0&0&0&0\\
0&0&1&1&1&0&0\\
0&1&1&1&1&1&0\\
0&1&1&1&0&0&0\\
0&1&1&1&1&1&0\\
0&0&1&1&1&0&0\\
0&0&0&0&0&0&0
\end{bmatrix}
$$

Input image

Output image

Note: output image is smaller than input image

If you want output size to be same as input, pad 0's to input

Convolution

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

Input image

*

0	**0**	**0**
0	**1**	**0**
0	**0**	**0**

=

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

Output image

Convolution

$$\begin{bmatrix} 0&0&0&0&0&0&0\\ 0&0&1&1&1&0&0\\ 0&1&1&1&1&1&0\\ 0&1&1&1&0&0&0\\ 0&1&1&1&1&1&0\\ 0&0&1&1&1&0&0\\ 0&0&0&0&0&0&0 \end{bmatrix} * \frac{1}{9}\begin{bmatrix} 1&1&1\\ 1&1&1\\ 1&1&1 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 3&5&6&5&3\\ 5&8&8&6&3\\ 6&9&8&7&4\\ 5&8&8&6&3\\ 3&5&6&5&3 \end{bmatrix}$$

Input image

Output image

Convolution

$$
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 1 & 1 & 0 \\
0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 1 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
*
\begin{bmatrix}
\mathbf{-1} & \mathbf{-1} & \mathbf{-1} \\
\mathbf{2} & \mathbf{2} & \mathbf{2} \\
\mathbf{-1} & \mathbf{-1} & \mathbf{-1}
\end{bmatrix}
=
\begin{bmatrix}
0 & 1 & 3 & 1 & 0 \\
1 & 1 & 1 & 3 & 3 \\
0 & 0 & -2 & -4 & -4 \\
1 & 1 & 1 & 3 & 3 \\
0 & 1 & 3 & 1 & 0
\end{bmatrix}
$$

Input image

Output image

Convolution

Very commonly used for:

- Blurring an image

*

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

=

- Finding edges

*

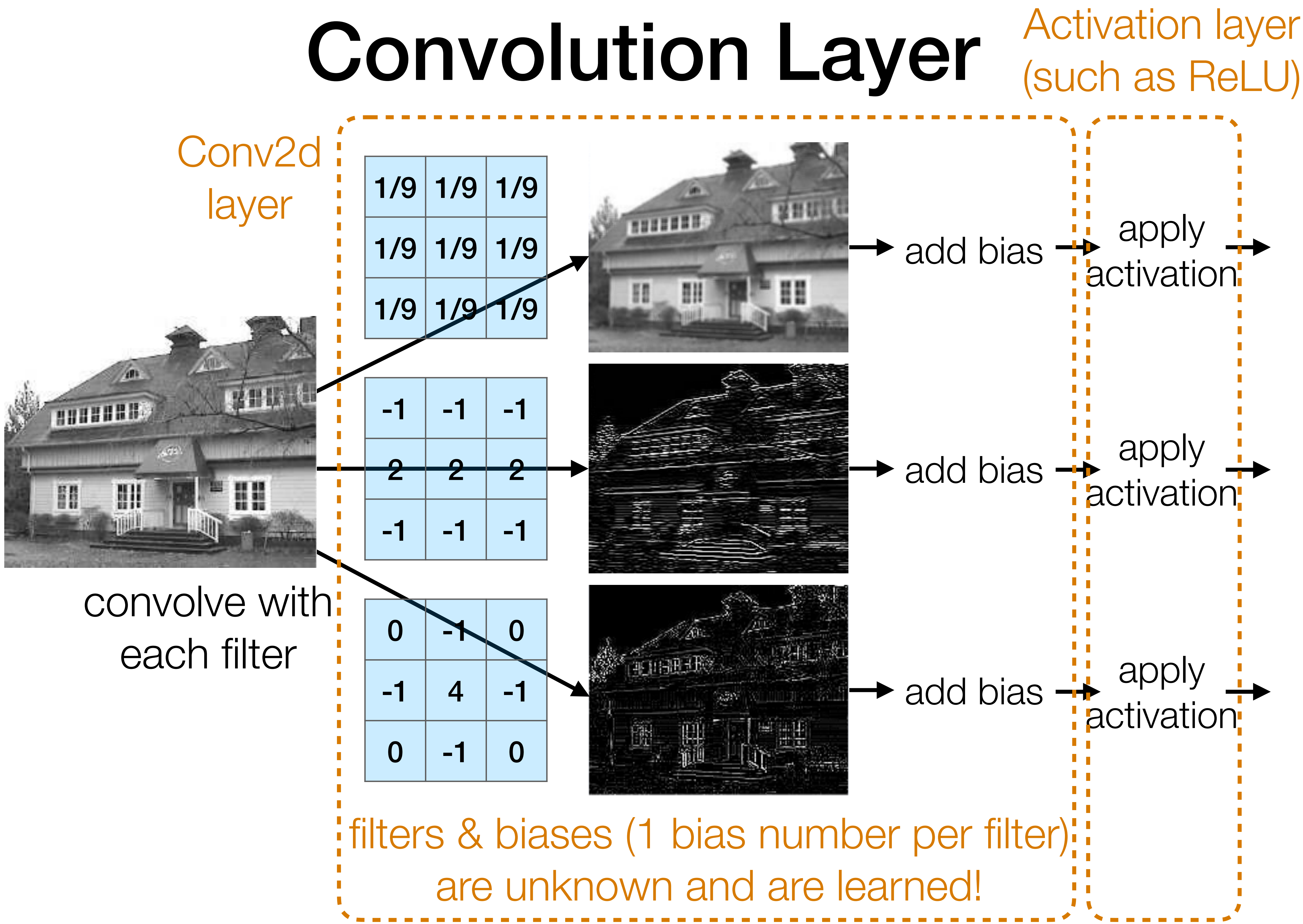
-1	-1	-1
2	2	2
-1	-1	-1

=

(this example finds horizontal edges)

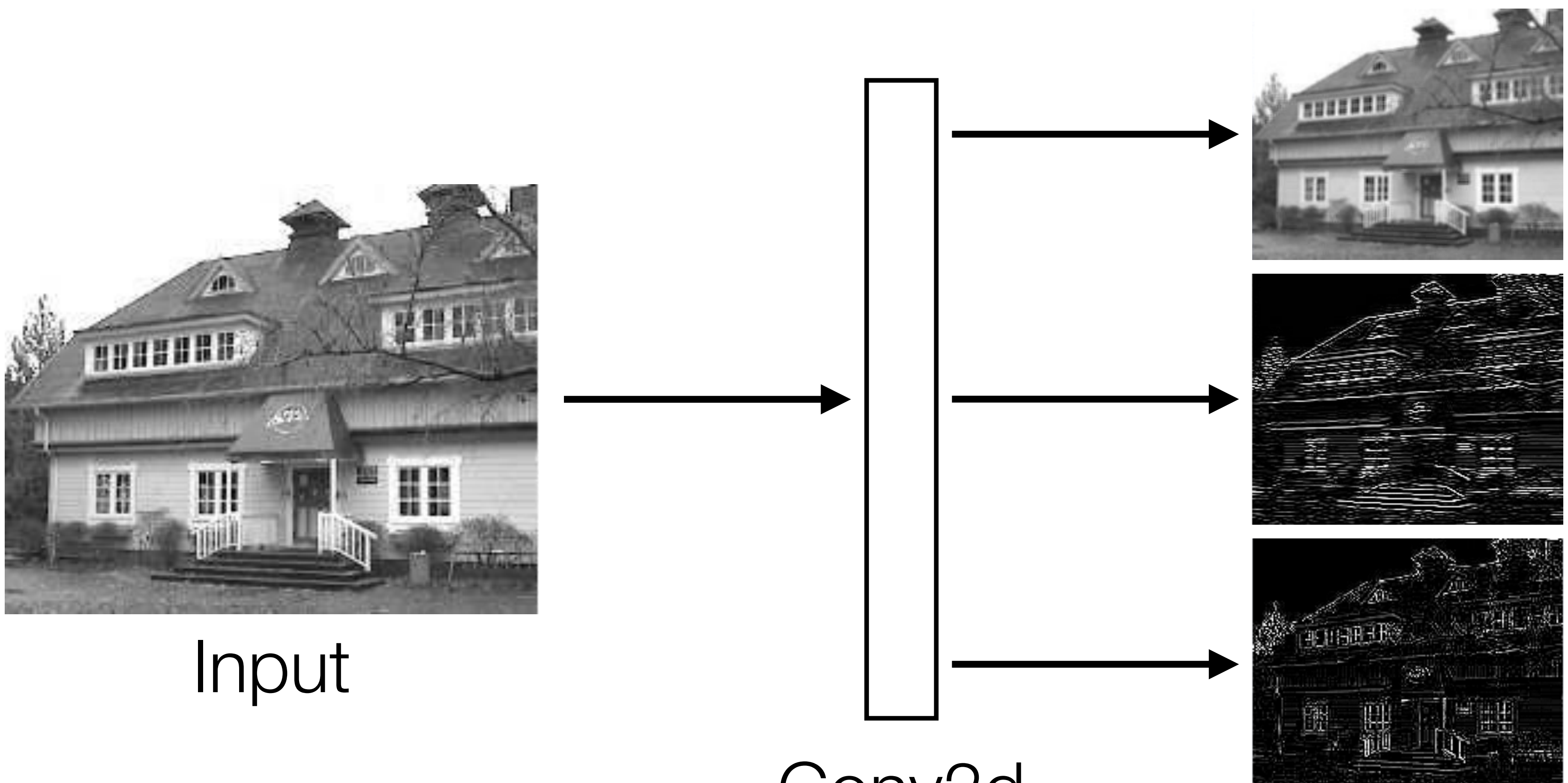
Images from: http://aishack.in/tutorials/image-convolution-examples/

Convolution Layer



Images from: http://aishack.in/tutorials/image-convolution-examples/

Convolution Layer



Images from: http://aishack.in/tutorials/image-convolution-examples/

Convolution Layer

Input

dimensions:
1 (# channels),
height,
width

Conv2d
(3 kernels,
each size 3x3),
ReLU activation

Stack output images into a single "output feature map"

dimensions:
3,
height-2,
width-2

Images from: http://aishack.in/tutorials/image-convolution-examples/

Convolution Layer

Input

dimensions:
1 (# channels),
height,
width

Conv2d
(*k* kernels
each size 3x3),
ReLU activation

Stack output images into a single "output feature map"

dimensions:
k,
height-2,
width-2

Images from: http://aishack.in/tutorials/image-convolution-examples/

Convolution Layer

Input

dimensions:
d (# channels)
height,
width

Conv2d
(*k* kernels
each size *d*x3x3),
ReLU activation

Stack output images into a single "output feature map"

dimensions:
k,
height-2,
width-2

Images from: http://aishack.in/tutorials/image-convolution-examples/

Convolution Layer

Input

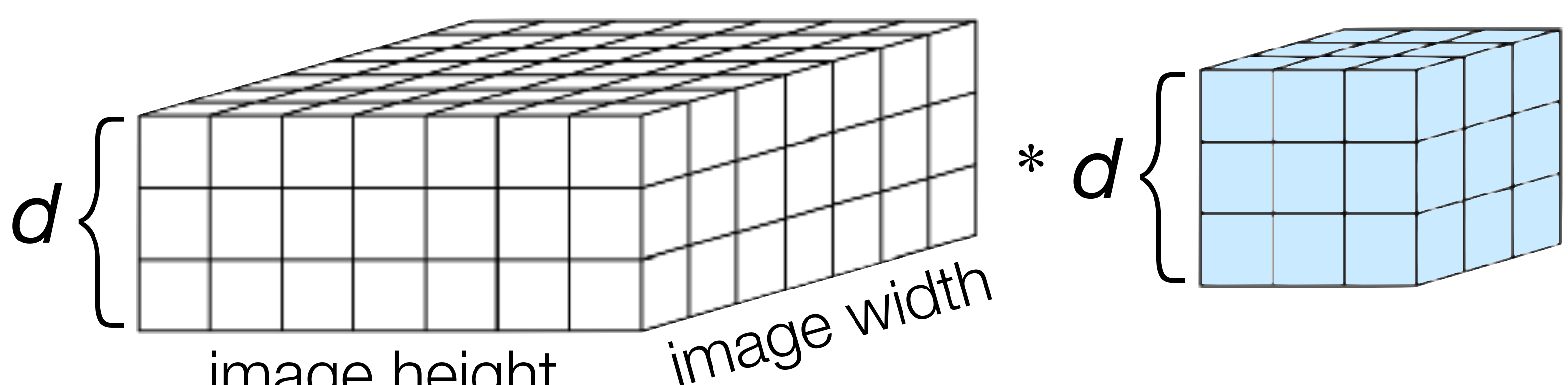
dimensions:
d (# channels)
height,
width

Conv2d
(*k* kernels
each size *d*x3x3),
ReLU activation

Stack output images into a single "output feature map"

dimensions:
k,
height-2,
width-2

Each filter:



Pooling

- Aggregate local information ("pool" together information)

- Produces a smaller image
  (each resulting pixel captures some "global" information)

- If "object" in input image shifts, want output to stay the same

Max Pooling

Convolution layer (1 filter, for simplicity no bias)

$$\underset{\text{Input}}{\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}} * \begin{bmatrix} \mathbf{-1} & \mathbf{-1} & \mathbf{-1} \\ \mathbf{2} & \mathbf{2} & \mathbf{2} \\ \mathbf{-1} & \mathbf{-1} & \mathbf{-1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 3 & 1 & 0 \\ 1 & 1 & 1 & 3 & 3 \\ 0 & 0 & -2 & -4 & -4 \\ 1 & 1 & 1 & 3 & 3 \\ 0 & 1 & 3 & 1 & 0 \end{bmatrix}$$

Max Pooling

Convolution layer (1 filter, for simplicity no bias)

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

Input

*

-1	**-1**	**-1**
2	**2**	**2**
-1	**-1**	**-1**

=

0	1	3	1	0
1	1	1	3	3
0	0	-2	-4	-4
1	1	1	3	3
0	1	3	1	0

0	1	3	1	0
1	1	1	3	3
0	0	0	0	0
1	1	1	3	3
0	1	3	1	0

Output image after ReLU

Output after max pooling

Max Pooling

Convolution layer (1 filter, for simplicity no bias)

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

Input

*

-1	**-1**	**-1**
2	**2**	**2**
-1	**-1**	**-1**

=

0	1	3	1	0
1	1	1	3	3
0	0	-2	-4	-4
1	1	1	3	3
0	1	3	1	0

0	1	3	1	0
1	1	1	3	3
0	0	0	0	0
1	1	1	3	3
0	1	3	1	0

Output image after ReLU

1	

Output after max pooling

Max Pooling

Convolution layer (1 filter, for simplicity no bias)

$$
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 1 & 1 & 0 \\
0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 1 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
*
\begin{bmatrix}
\mathbf{-1} & \mathbf{-1} & \mathbf{-1} \\
\mathbf{2} & \mathbf{2} & \mathbf{2} \\
\mathbf{-1} & \mathbf{-1} & \mathbf{-1}
\end{bmatrix}
=
\begin{bmatrix}
0 & 1 & 3 & 1 & 0 \\
1 & 1 & 1 & 3 & 3 \\
0 & 0 & -2 & -4 & -4 \\
1 & 1 & 1 & 3 & 3 \\
0 & 1 & 3 & 1 & 0
\end{bmatrix}
$$

Input

$$
\begin{bmatrix}
0 & 1 & 3 & 1 & 0 \\
1 & 1 & 1 & 3 & 3 \\
0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 3 & 3 \\
0 & 1 & 3 & 1 & 0
\end{bmatrix}
$$

Output image after ReLU

1	3

Output after max pooling

Max Pooling

Convolution layer (1 filter, for simplicity no bias)

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

Input

$$* \begin{bmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 3 & 1 & 0 \\ 1 & 1 & 1 & 3 & 3 \\ 0 & 0 & -2 & -4 & -4 \\ 1 & 1 & 1 & 3 & 3 \\ 0 & 1 & 3 & 1 & 0 \end{bmatrix}$$

0	1	3	1	0
1	1	1	3	3
0	0	0	0	0
1	1	1	3	3
0	1	3	1	0

Output image after ReLU

1	3
1	

Output after max pooling

Max Pooling

Convolution layer (1 filter, for simplicity no bias)

Input

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

*

-1	**-1**	**-1**
2	**2**	**2**
-1	**-1**	**-1**

=

0	1	3	1	0
1	1	1	3	3
0	0	-2	-4	-4
1	1	1	3	3
0	1	3	1	0

0	1	3	1	0
1	1	1	3	3
0	0	0	0	0
1	1	1	3	3
0	1	3	1	0

Output image after ReLU

1	3
1	3

Output after max pooling

Max Pooling

Convolution layer (1 filter, for simplicity no bias)

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

Input

*

-1	**-1**	**-1**
2	**2**	**2**
-1	**-1**	**-1**

=

0	1	3	1	0
1	1	1	3	3
0	0	-2	-4	-4
1	1	1	3	3
0	1	3	1	0

0	1	3	1	0
1	1	1	3	3
0	0	0	0	0
1	1	1	3	3
0	1	3	1	0

Output image after ReLU

What numbers were involved in computing this 1?

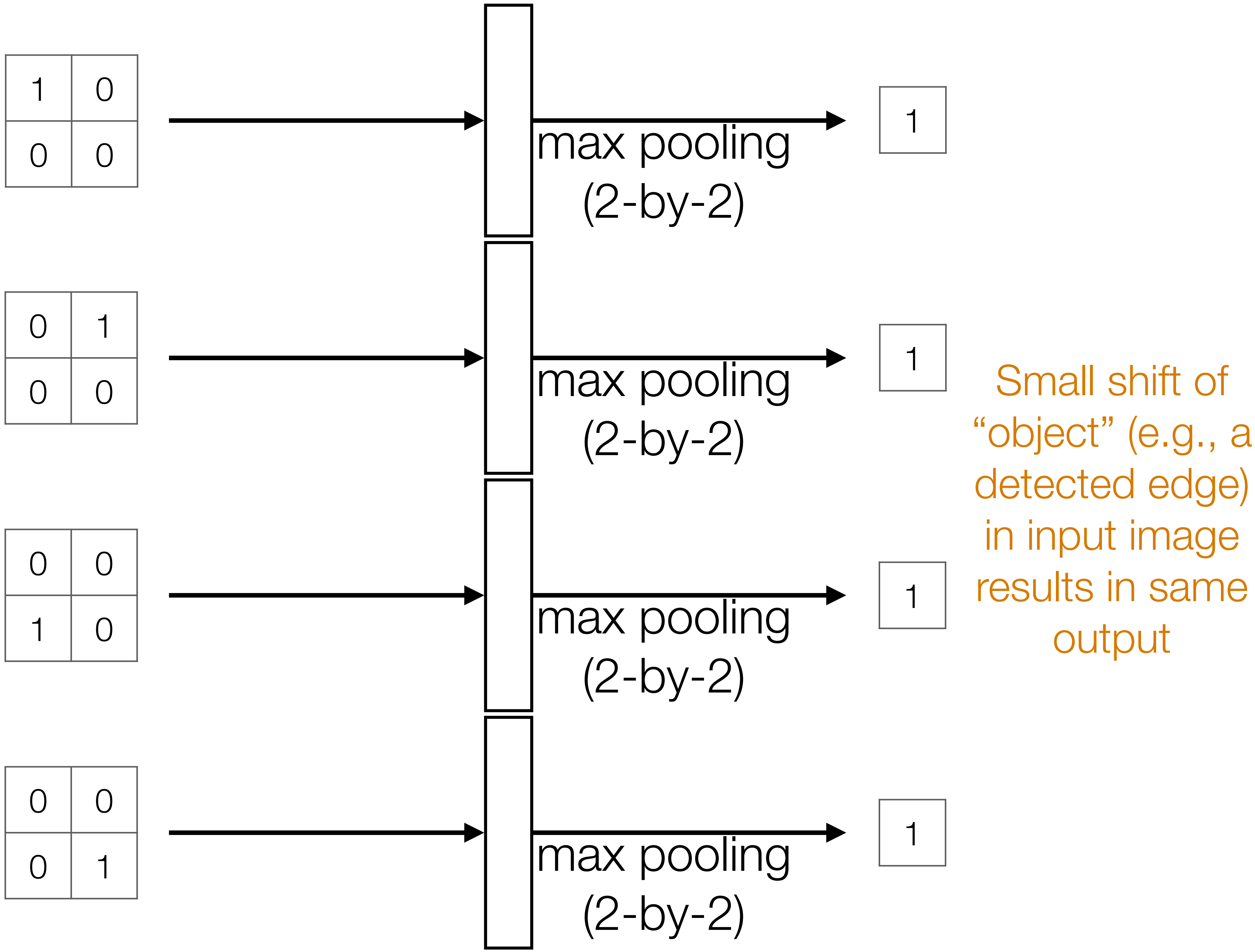
In this example: 1 pixel in max pooling output captures information from 16 input pixels!

Example: applying max pooling again results in a single pixel that captures info from entire input image!

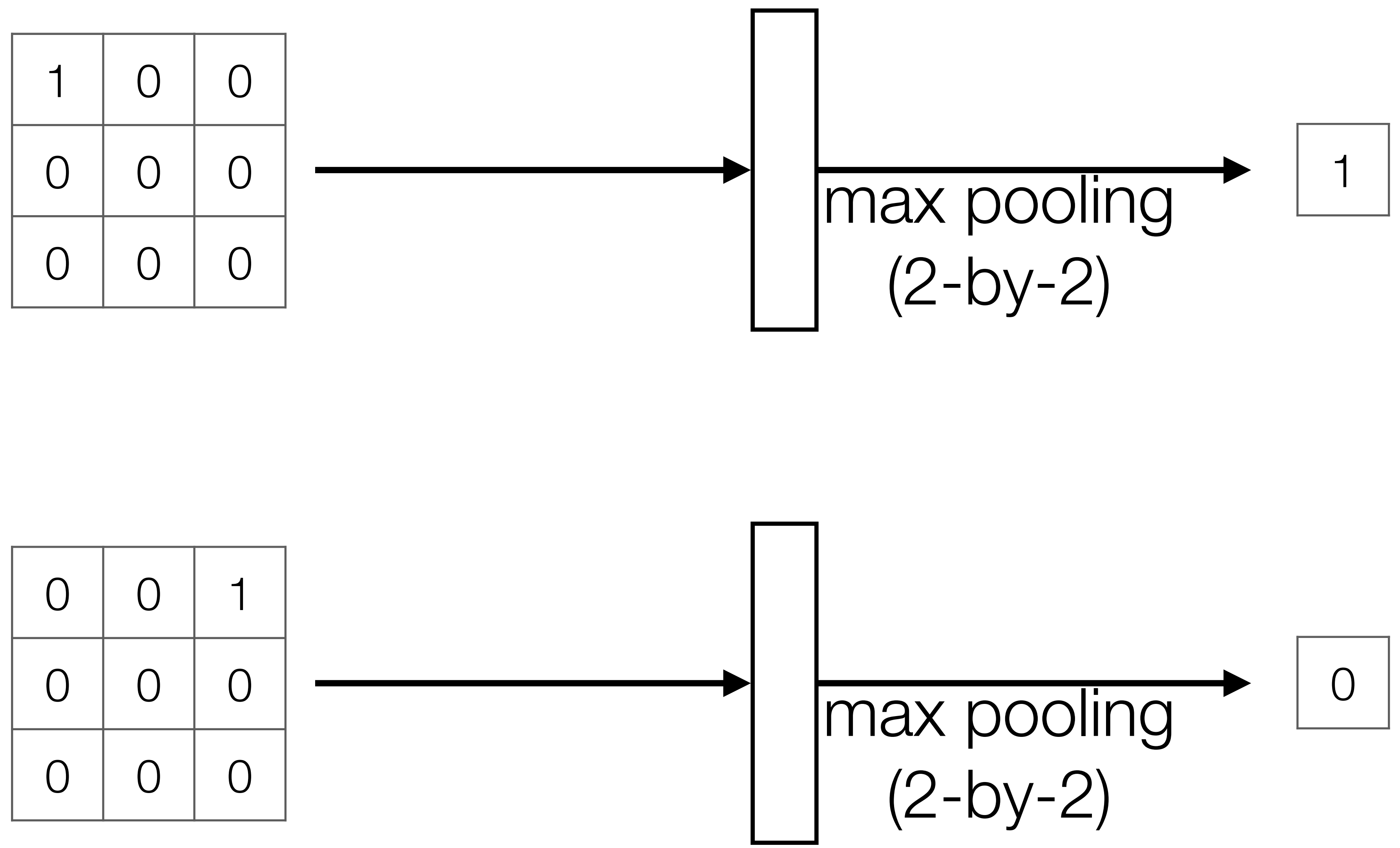
1	3
1	3

Output after max pooling

Max Pooling and (Slight) Shift Invariance

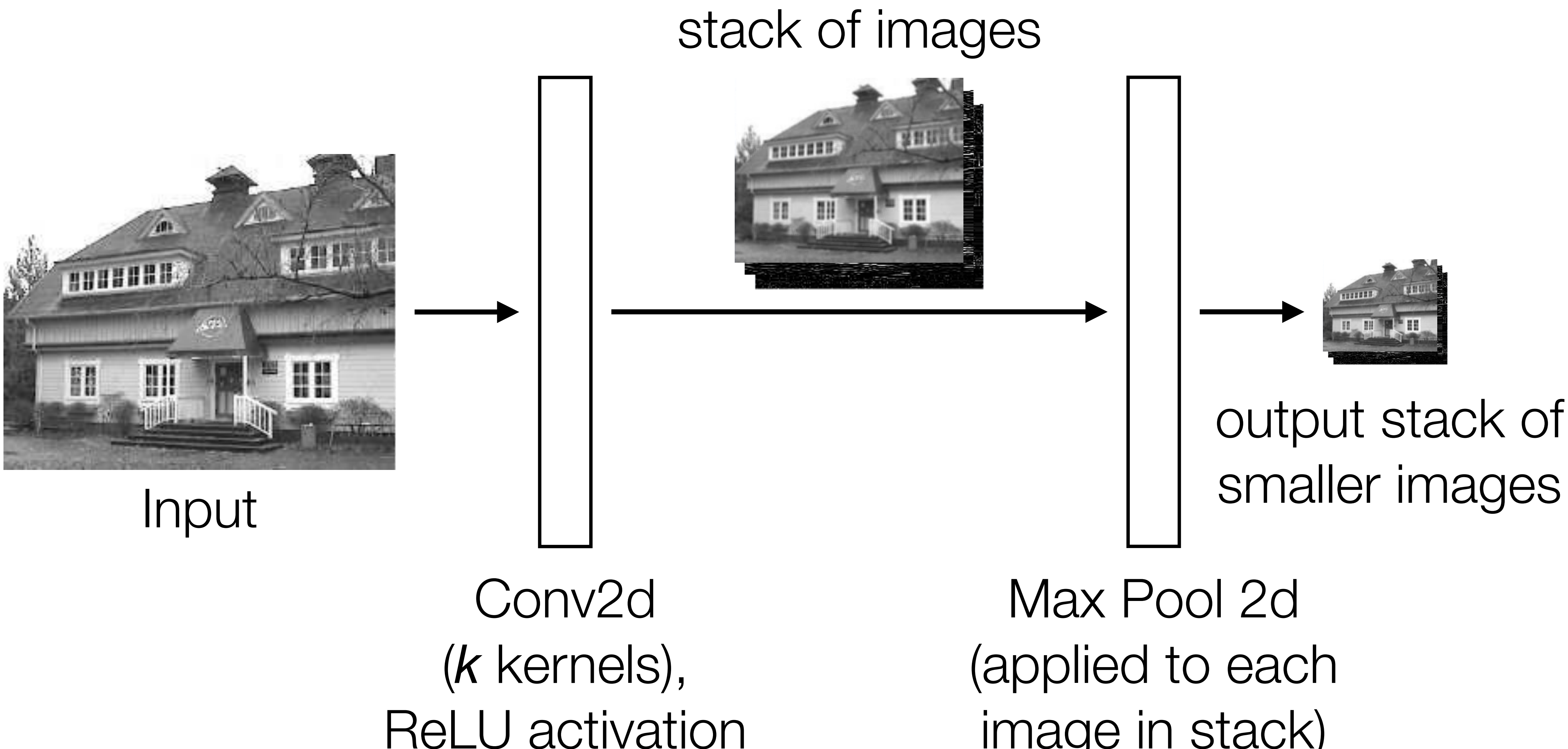


Max Pooling and (Slight) Shift Invariance



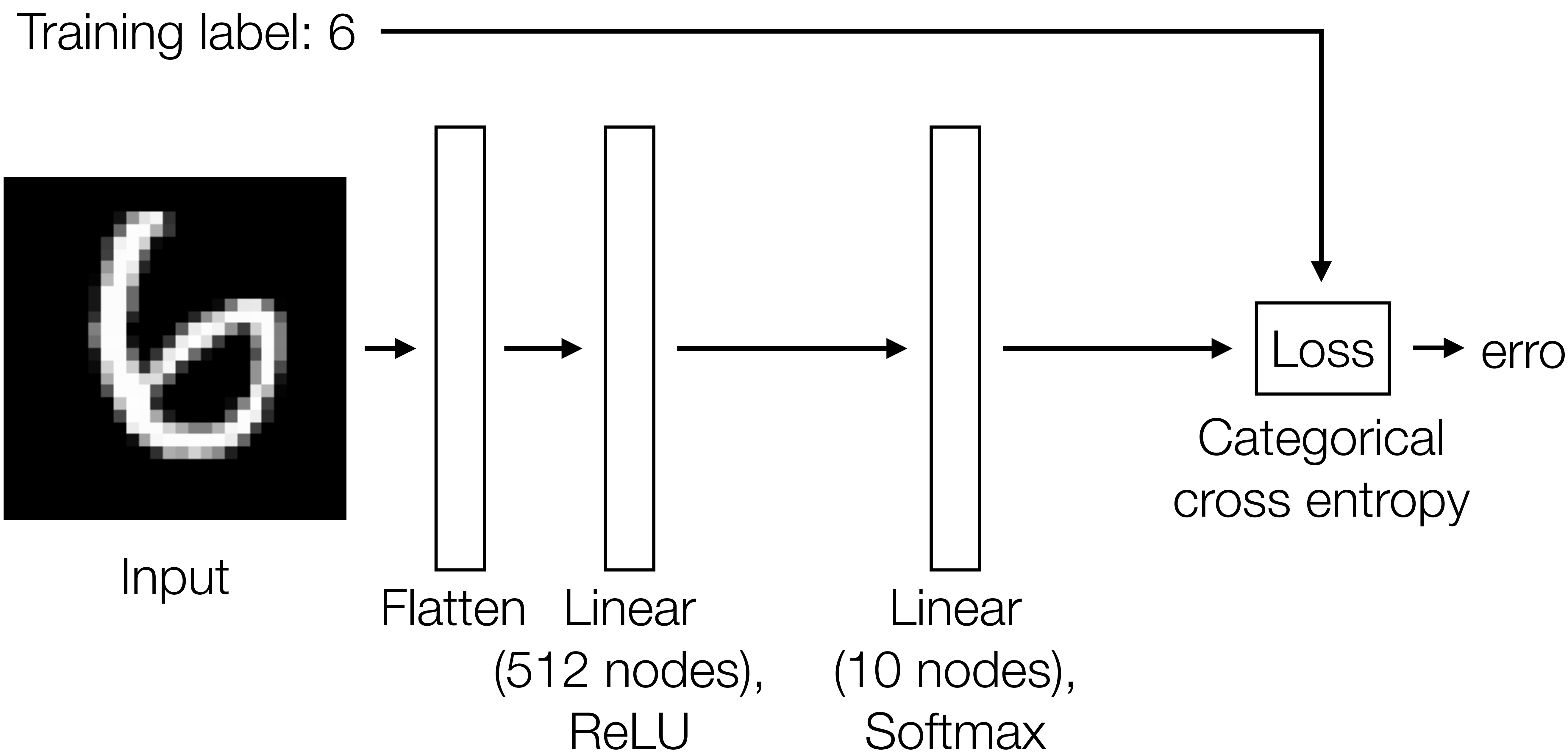
Bigger shift in input can still change output

Basic Building Block of CNNs



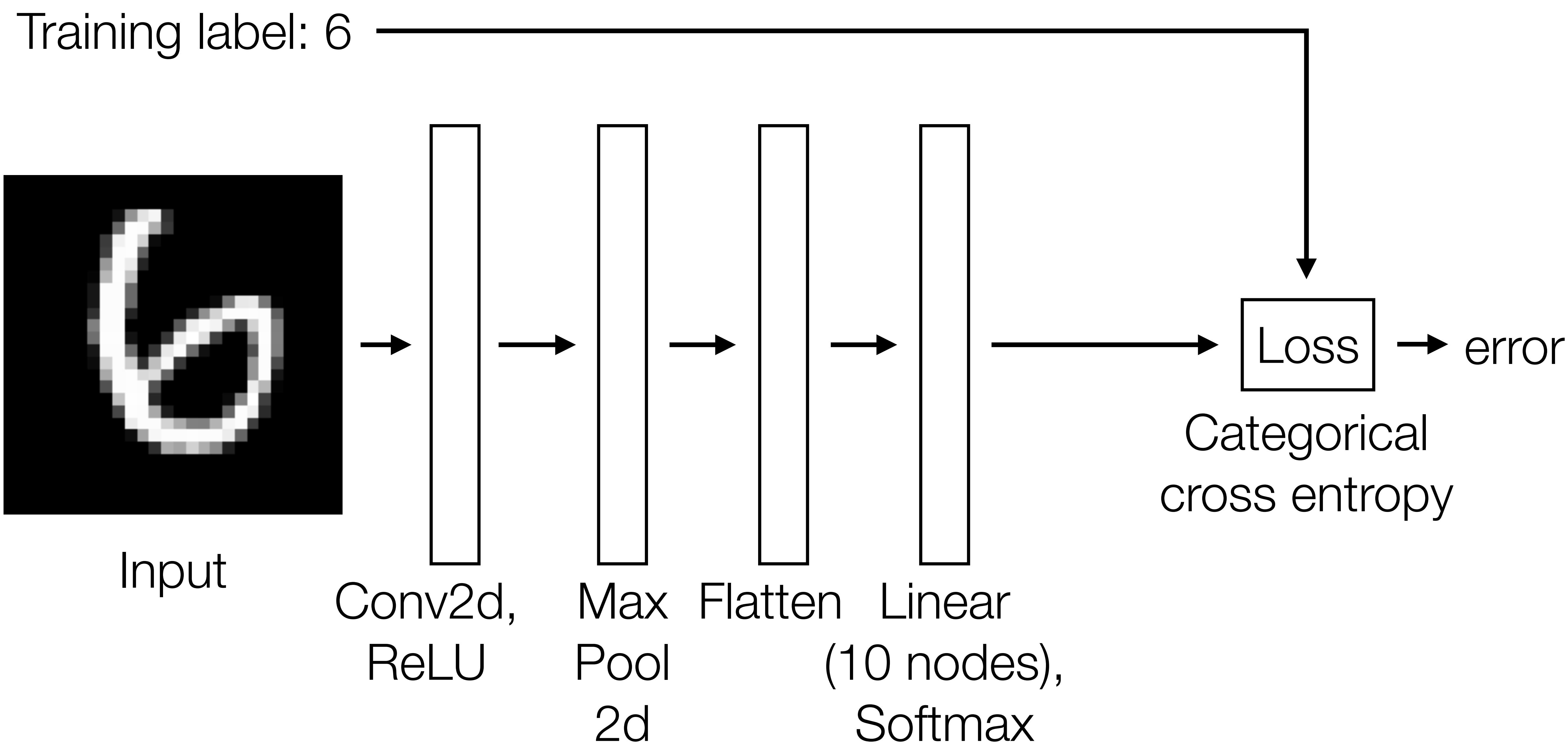
Images from: http://aishack.in/tutorials/image-convolution-examples/

Handwritten Digit Recognition

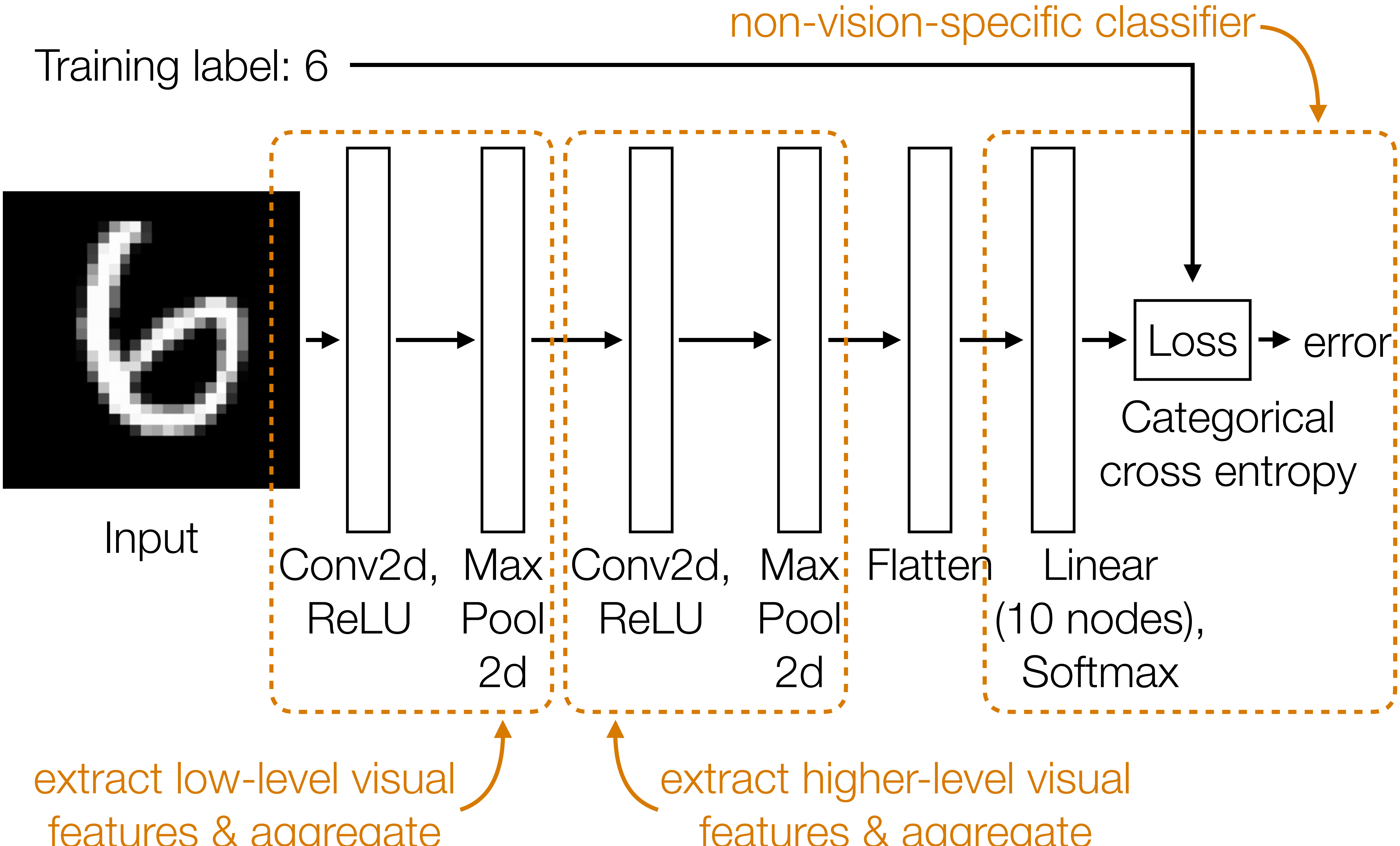


Important: in lecture, I will some times use this notation instead

Handwritten Digit Recognition



Handwritten Digit Recognition



CNNs

Demo

CNNs

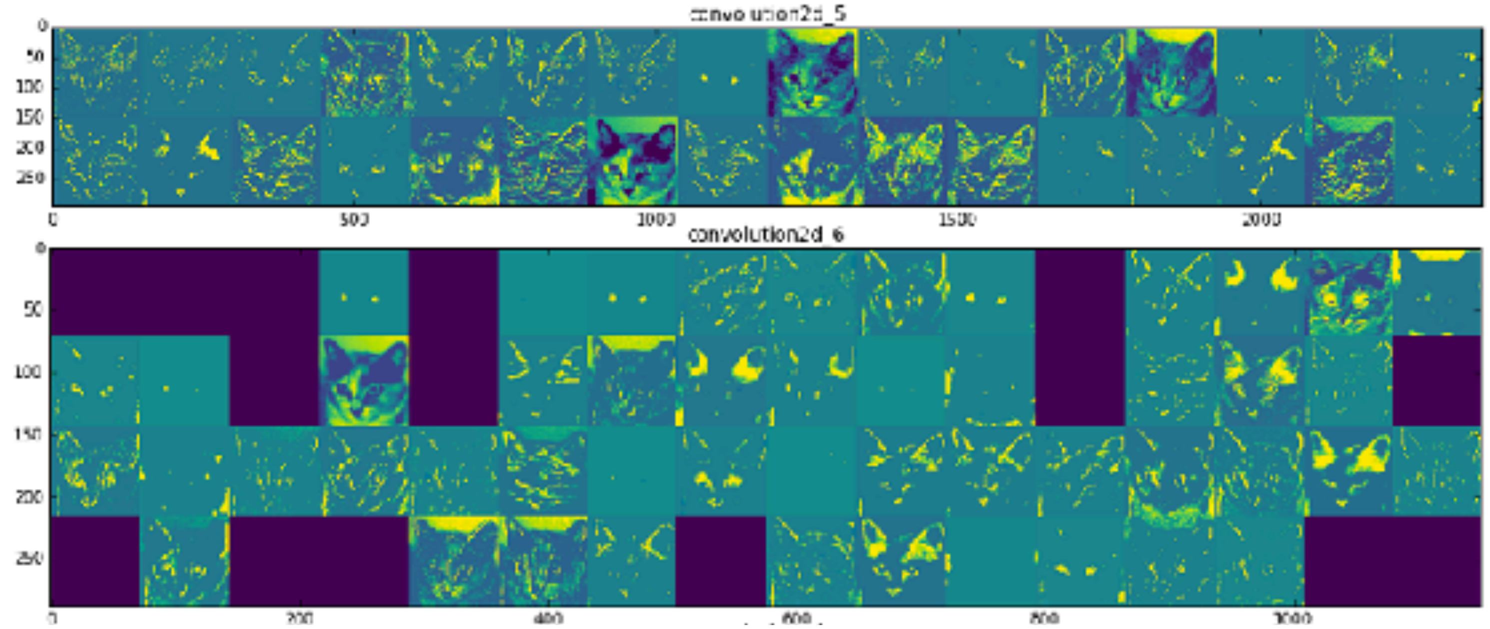
- Learn convolution filters for extracting simple features

- Max pooling produces a *smaller* summary output and is somewhat invariant to small shifts in input "objects"
  - For examples where max pooling fails to achieve this and for a better way to do pooling, see Richard Zhang's fix for max pooling linked on the course webpage

- Repeat convolution→activation→pooling to learn increasingly higher-level features

Visualizing What a CNN Learned

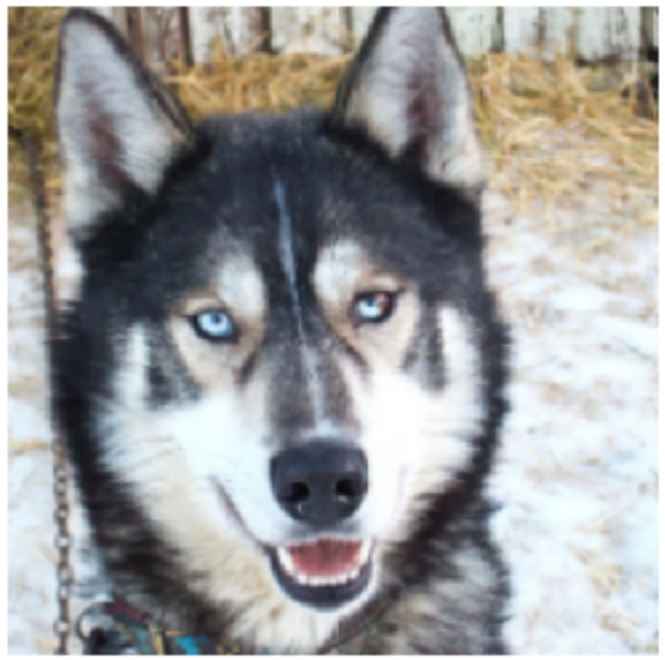
- Plot filter outputs at different layers



- Plot regions that maximally activate an output neuron

Images: Francois Chollet's "Deep Learning with Python" Chapter 5

Example: Wolves vs Huskies



(a) Husky classified as wolf

(b) Explanation

Turns out the deep net learned that wolves are wolves because of snow…

➔ visualization is crucial!

Source: Ribeiro et al. "Why should I trust you? Explaining the predictions of any classifier." KDD 2016.