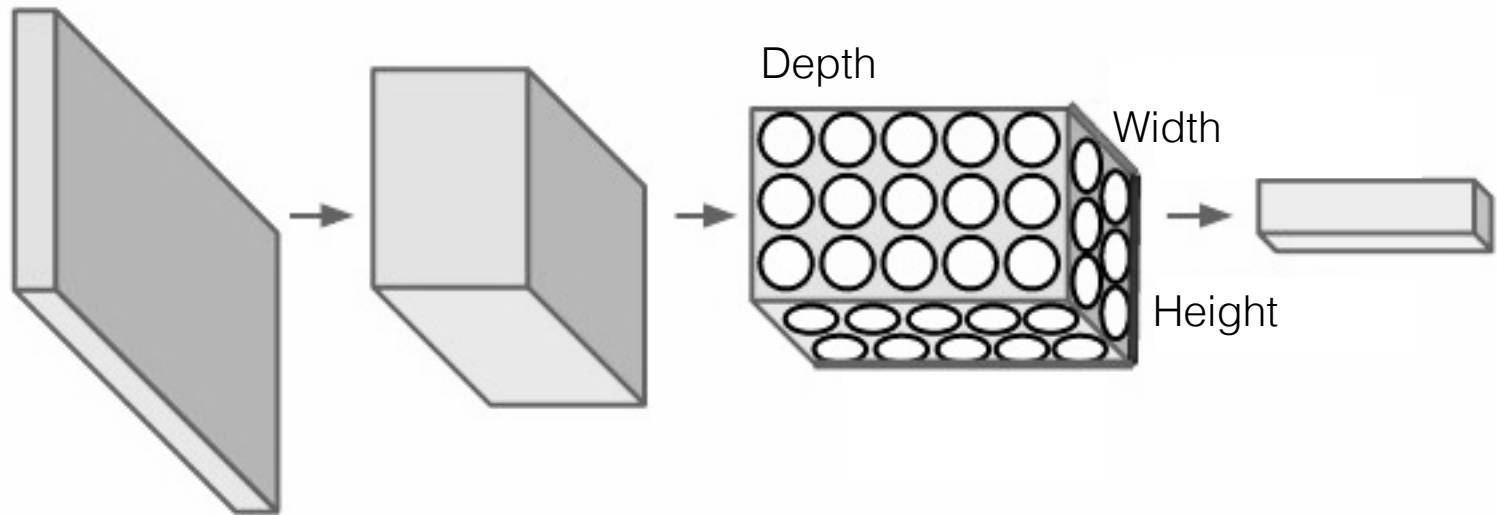


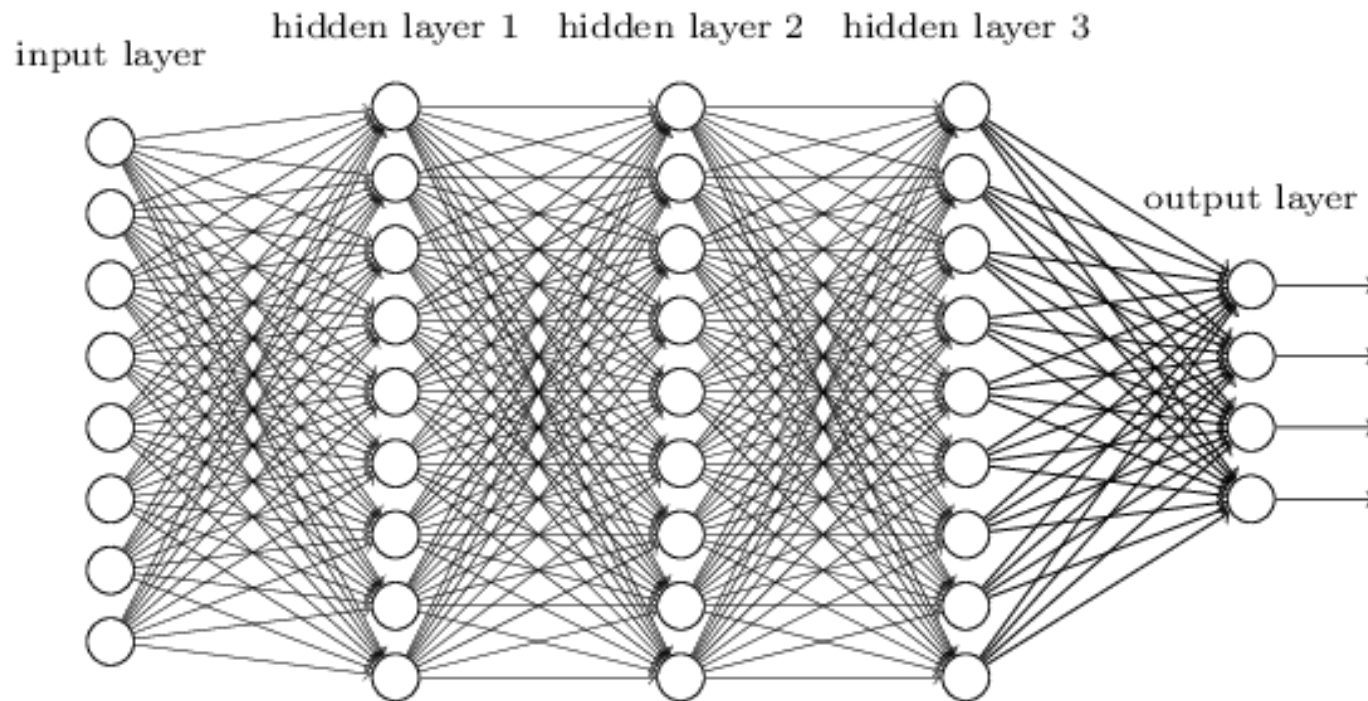
Convolutional Neural Networks



What Is It?

Convolutional Neural Networks or CNN is a class of deep, feed-forward artificial neural networks that are applied to analyzing visual imagery.

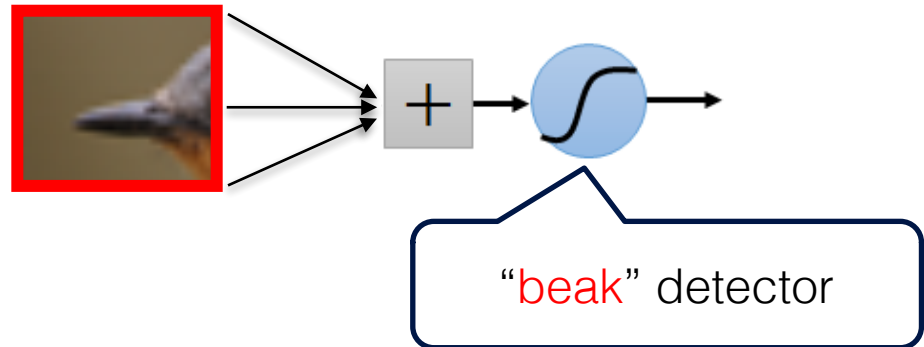
Neural Networks



- We try to reduce the number of parameters from a model for ease of processing.
- From this fully connected model, do we really need all the edges?
- Can some of these be shared?

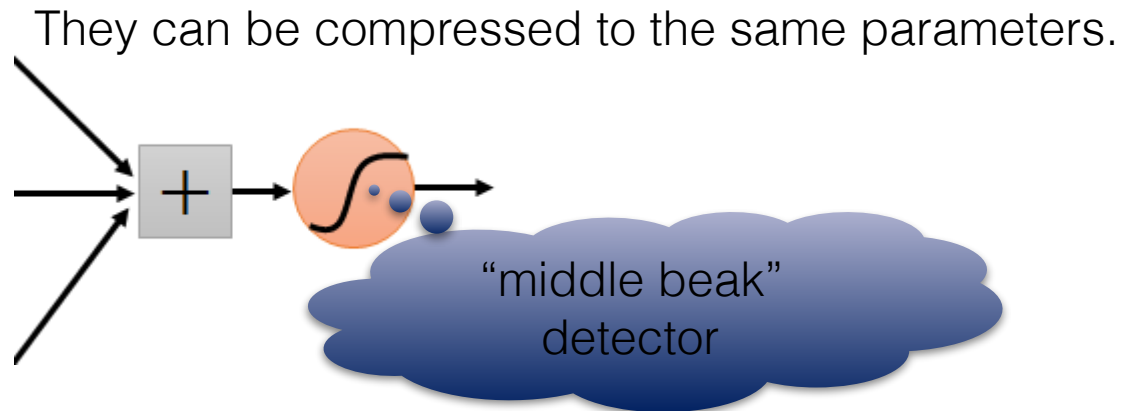
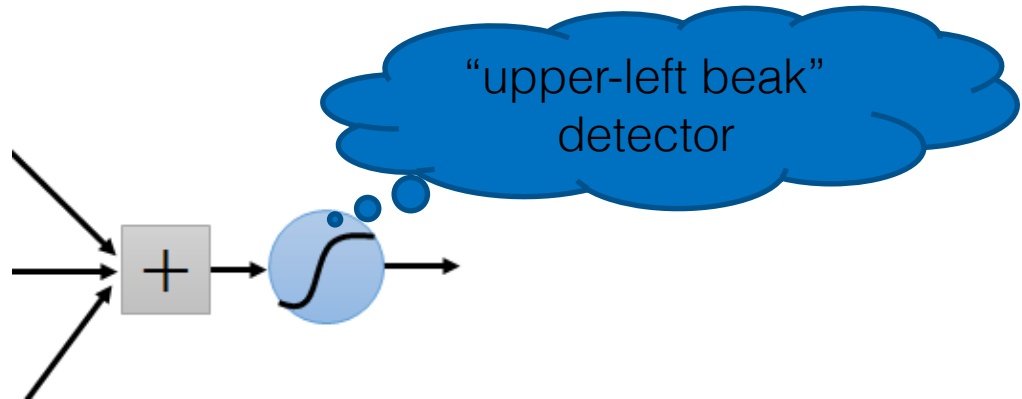
Learning an image

Some patterns are much smaller than the whole image
Can represent a small region with fewer parameters



Learning an image

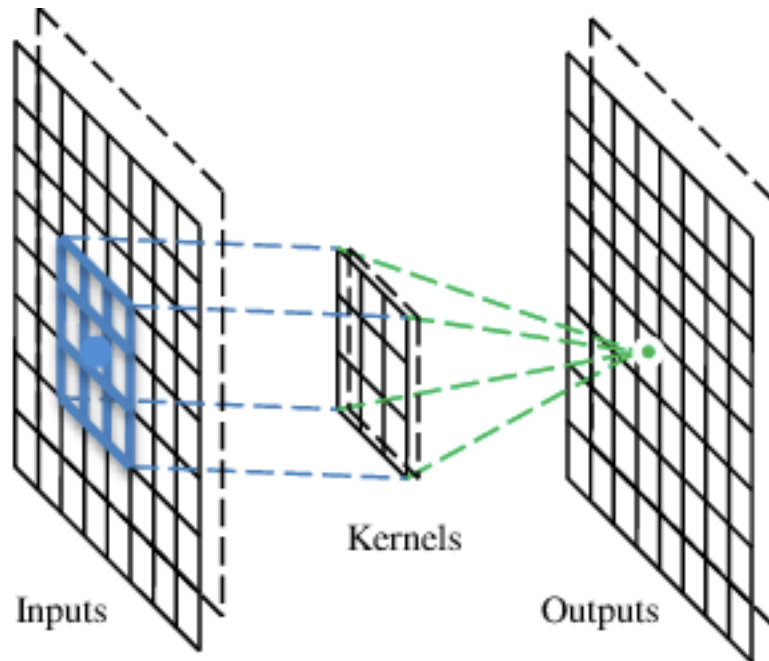
Same pattern appears in different places: They can be compressed!
What about training a lot of such “small” detectors and each detector must “move around”.



A convolutional layer

A CNN is a neural network with some convolutional layers (and some other layers).

A convolutional layer has a number of filters that does convolutional operation.



Convolution operation

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

6 x 6 image

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

-1	1	-1
-1	1	-1
-1	1	-1

Filter 2

⋮ ⋮

Each filter detects a small pattern (3 x 3).
Hyper parameters

Convolution operation

stride=1

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

Dot product

3 -1

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

6 x 6 image

Convolution operation

stride=2

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

Dot product

3 -3

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

6 x 6 image

Convolution operation

stride = 1

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

6 x 6 image

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

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

Convolution operation

stride=1

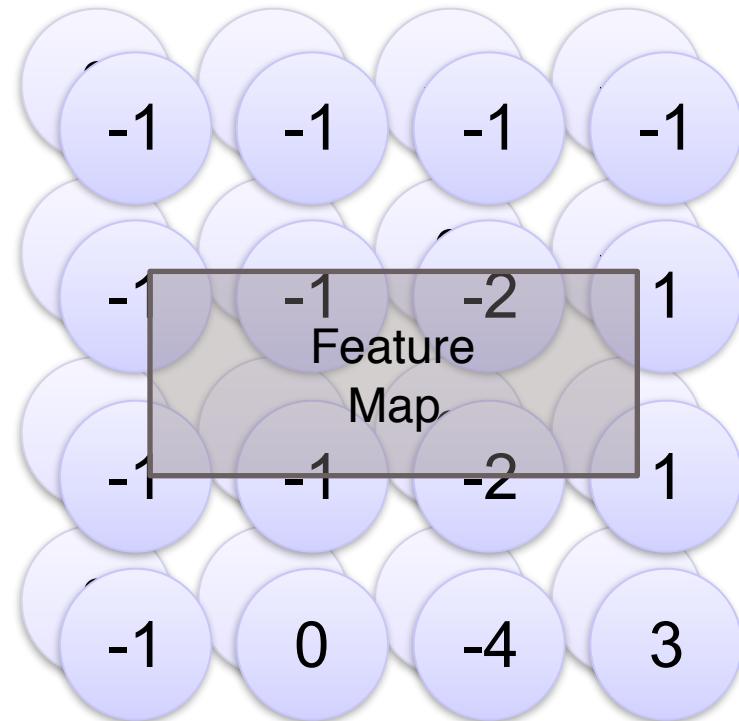
1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image

-1	1	-1
-1	1	-1
-1	1	-1

Filter 2

Repeat this for each filter



100

The image displays three overlapping windows, each showing a different color channel of a grayscale image of a dog. The top window shows the red channel, the middle window shows the green channel, and the bottom window shows the blue channel. Each window has a title bar and standard window controls (minimize, maximize, close).

A diagram showing a 3x3 grid of values. The grid is composed of three overlapping 3x3 sub-grids, each with a different background color (light blue, light green, and light red). The values in the grid are as follows:

1	-1	-1
-1	1	-1
-1	-1	1



The diagram shows a 6x6 grid of cells. Each cell contains either a blue '1' or a black '0'. The grid is as follows:

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

On the left side of the grid, there is a vertical stack of 6 cards. Each card has a black '0' on its face. The cards are stacked such that the top card is slightly offset to the left, and the bottom card is slightly offset to the right, creating a 3D effect. The cards are positioned to the left of the first column of the grid.

Convolution vs Fully Connected

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

image

1	-1	-1
-1	1	-1
-1	-1	1

-1	1	-1
-1	1	-1
-1	1	-1

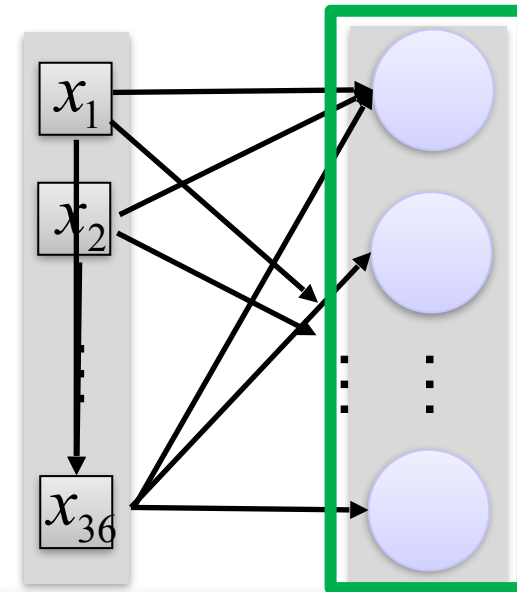


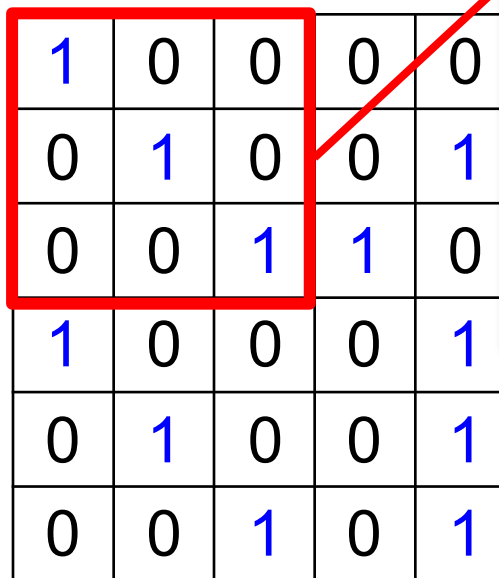
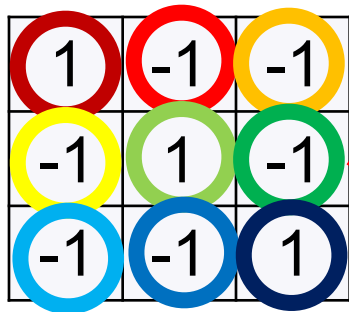
convolution

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

Fully-connected

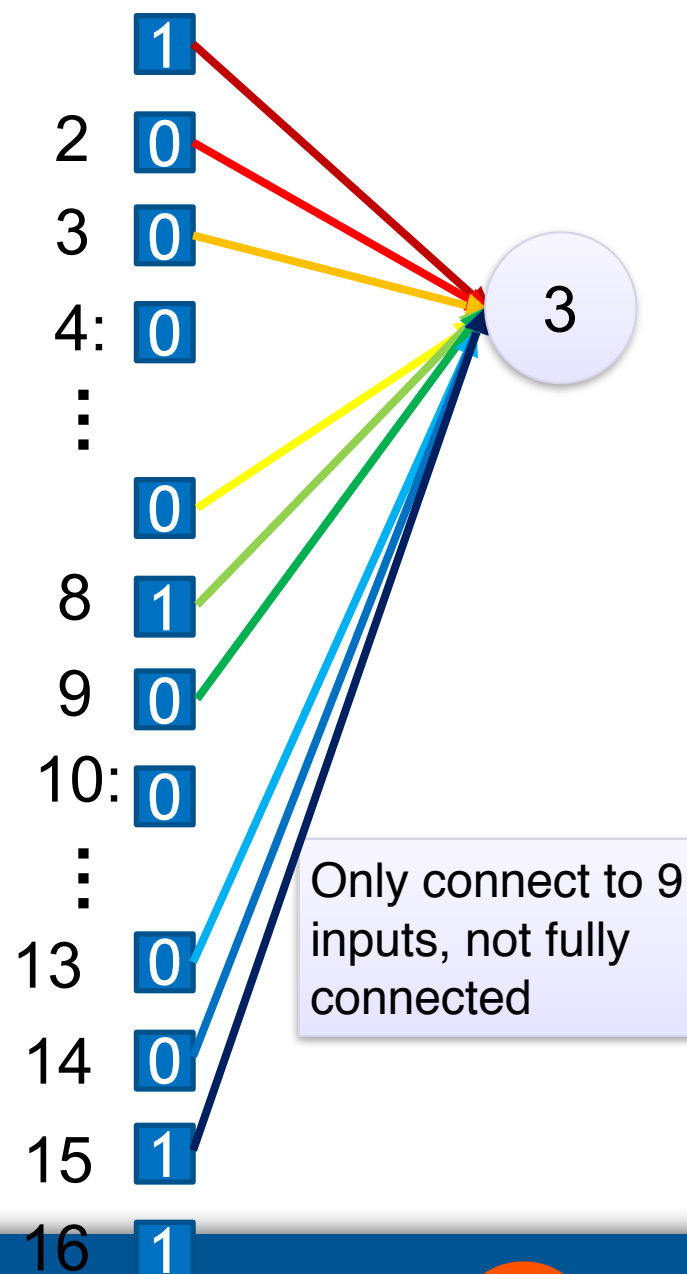
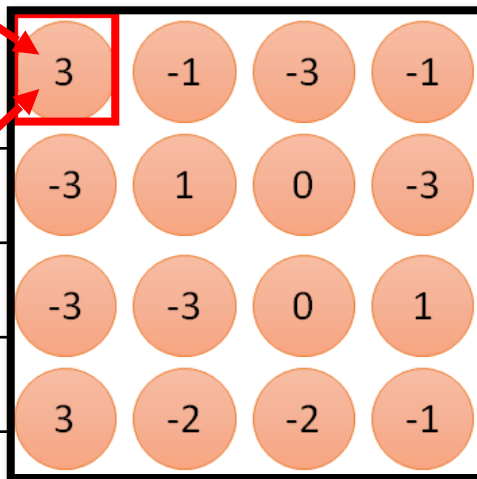
1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0



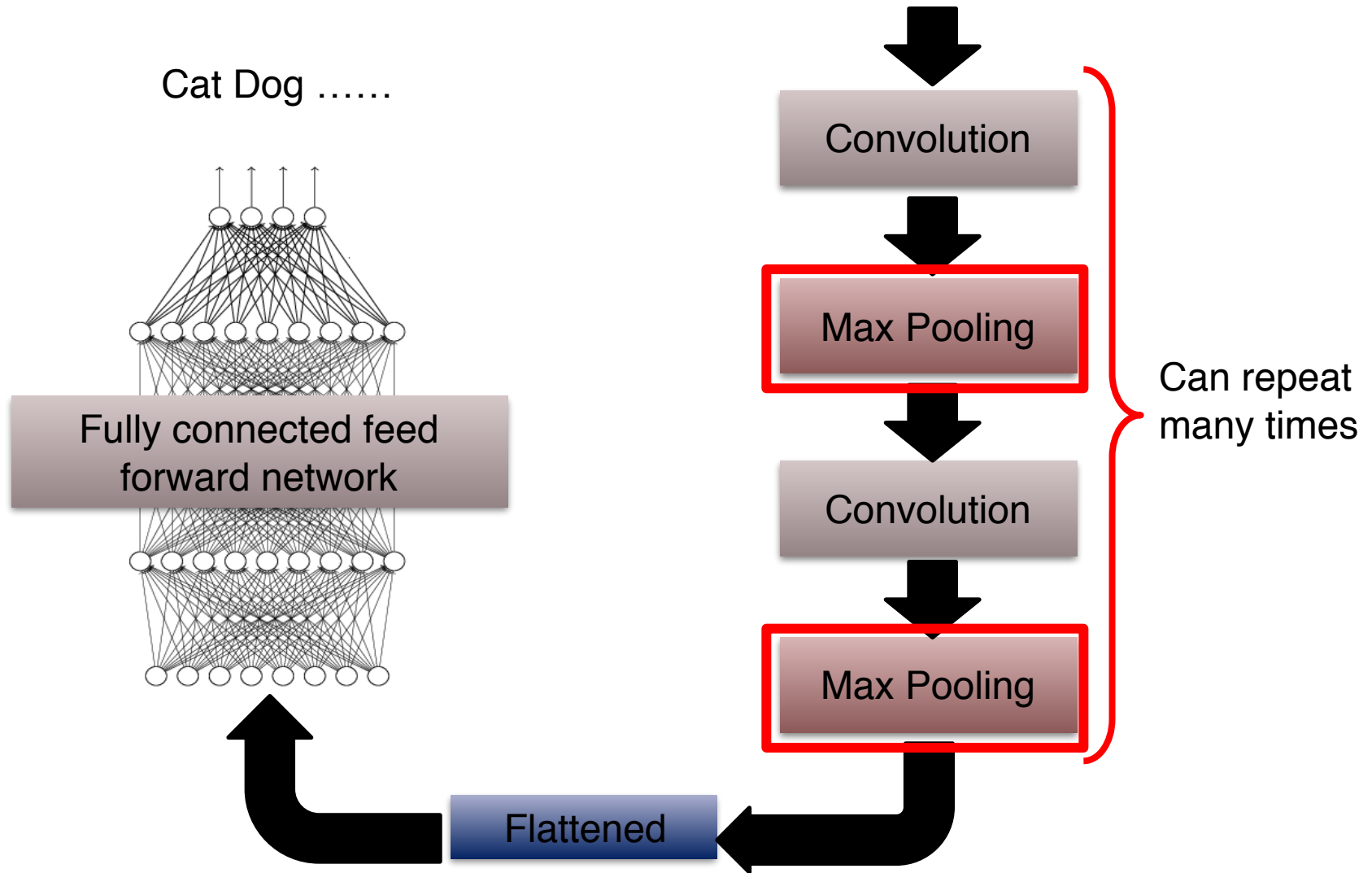


6 x 6 image

fewer parameters!



The CNN



Pooling Layer

A pooling layer is another building block of a CNN.

Its function is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network.

Pooling layer operates on each feature map independently.

The most common approach used in pooling is max pooling.

Max Pooling

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

-1	1	-1
-1	1	-1
-1	1	-1

Filter 2

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

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

Reasons for pooling

Subsampling pixels will not change the object

bird



Subsampling

bird



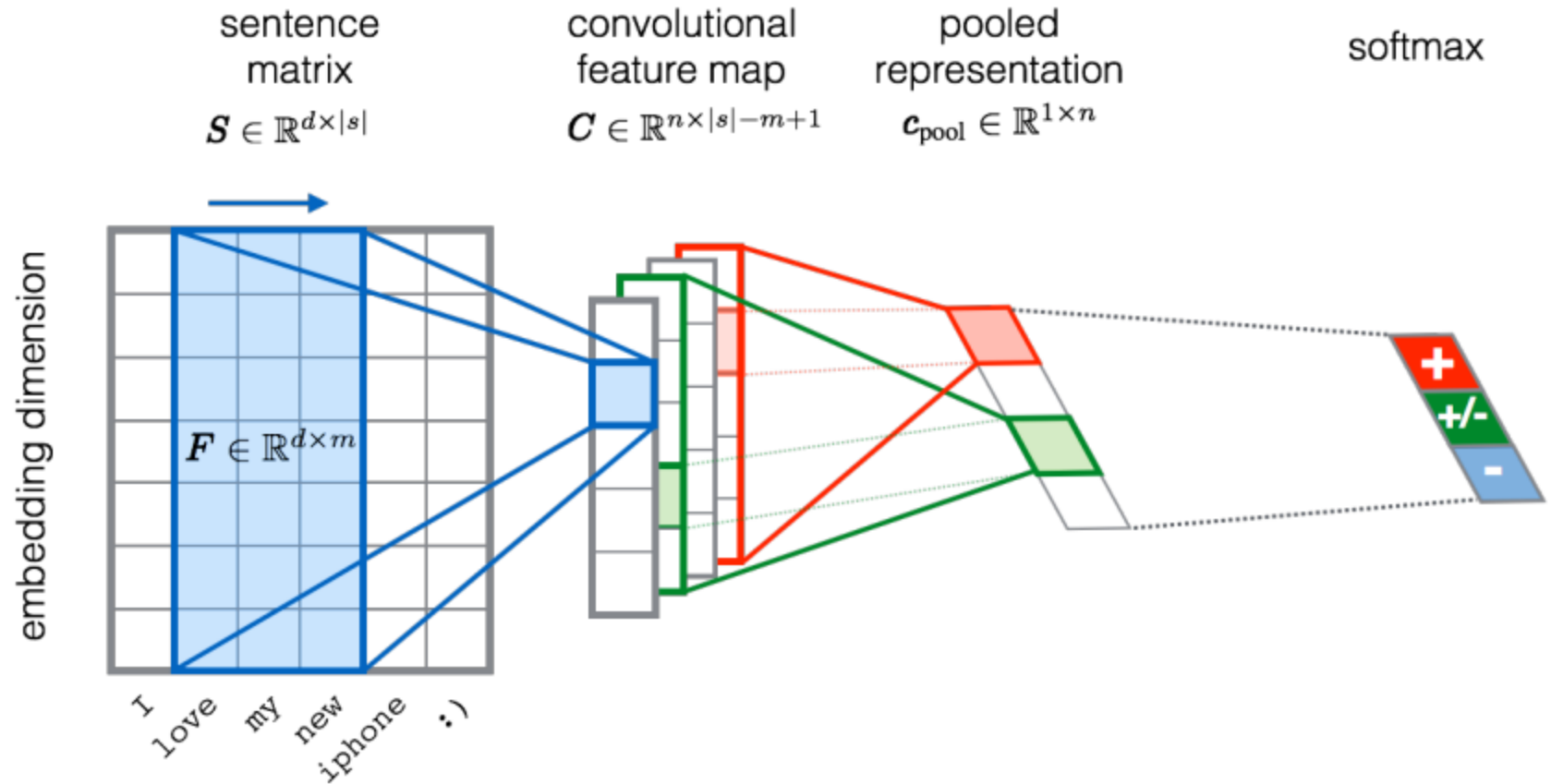
We can subsample the pixels to make image smaller
fewer parameters to characterize the image

Learning an image

A CNN compresses a fully connected network in two ways:

- Reducing number of connections
- Shared weights on the edges
- Max pooling further reduces the complexity

CNN in text classification



CNN in speech recognition

