Convolutional Neural Network (CNN)

卷积神经网络

Network Architecture designed for Image

Image Classification

one-hot vector
向量的长度 = 类别数量

100 x 100

Model

$$\begin{bmatrix} \vdots \\ 0.2 \\ 0.7 \\ 0.1 \\ \vdots \end{bmatrix} \quad y'$$

prediction

Cross entropy

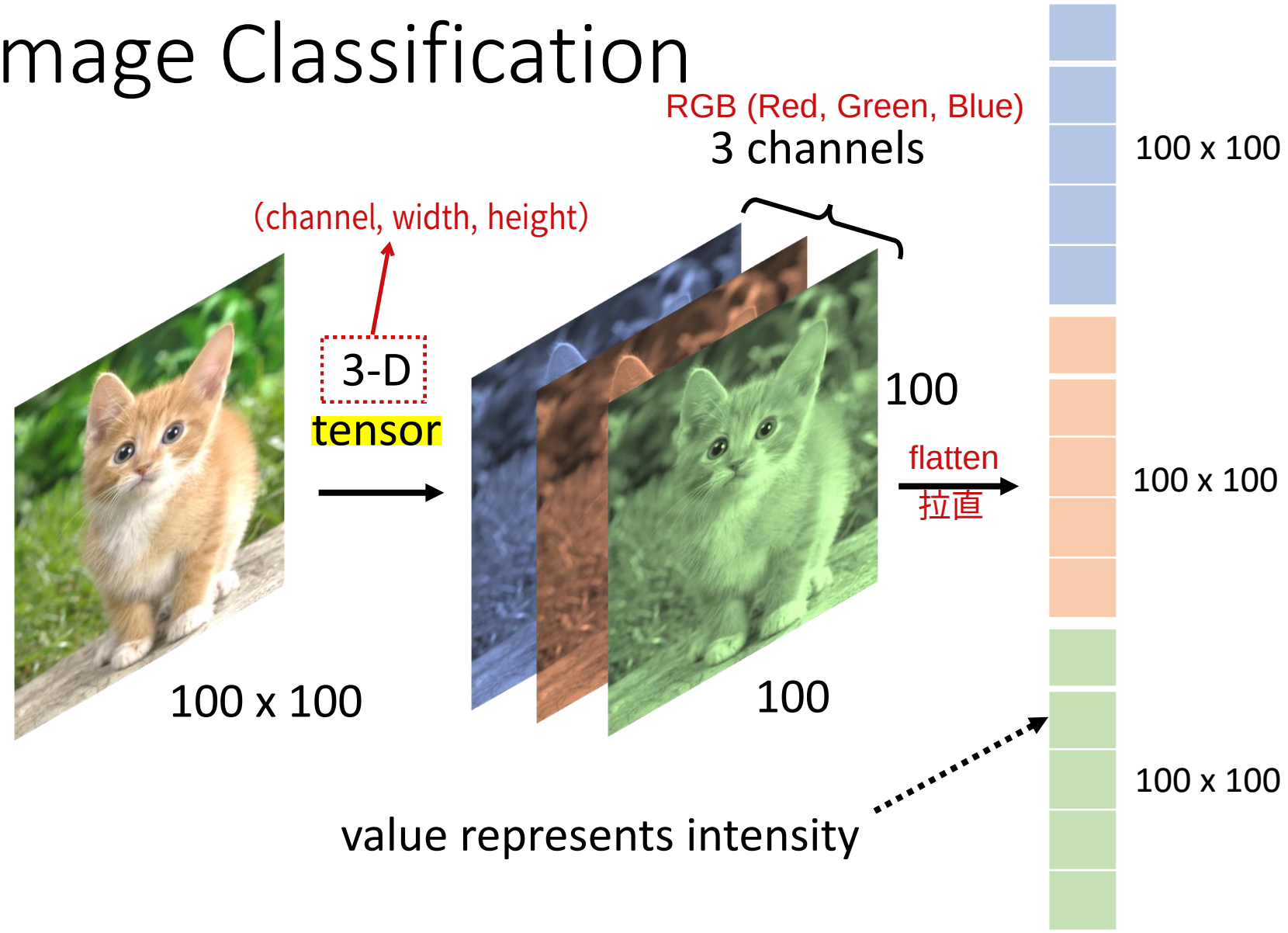
$$\begin{matrix} \\ \text{dog} \\ \text{cat} \\ \text{tree} \\ \\ \end{matrix} \begin{bmatrix} \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix} \quad \hat{y}$$

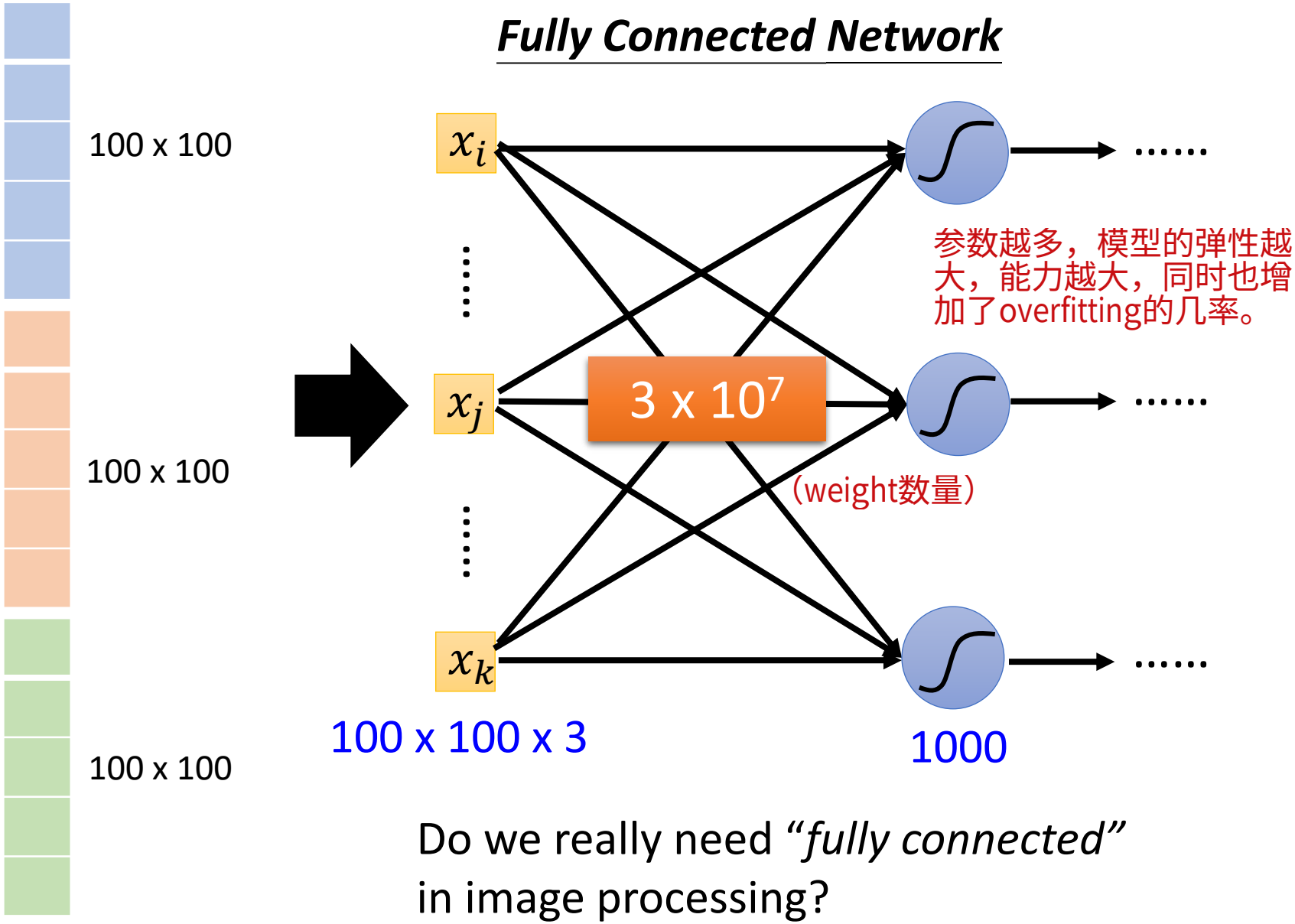
target

(All the images to be classified have the same size.)

输入图片的大小都要一样

Image Classification

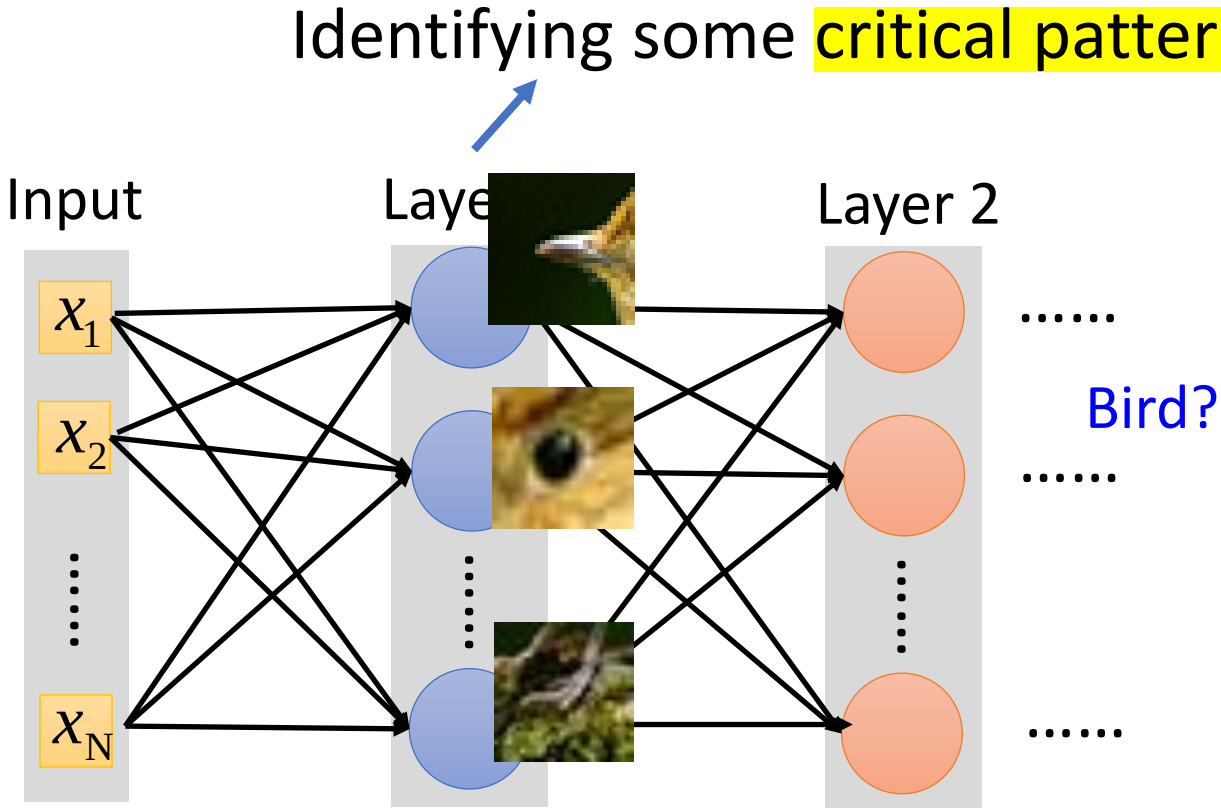




Do we really need "*fully connected*" in image processing?

观察到某个重要的特性来做判断

Observation 1



Perhaps human also identify birds in a similar way … ☺

人也是抓住某个事物的特性来判断。

鸟的眼睛

鸟嘴

全身都是黑的

所有特征加起来，让有的人第一眼会觉得：这是乌鸦。

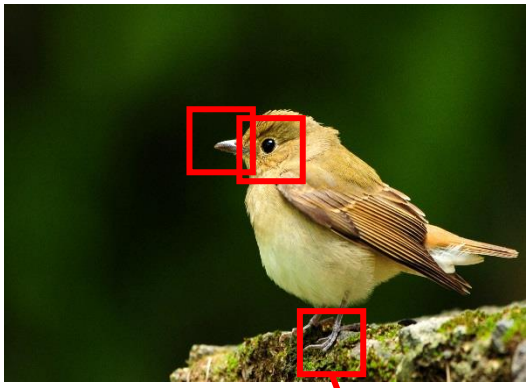
https://www.dcard.tw/f/funny/p/233833012

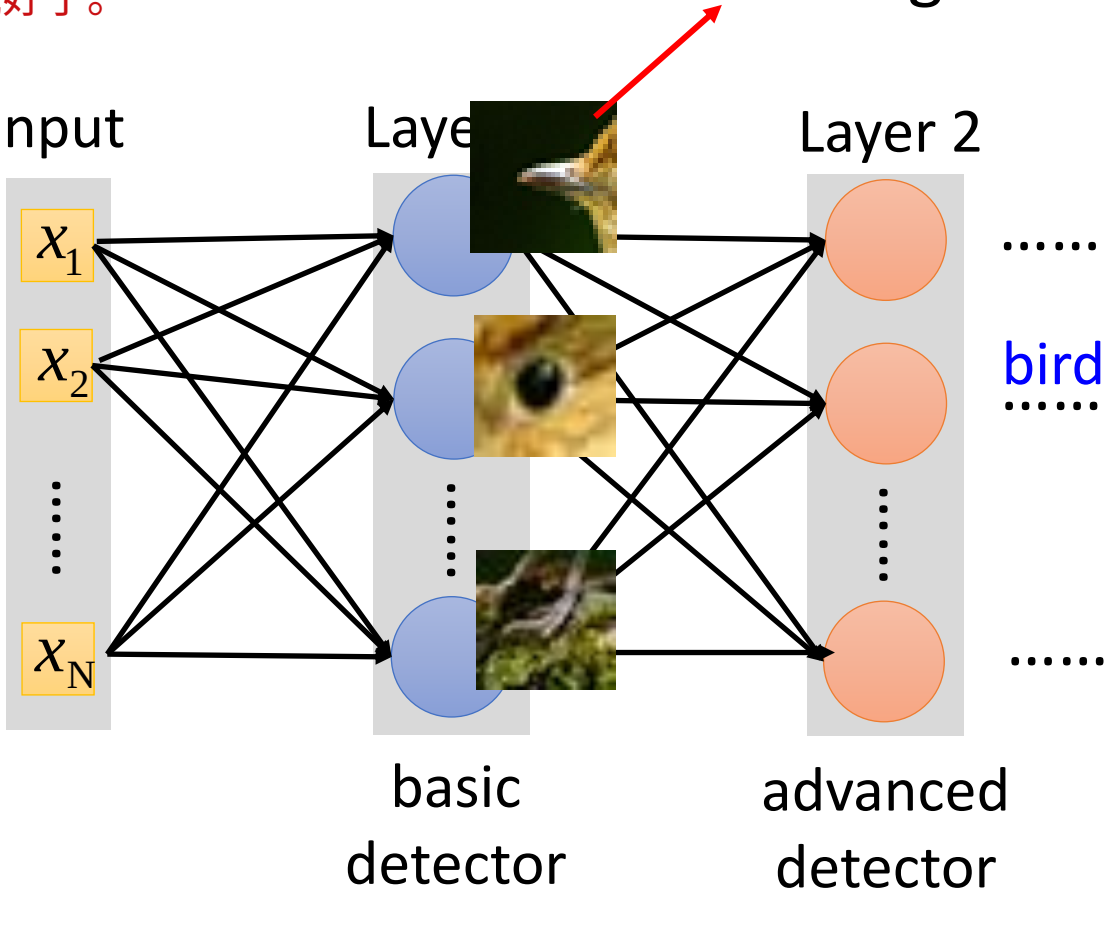
Observation 1

A neuron does not have to see the whole image.

不需要看整张图片，看局部就好了。
不需要将整个图片作为输入。

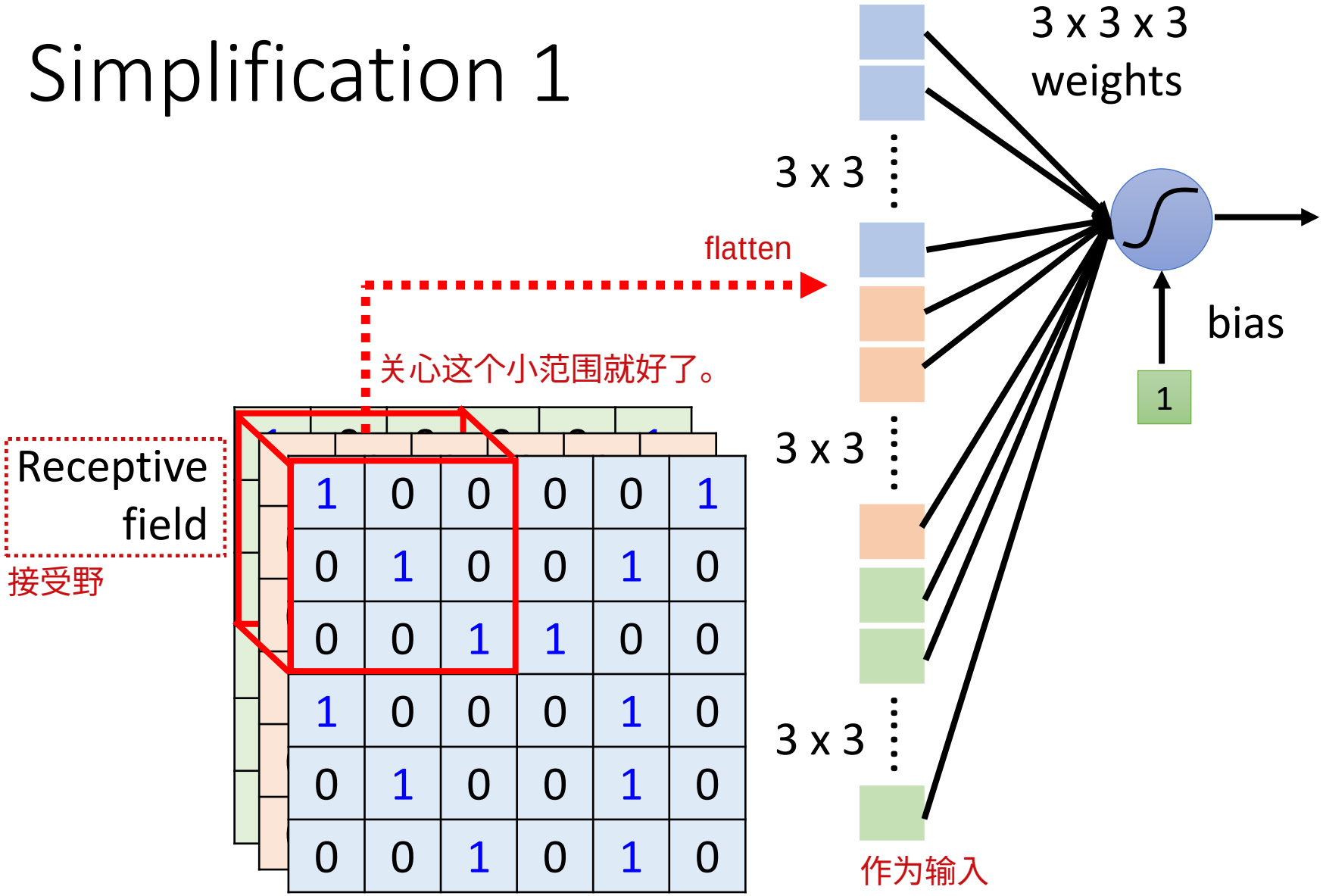
Need to see the whole image?





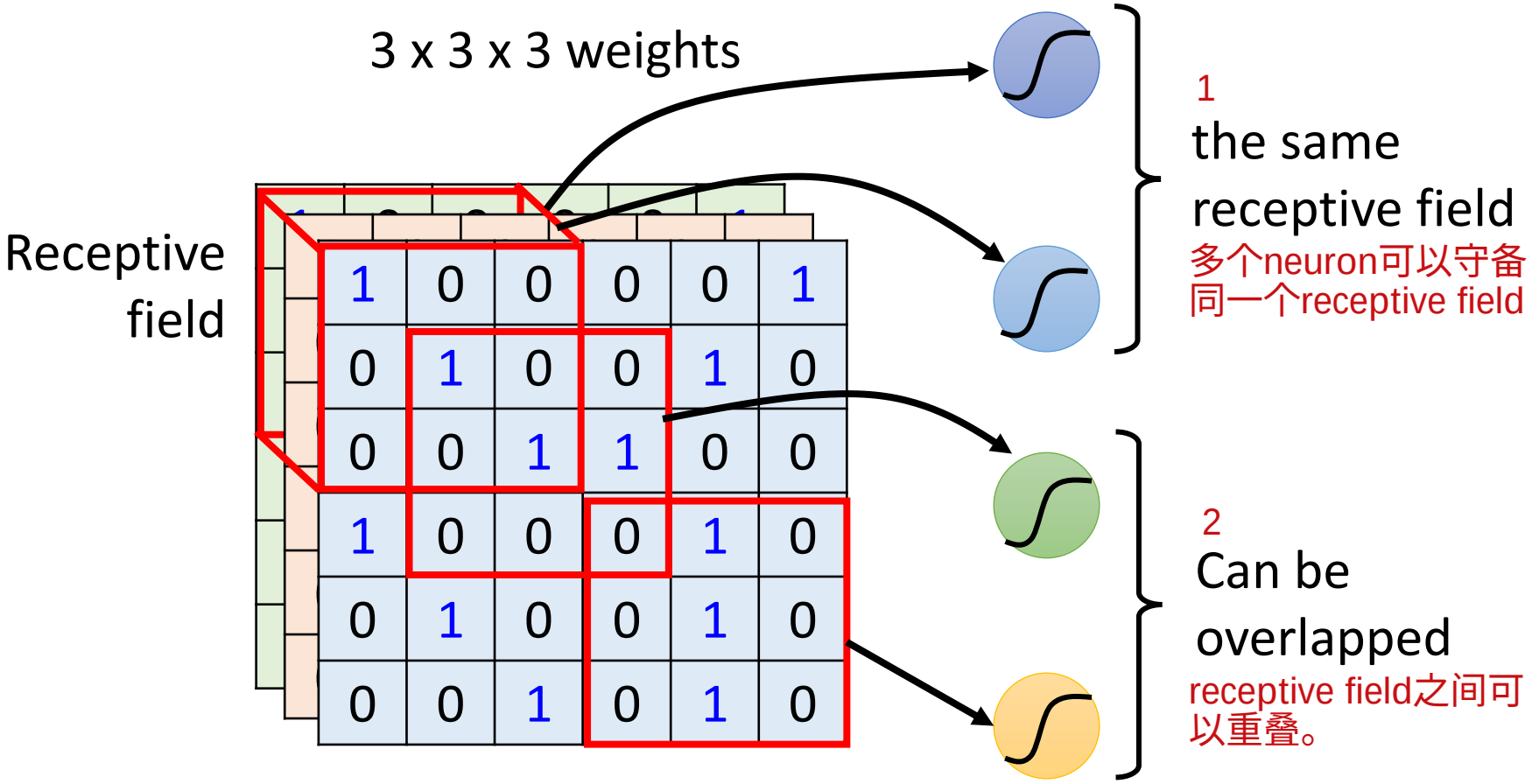
Some patterns are much smaller than the whole image.

Simplification 1



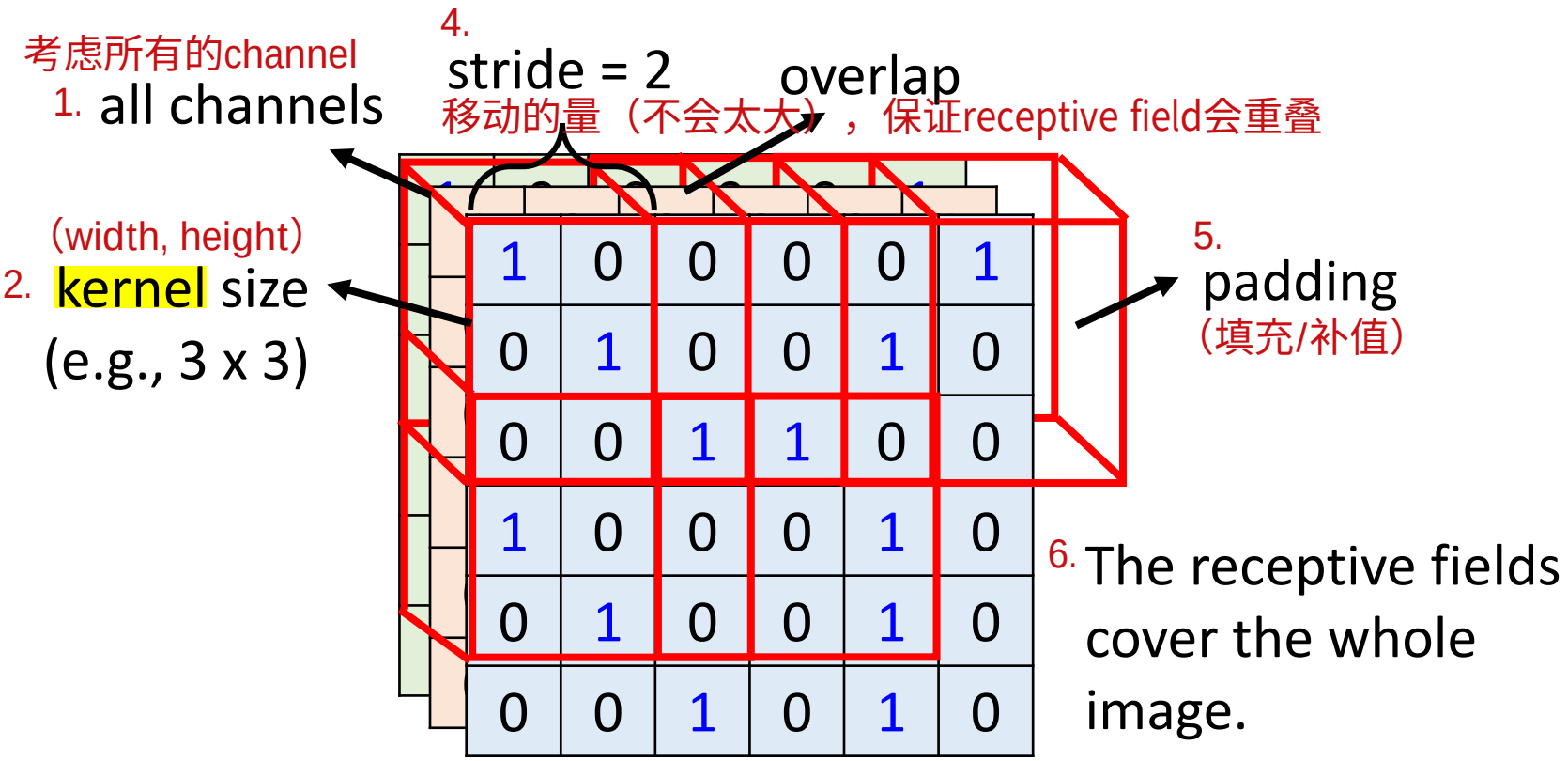
Simplification 1

- Can different neurons have different sizes of receptive field?
  可以，常见招式，选择不同大小的receptive field。
- Cover only some channels?
  可以，在network compression里会讲。
- Not square receptive field?
  可以的，自己决定。



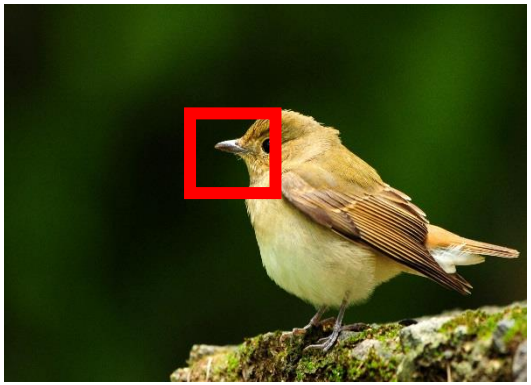
Simplification 1 – Typical Setting

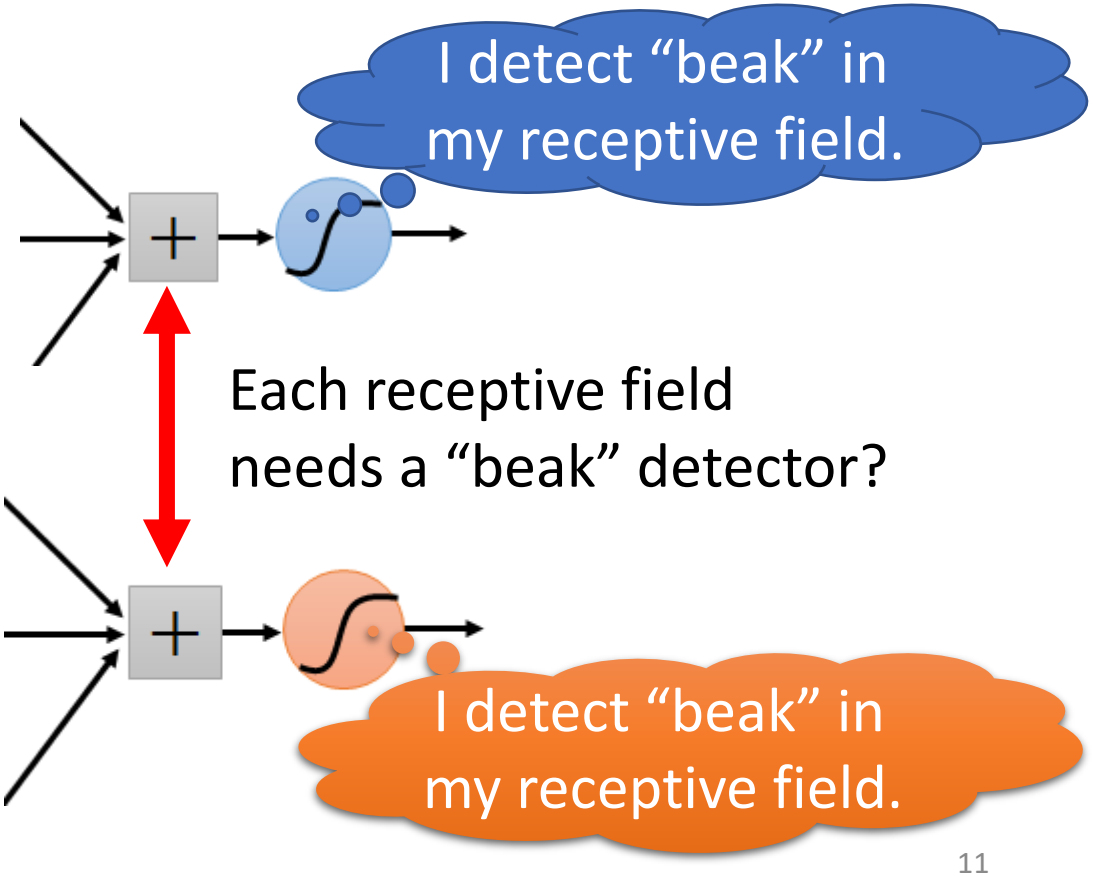
3. Each receptive field has a set of neurons (e.g., 64 neurons).



Observation 2 特征在不同的区域

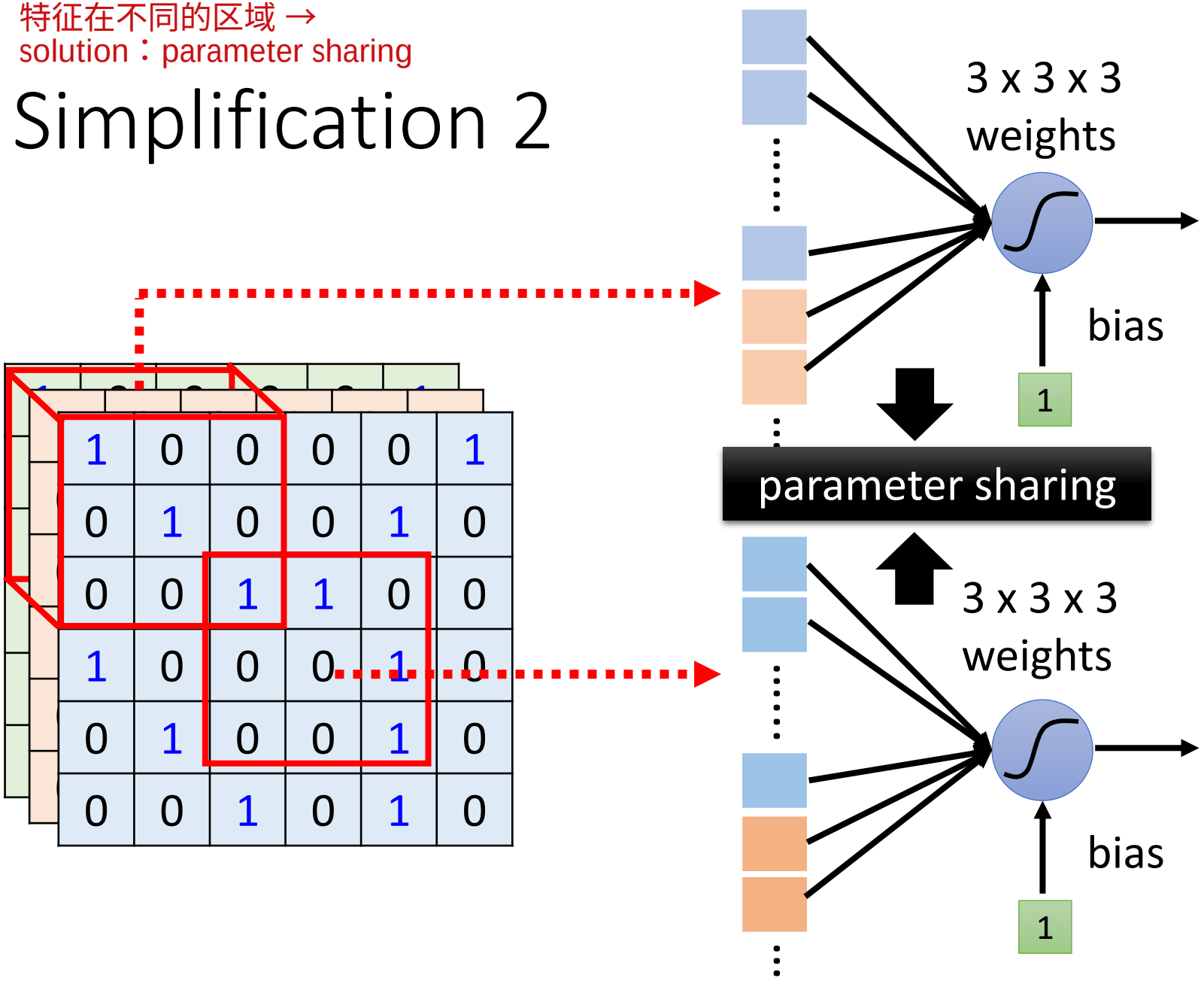
- The same patterns appear in different regions.





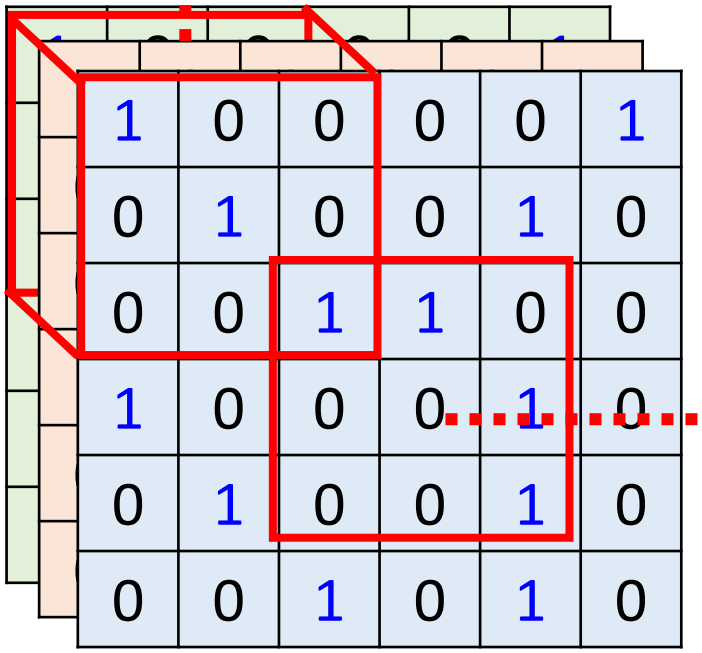
特征在不同的区域 →
solution：parameter sharing

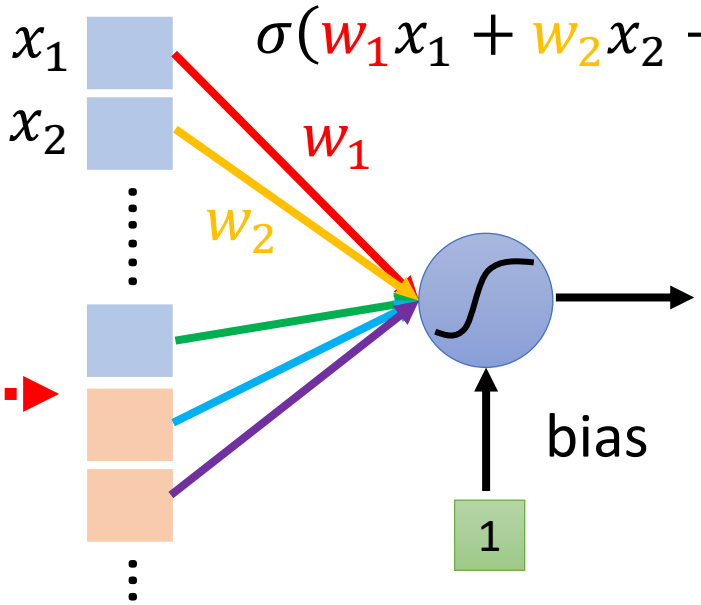
Simplification 2

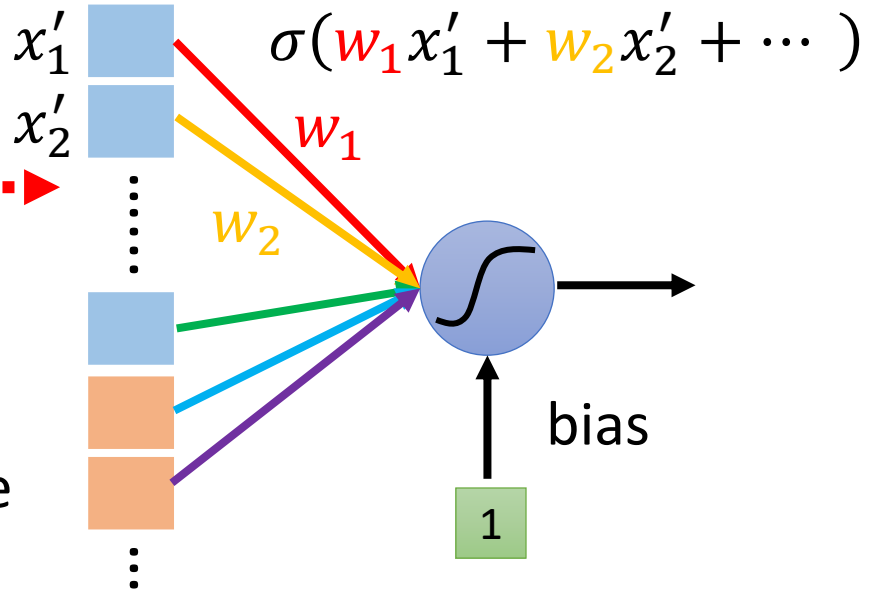


相同receptive field → 不共享参数 → 不同output
不同receptive field → 共享参数 → 不同output

Simplification 2



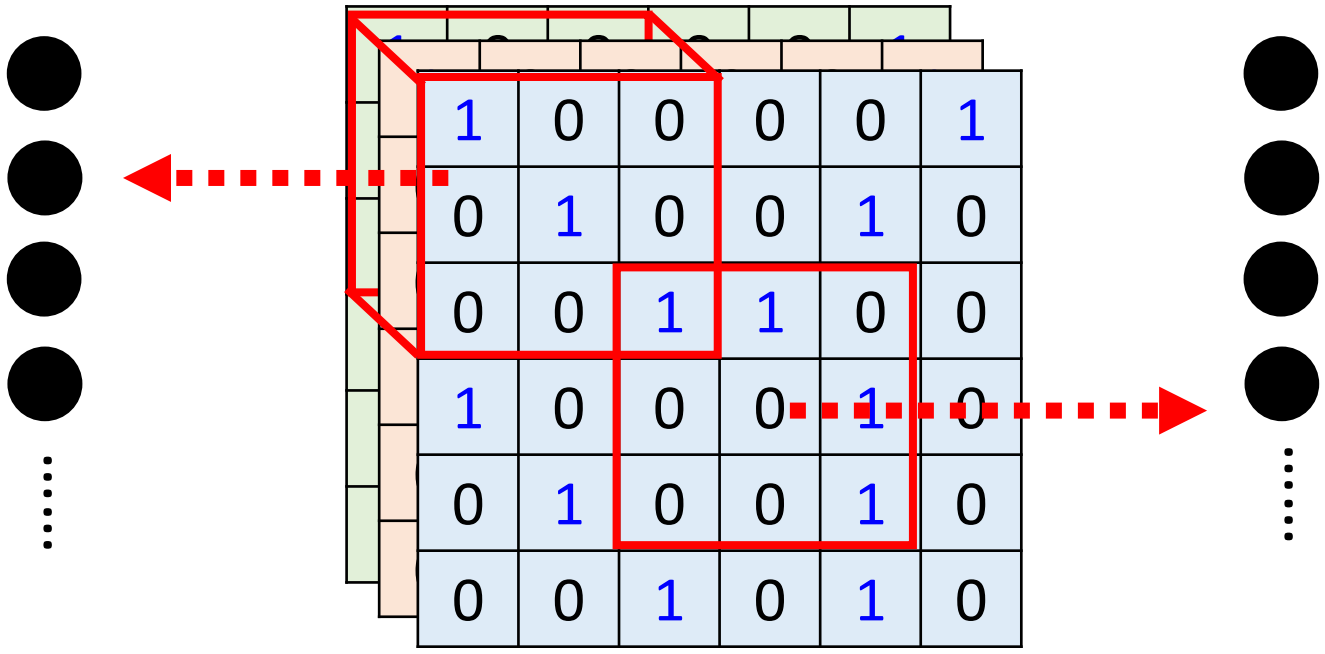




Two neurons with the same receptive field would not share parameters.

Simplification 2 – Typical Setting

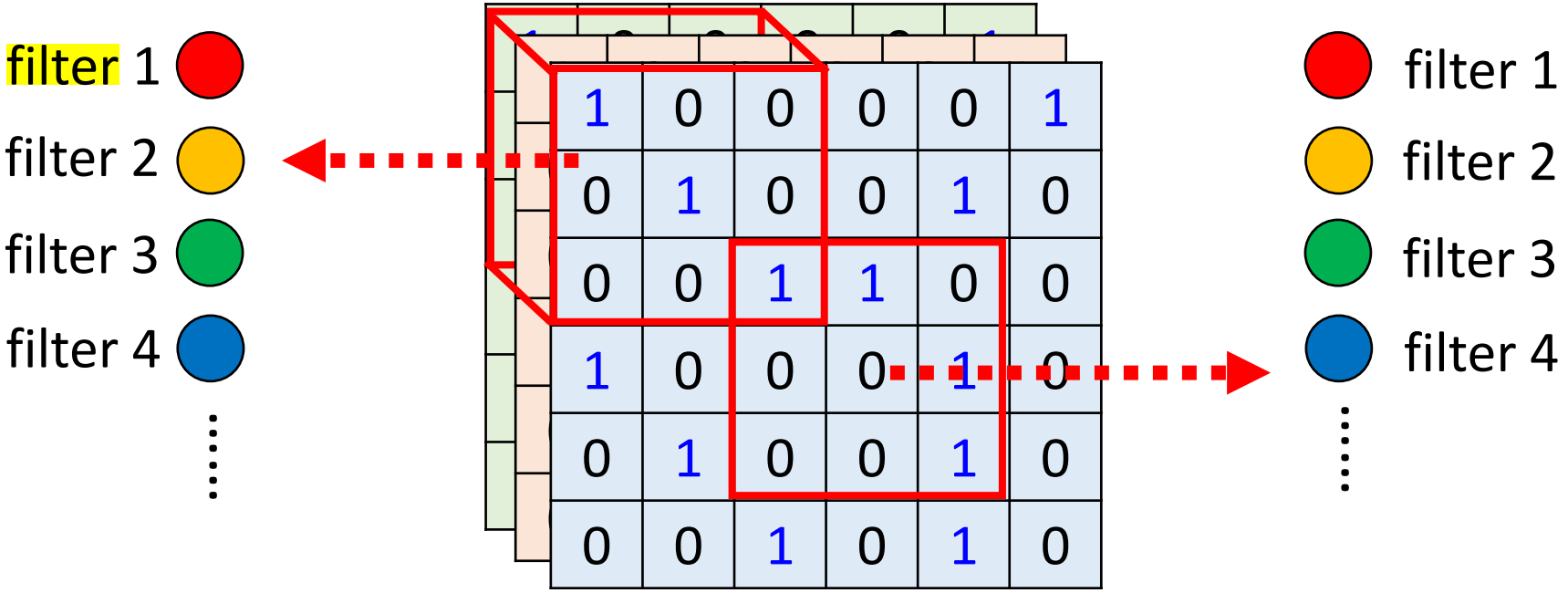
Each receptive field has a set of neurons (e.g., 64 neurons).



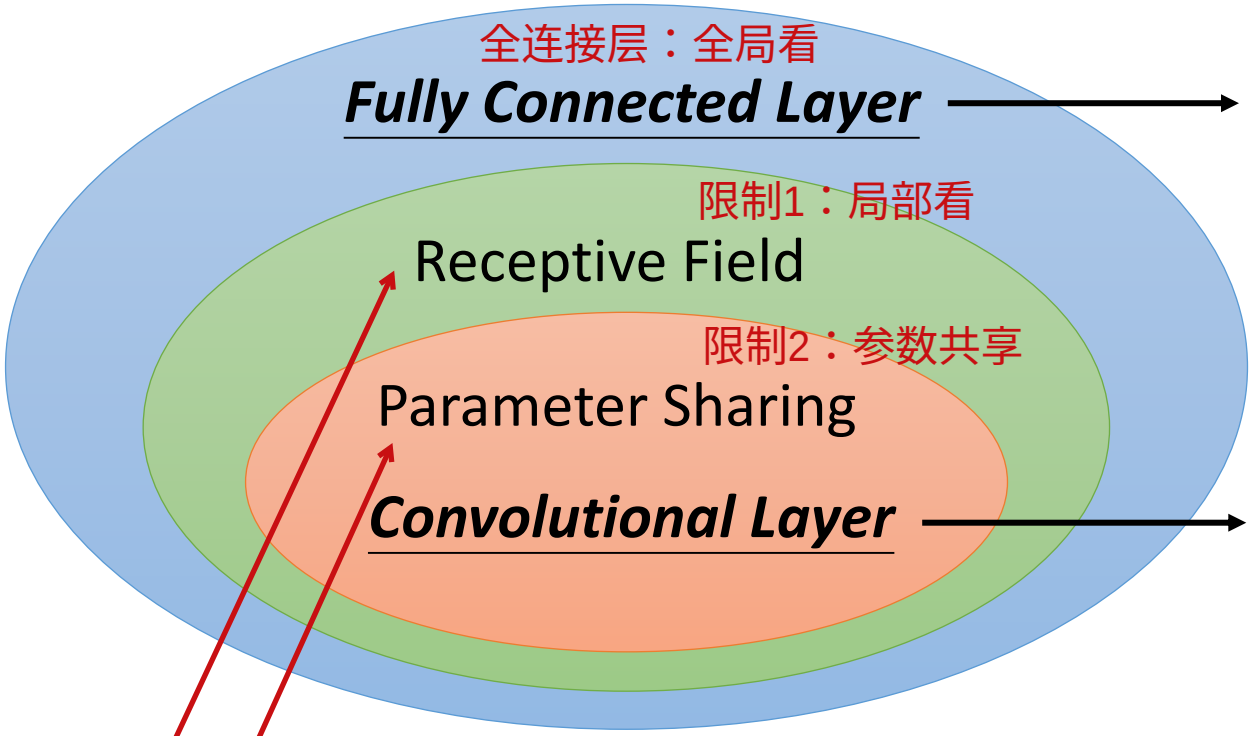
Simplification 2 – Typical Setting

Each receptive field has a set of neurons (e.g., 64 neurons).

Each receptive field has the neurons with the same set of parameters.



Benefit of Convolutional Layer



弹性最大

Jack of all trades, master of none

限制大，model bias大，flexibility小，但是特别适合image

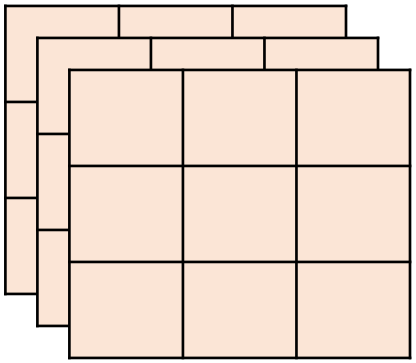
Larger model bias (for image)

- Some patterns are much smaller than the whole image.
- The same patterns appear in different regions.

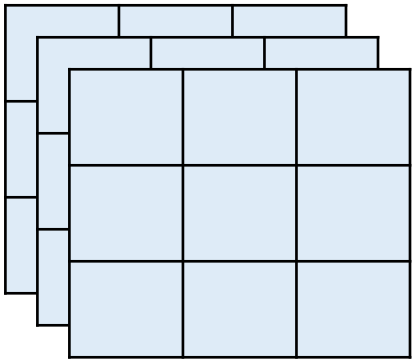
Another story based on ***filter*** ☺

Convolutional Layer

Convolution



Filter 1
3 x 3 x channel
tensor



Filter 2
3 x 3 x channel
tensor

channel = 3 (colorful) 彩色

channel = 1 (black and white) 黑白

Each filter detects a small pattern (3 x 3 x channel).

Convolutional Layer

Consider channel = 1
(black and white image)

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

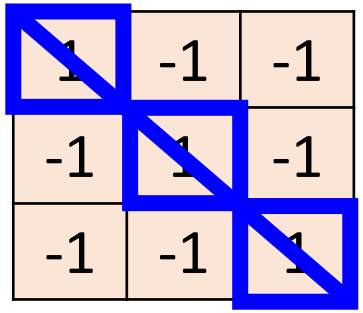
Filter 1

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

Filter 2

(The values in the filters are unknown parameters.)

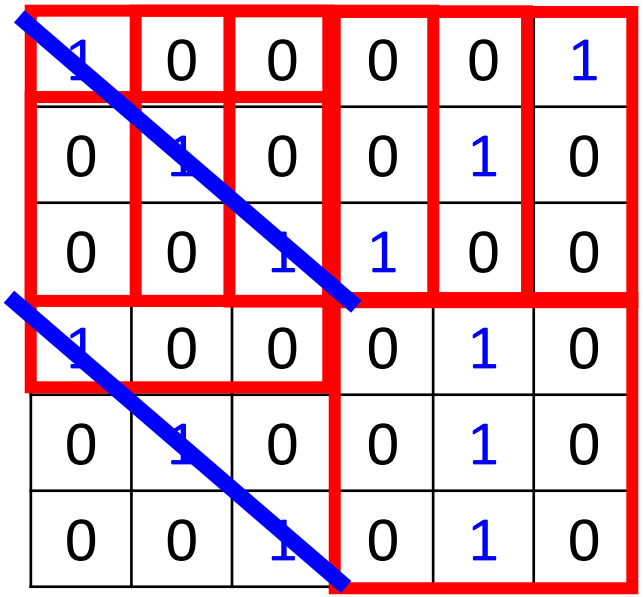
Convolutional Layer



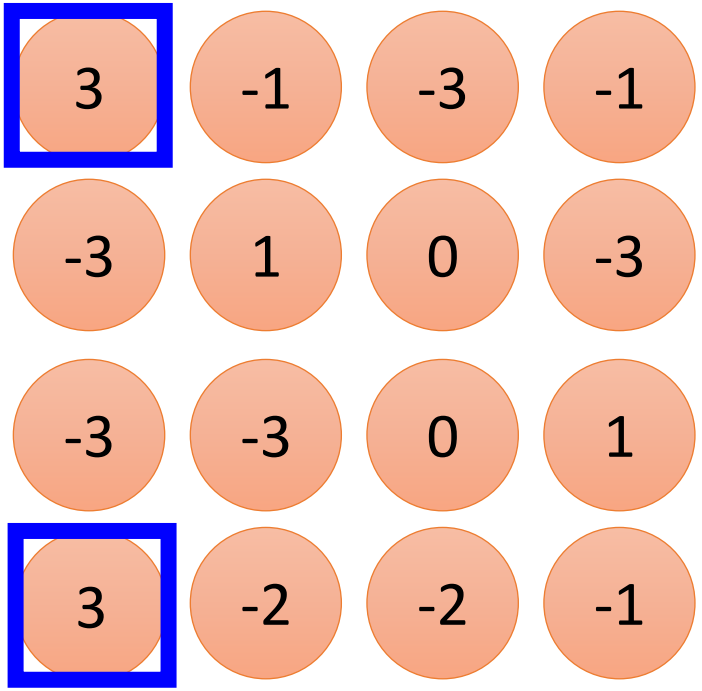
Filter 1

Inner Product

stride=1



6 x 6 image



左上角和左下角的值最大 → 有filter的特征

feature map

Convolutional Layer

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

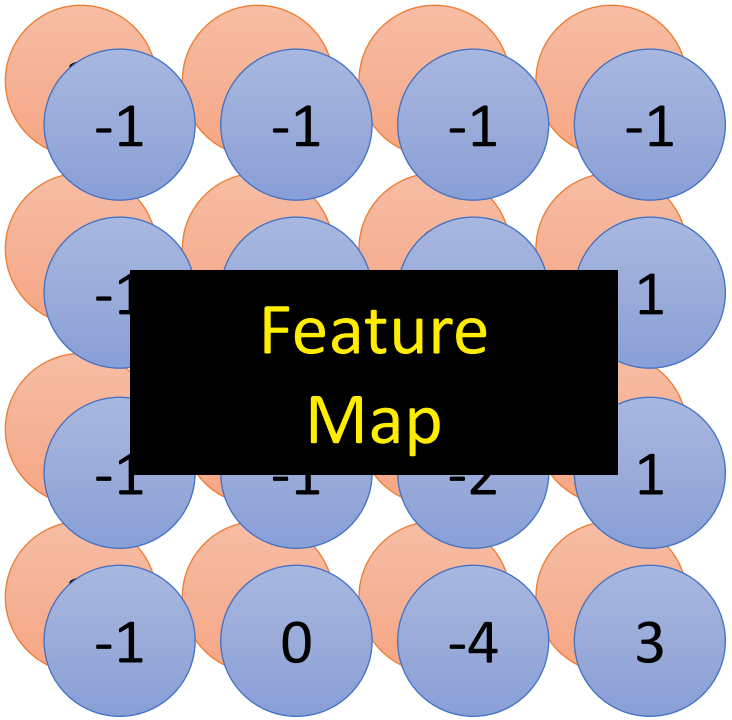
Filter 2

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

Do the same process for every filter

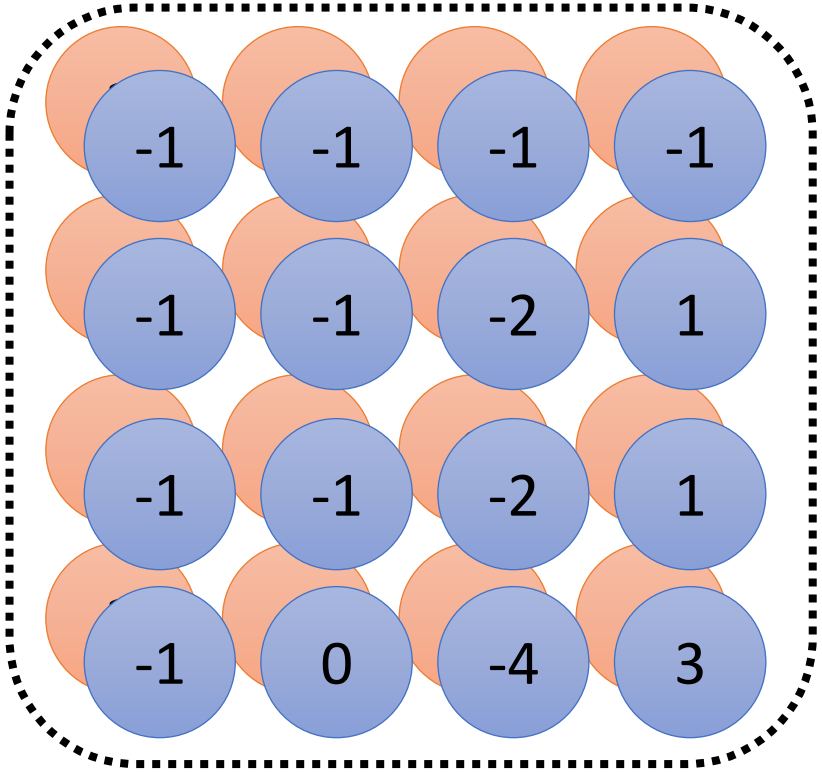


Convolutional Layer

64 filters

Convolution

Convolution



"Image" with 64 channels

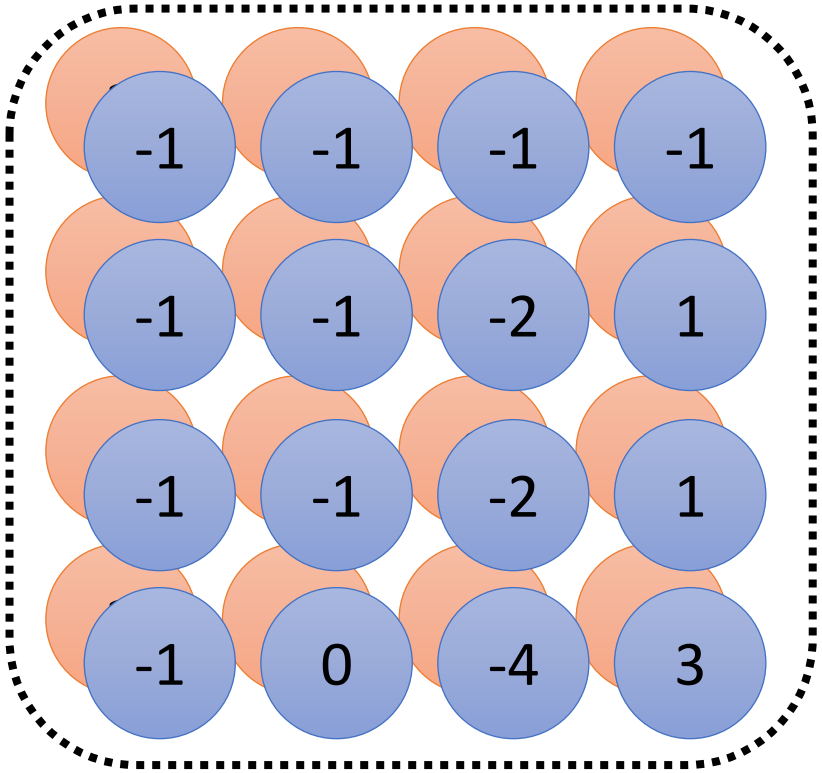
可以看成是另外一张图片，但是是64个channels。
# of channels = # of filters of the last layer

Multiple Convolutional Layers

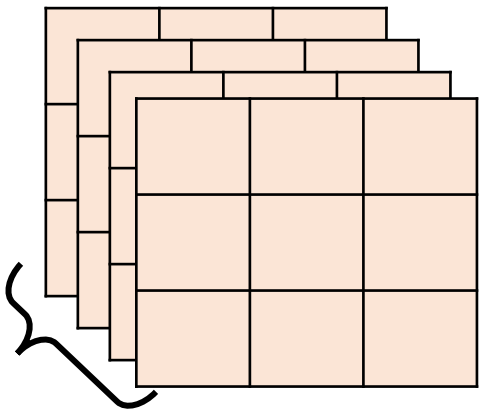
64 filters

Convolution

Convolution



"Image" with 64 channels



Filter:
3 x 3 x 64

Multiple Convolutional Layers

64 filters

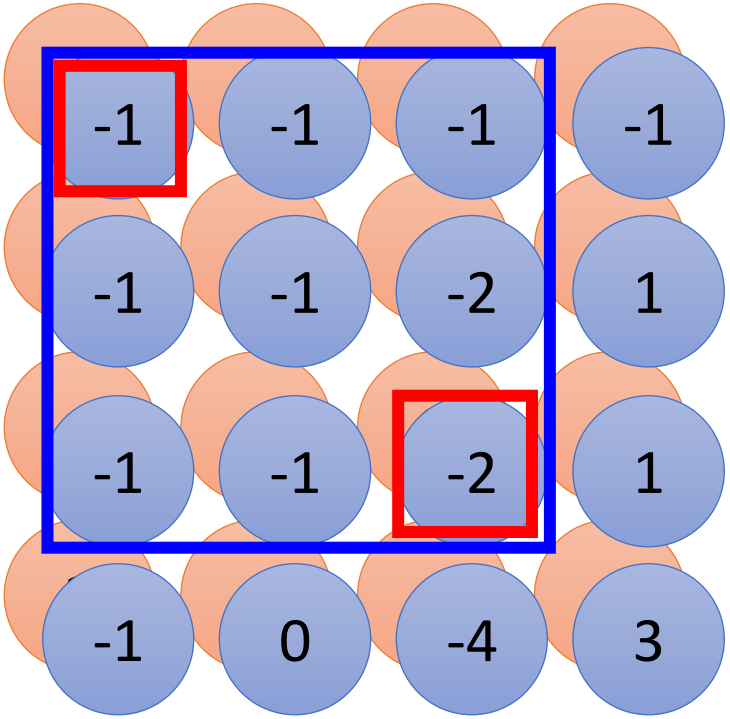
Convolution

Convolution

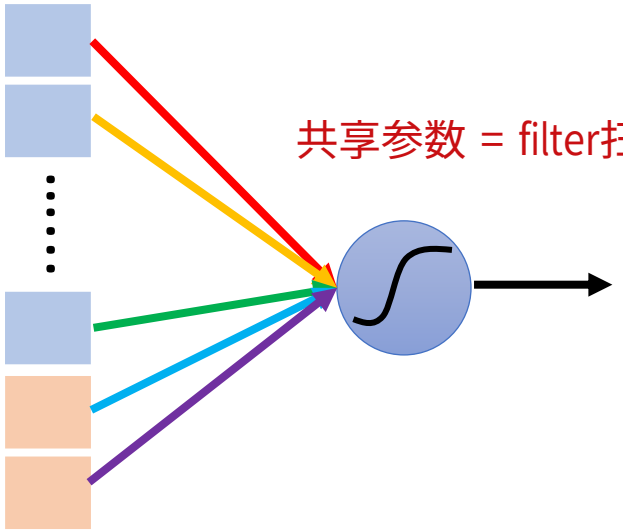
实际上考虑的是5×5

看起来是3×3

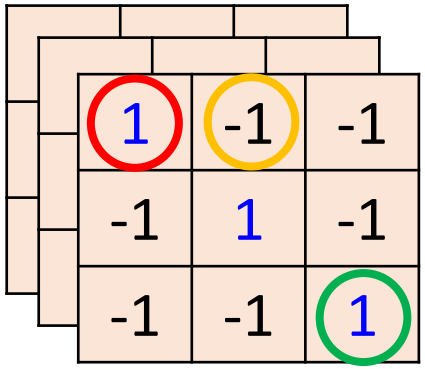
1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0



Comparison of Two Stories



共享参数 = filter扫过



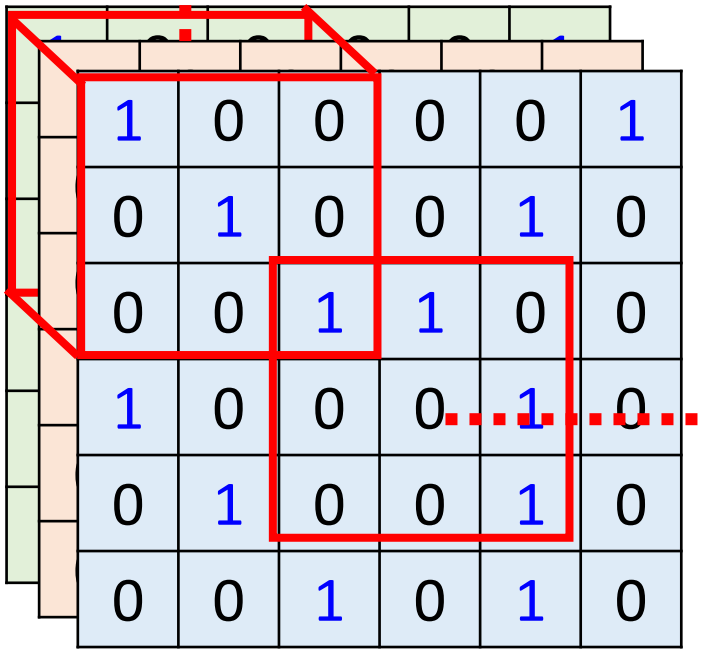
Filter
3 x 3 x channel
tensor

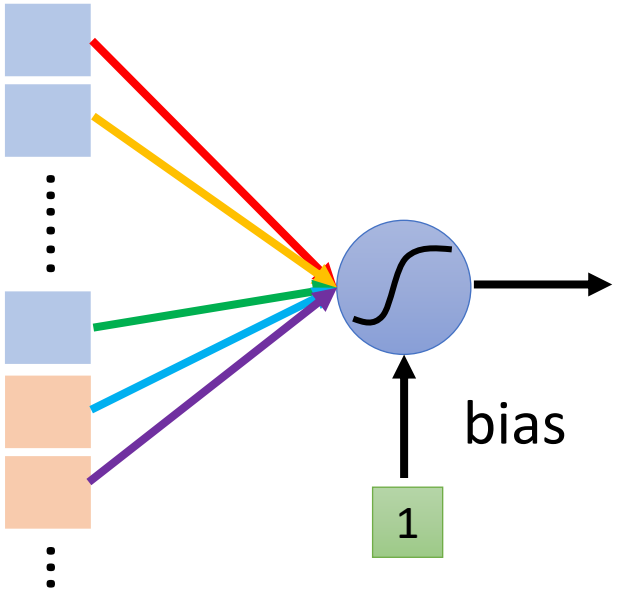
Receptive field

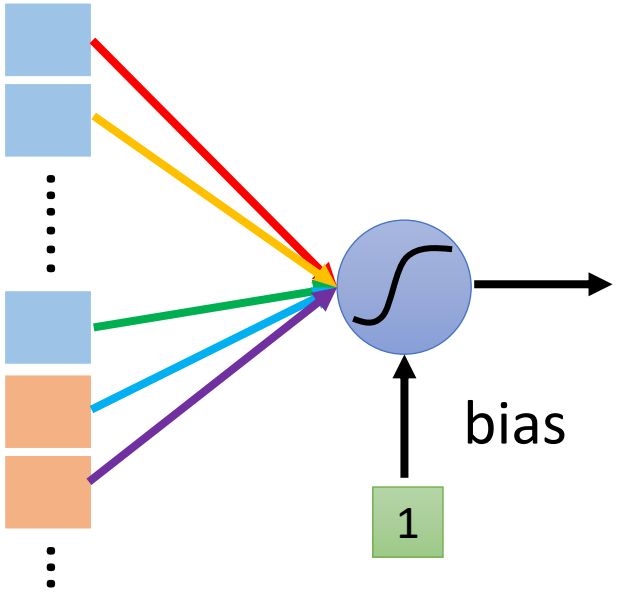
(ignore bias in this slide)

convolution = filter扫过整张图片 = receptive field共享参数

The neurons with different receptive fields **share the parameters**.







Each filter convolves over the input image.

Convolutional Layer

Neuron Version Story	***Filter Version Story***
Each neuron only considers a receptive field.	There are a set of filters detecting small patterns.
The neurons with different receptive fields share the parameters.	Each filter convolves over the input image.

They are the same story.

pooling

Observation 3

- Subsampling the pixels will not change the object

bird

subsampling

下采样

bird

pooling不会学习；没有参数

Pooling – Max Pooling

选的代表：最大的那个

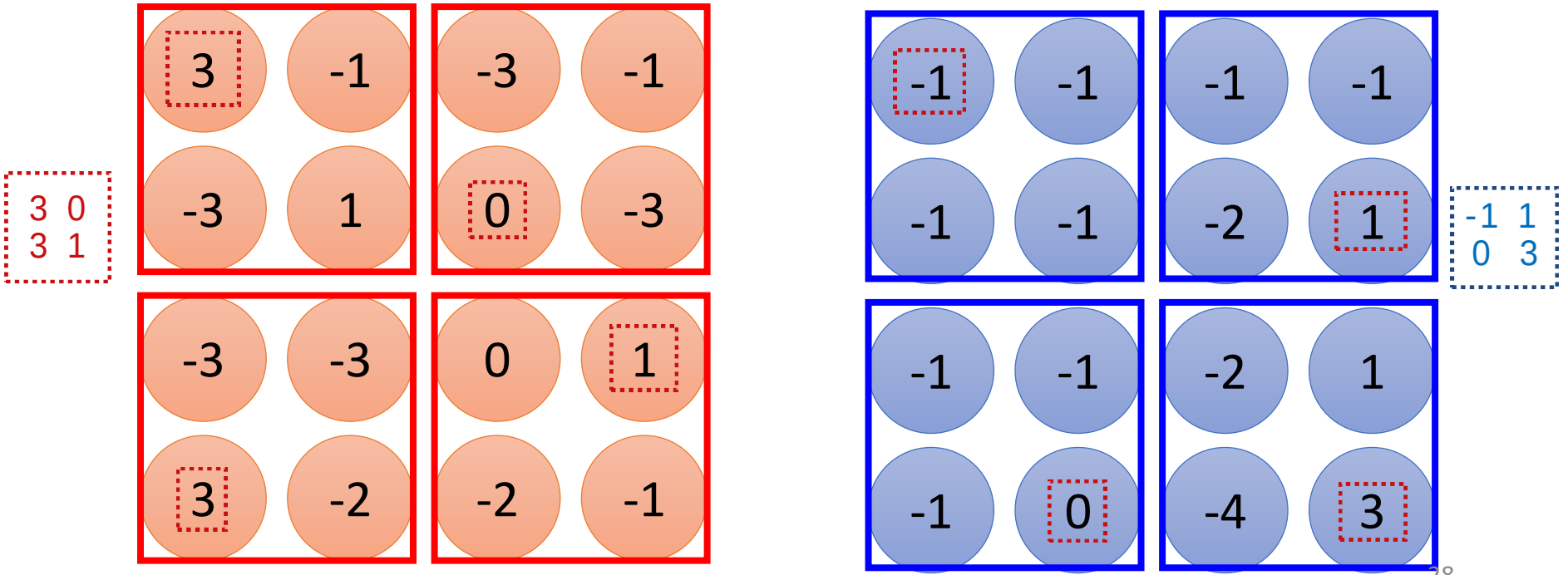
有很多pooling的方法；
e.g., mean pooling

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

Filter 1

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

Filter 2



Convolutional Layers + Pooling

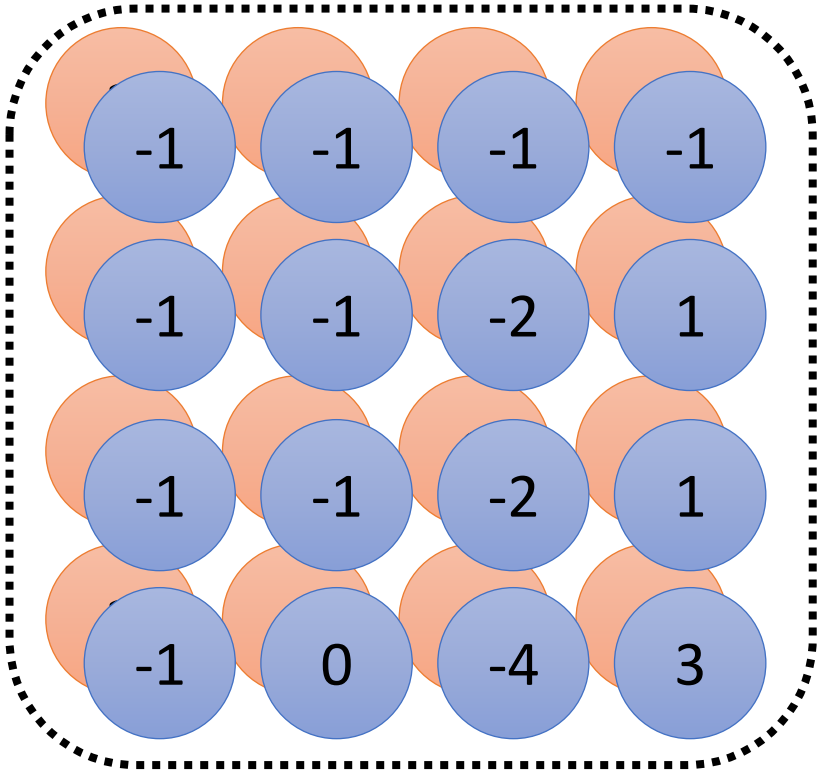
Repeat

Convolution

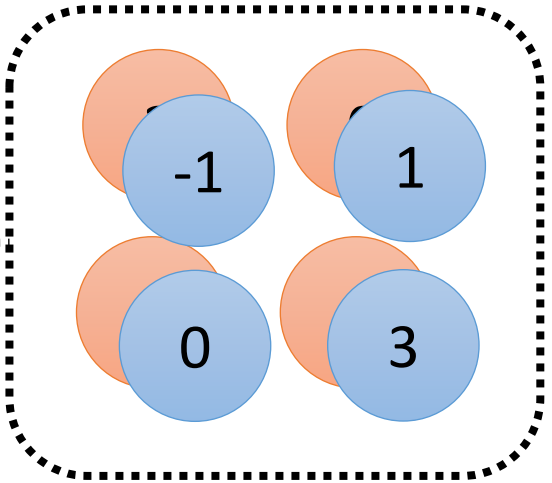
Pooling

（会让图片失真）

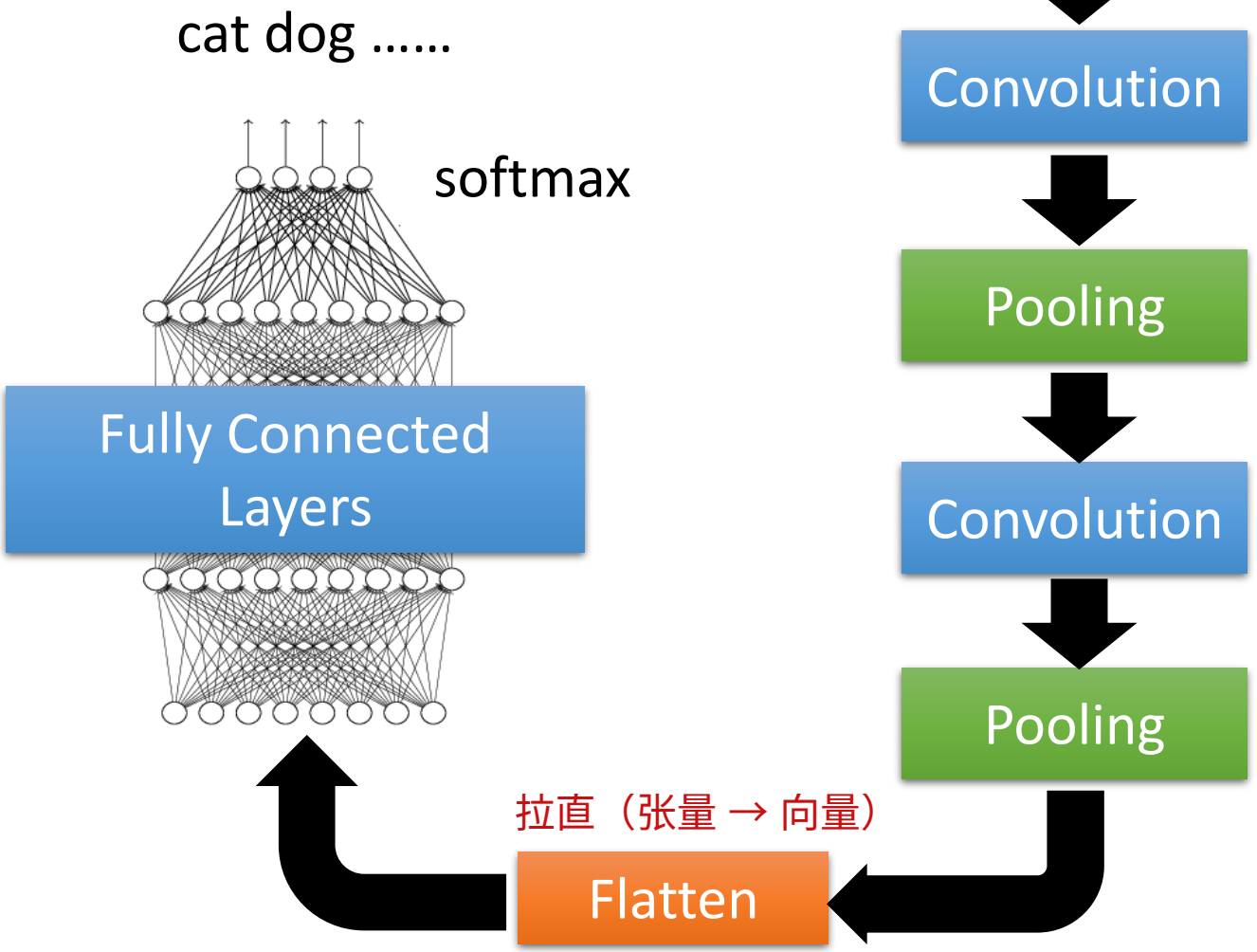
pooling的目的：减少运算量



“Image” with 64 channels



The whole CNN



Application: Playing Go 下围棋 = 分类问题

19 x 19 matrix (image)

Network

Next move (19 x 19 positions) 小的图片

19 x 19 classes

(19, 19, 48)

domain knowledge

48 channels in Alpha Go

Black: 1

white: -1

none: 0

Fully-connected network can be used

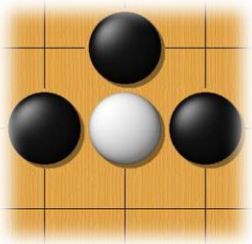
But CNN performs much better.

当任务和图像处理有类似特性时，可以用CNN。

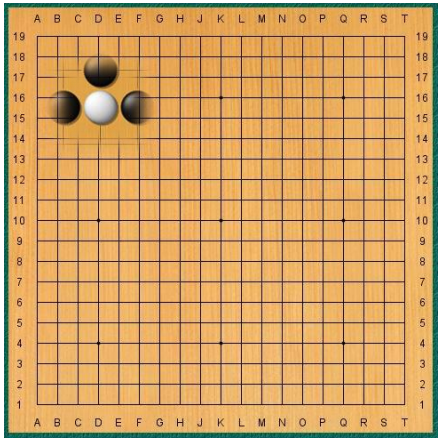
Why CNN for Go playing?

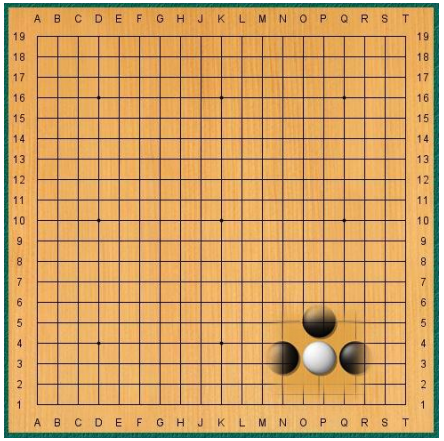
1. Some patterns are much smaller than the whole image

Alpha Go uses 5 x 5 for first layer



2. The same patterns appear in different regions.

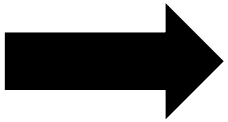




AlphaGo的CNN不用Pooling，因为Pooling会掉信息，所以不适合下棋的任务。

Why CNN for Go playing?

- Subsampling the pixels will not change the object



Pooling

How to explain this???

Neural network architecture. The input to the policy network is a $19 \times 19 \times 48$ image stack consisting of 48 feature planes. The first hidden layer zero pads the input into a 23×23 image, then convolves k filters of kernel size 5×5 with stride 1 with the input image and applies a rectifier nonlinearity. Each of the subsequent hidden layers 2 to 12 zero pads the respective previous hidden layer into a 21×21 image, then convolves k filters of kernel size 3×3 with stride 1, again followed by a rectifier nonlinearity. The final layer convolves 1 filter of kernel size 1×1 with stride 1, with a different bias for each position, and applies a softmax function. The match version of AlphaGo used $k = 192$ filters; Fig. 2b and Extended Data Tabl 256 and 384 filters

ReLU

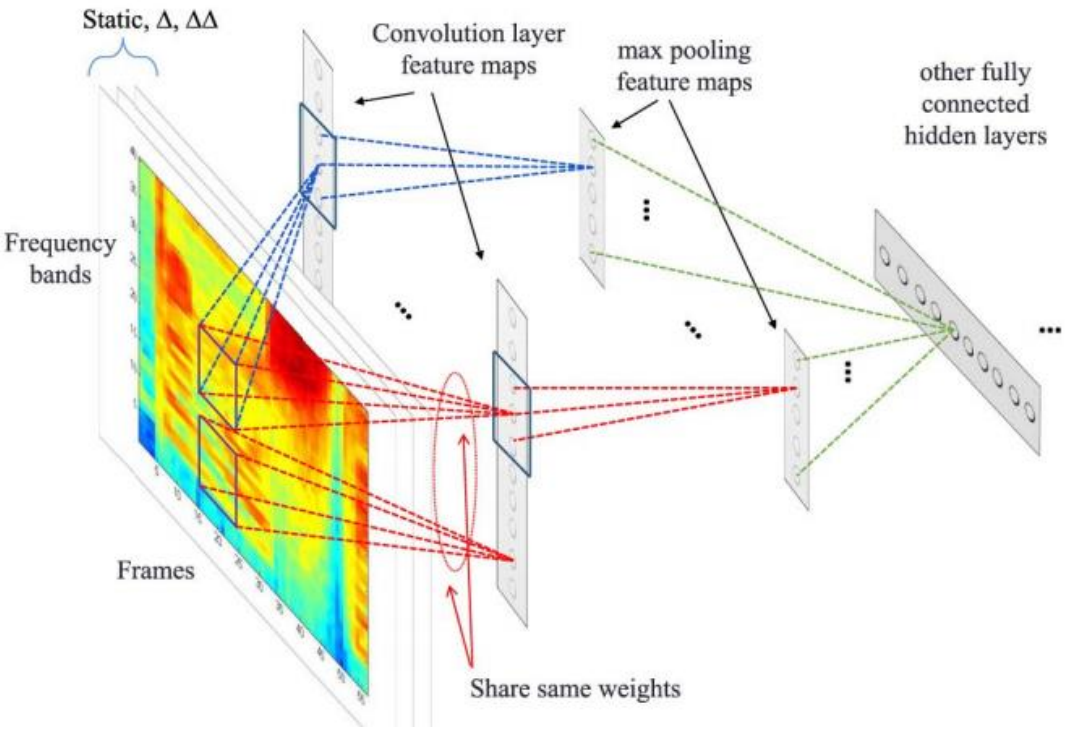
Alpha Go does not use Pooling ……

More Applications

语音和NLP上也都有应用。

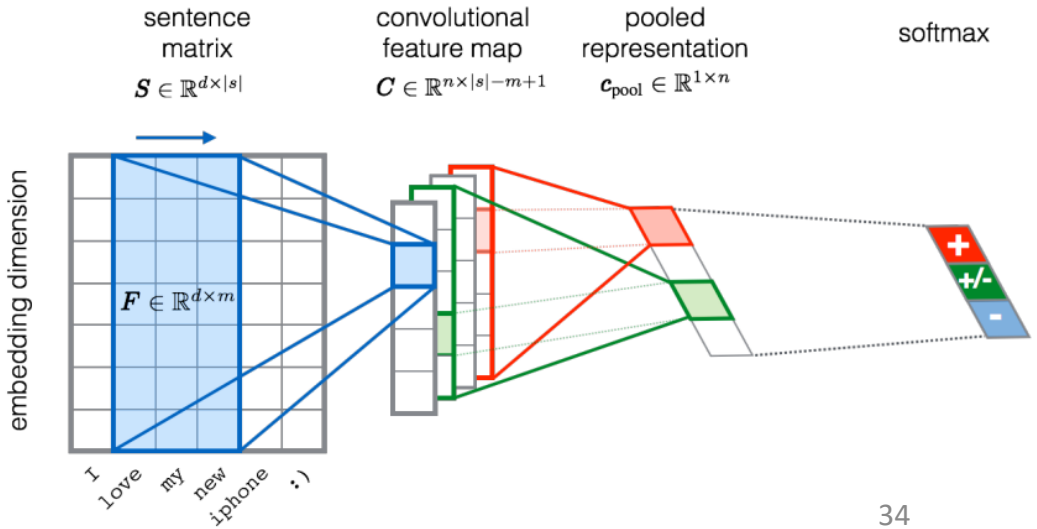
Speech

https://dl.acm.org/doi/10.1109/TASLP.2014.2339736



Natural Language Processing

https://www.aclweb.org/anthology/S15-2079/



To learn more …

- CNN is not invariant to scaling and rotation (we need data augmentation ☺).

放大/放小　　旋转

Spatial Transformer Layer

https://youtu.be/SoCywZ1hZak

(in Mandarin)