



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



学生创新中心
Student Innovation Center

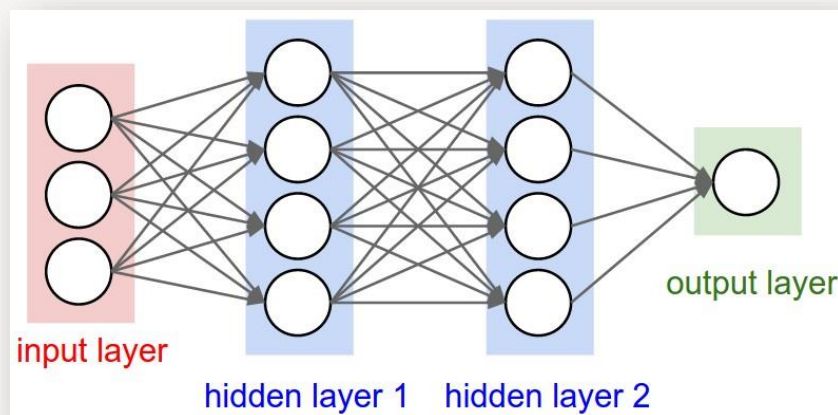
卷积神经网络与经典模型

肖雄子彦

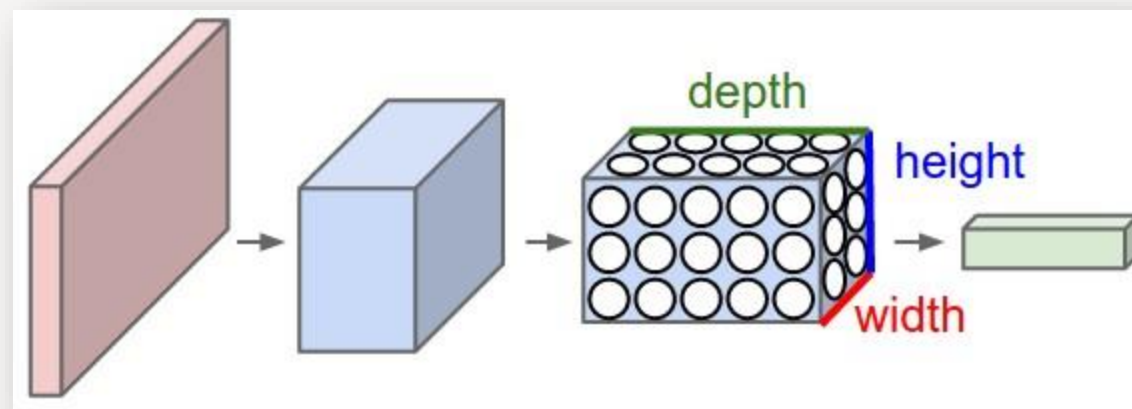


CNN 卷积神经网络

卷积神经网络（Convolutional Neural Network, CNN）是深度学习算法代表之一。CNN是包含卷积计算且具有深度结构的神经网络，对于大型图像处理有出色表现。



线性变化、激活函数、全连接层
传统的全连接神经网络（平面）



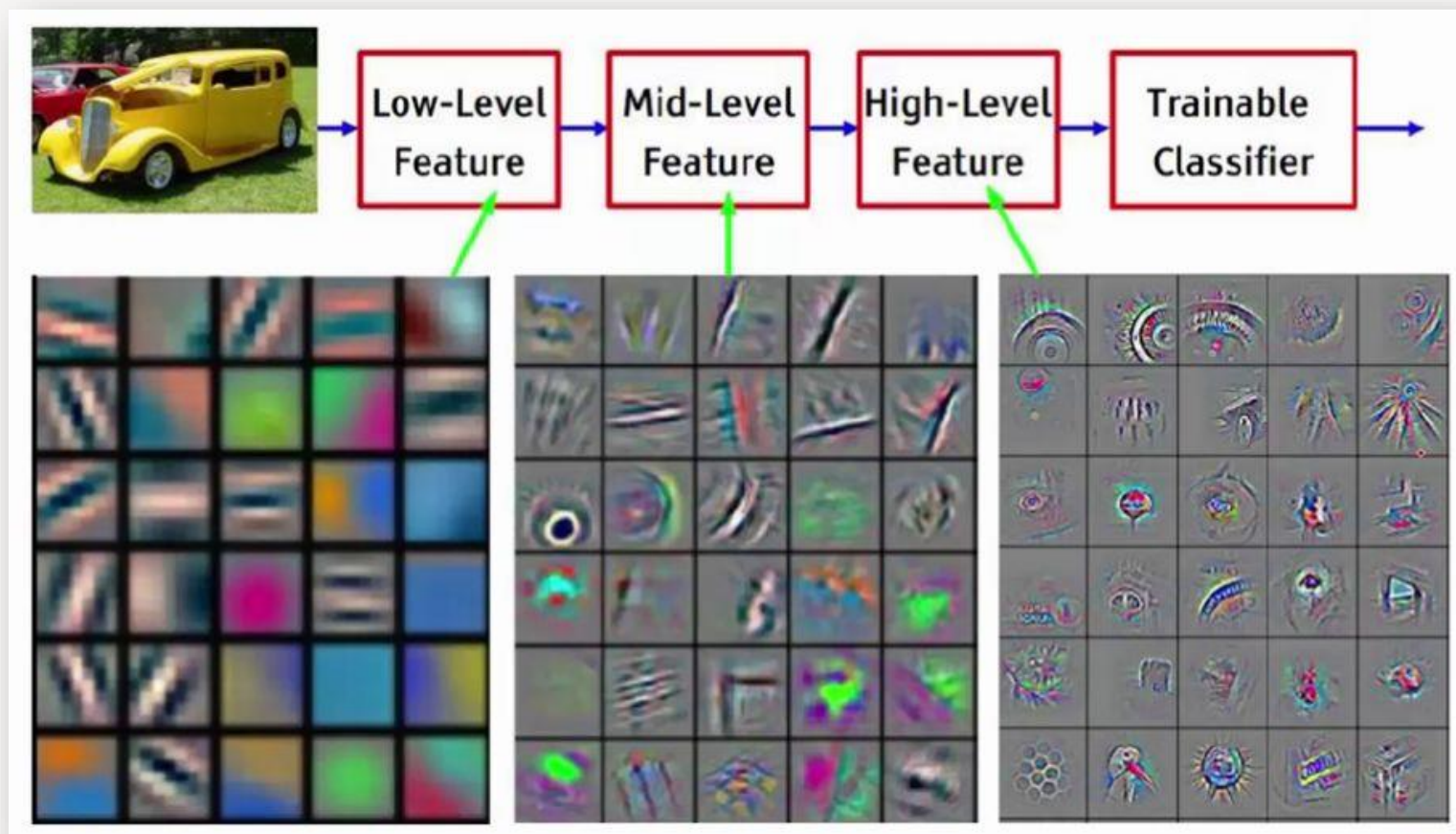
卷积层、激活函数、池化层、全连接层
卷积神经网络CNN（立体）

相对于FC结构来说，擅长图像处理卷积结构
架构更清晰，参数更少，反而是更加简单的

Convolutional Neural Network

卷积神经网络

卷积层的作用一：提取图片特征——随着卷积层的增加，能提取的特征也更复杂。



第一轮：提取出低层次的特征

第二轮：提取出中层次的特征

第三轮：提取出高层次的特征

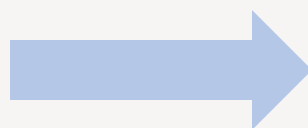
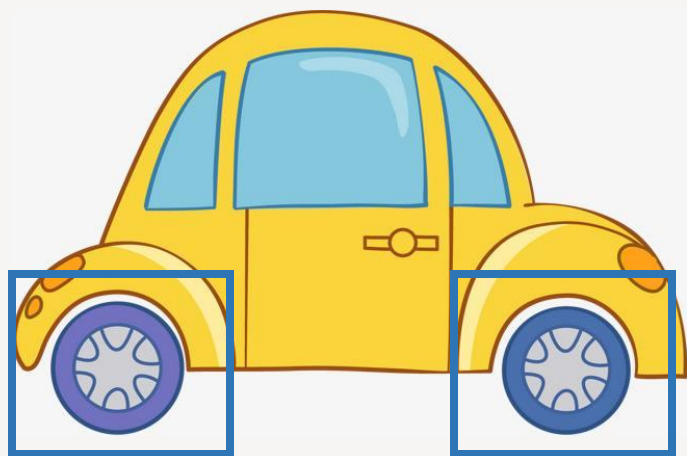
.....

经过一次次的卷积运算
提取到图像高层次特征

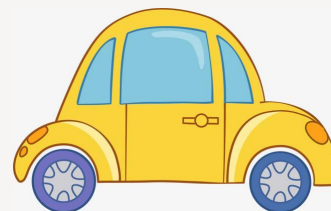
Convolutional Neural Network

卷积神经网络

Why CNN? ——其次，图片中会出现相似的特征，CNN用同一组卷积核提取不同区域的相似特性。此外，一定范围内缩小图像尺寸不影响图像的识别。



subsampling



池化
pooling

卷积层优点二：参数共享
用同一组卷积核，识别相似的特性

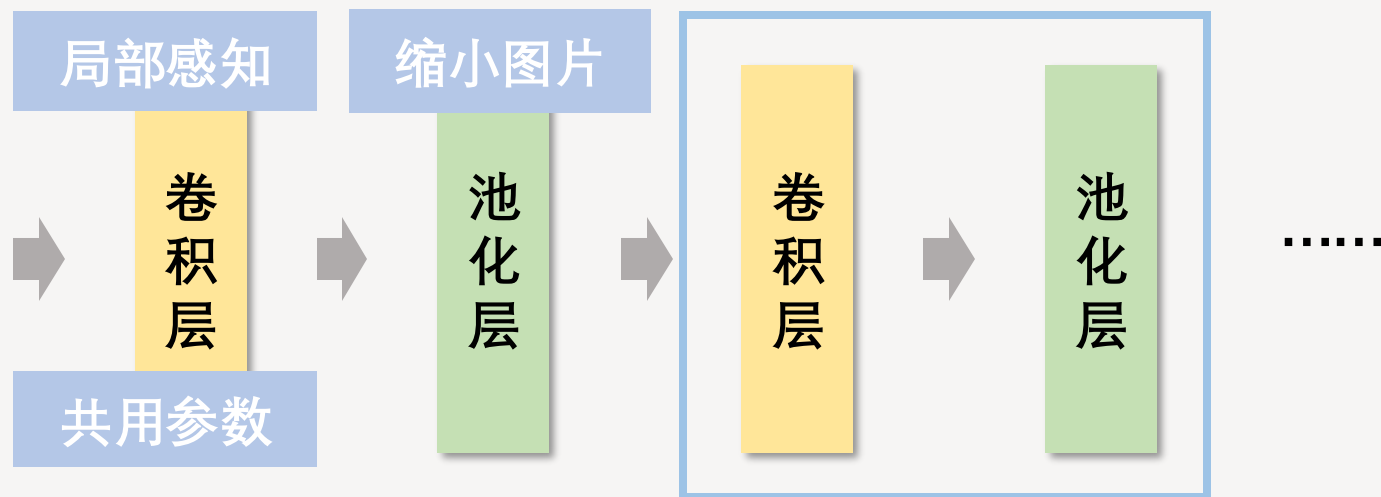
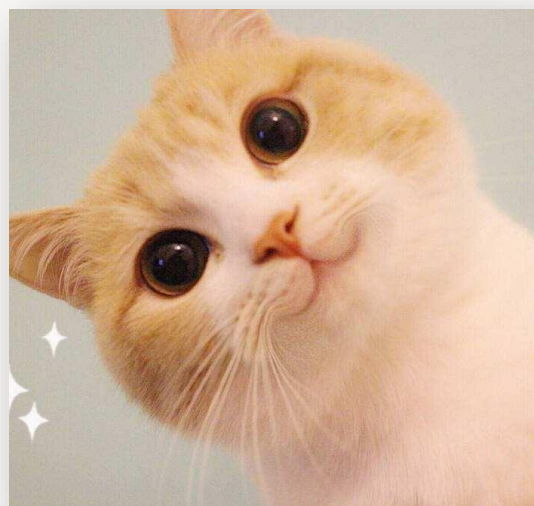
image缩小也可被识别
但可大大减少后续参数计算量

Convolutional Neural Network

卷积神经网络

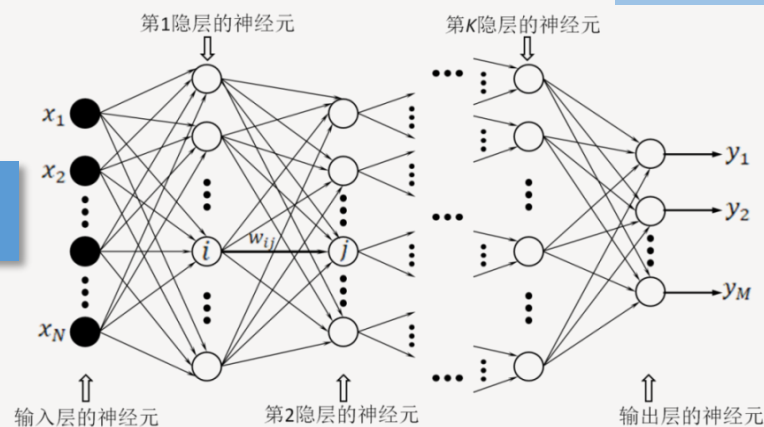
总结来说，CNN的卷积层和池化层使得对图像数据的处理更加高效

可以重复的



Fully Connected 全连接层

猫？狗？羊驼？...

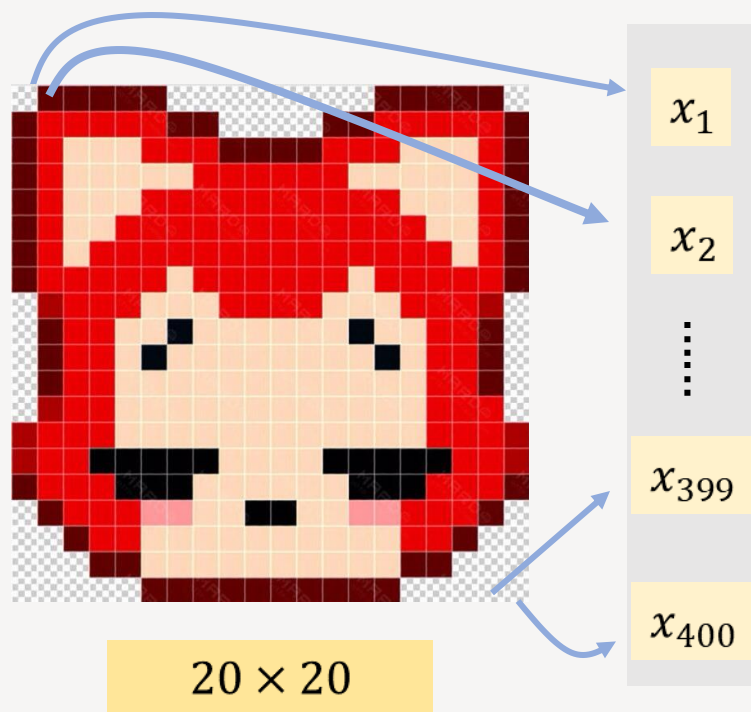


Flatten展开

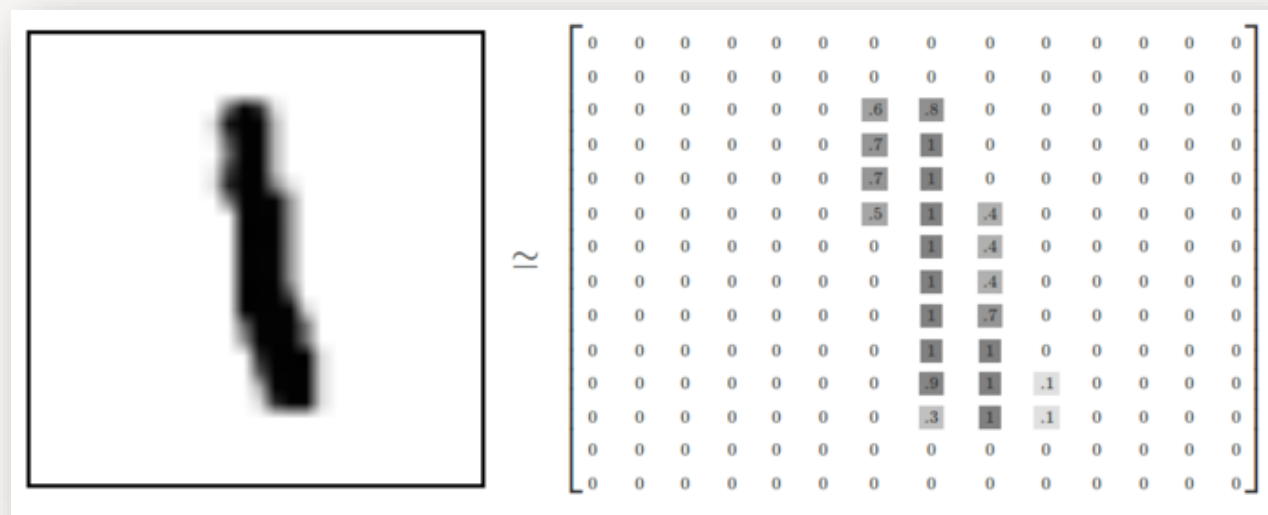
Convolutional Neural Network

卷积神经网络

■ 输入：计算机怎么读取图片？会把图像打碎成像素矩阵，每个输入都是像素的颜色码。



全连接神经网络的输入



保持图片的2D\3D的shape
如果是彩色图，会有RGB三个通道，一个点储存3个值。

卷积神经网络的输入

CNN 卷积神经网络

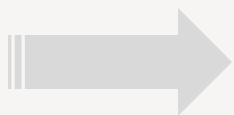
卷积层 (Convolutional Layer) 如何操作？

- ① **卷积核(kernel):** 可以理解成NN中的神经元，只不过只和局部信息进行连接。

卷积方法？卷积核在整张输入图片上滑动平移，与覆盖的部分做点积，得到Feature Map。

1	0	1
0	1	0
1	0	1

3*3
kernel



1 _{x1}	1 _{x0}	1 _{x1}	0	0
0 _{x0}	1 _{x1}	1 _{x0}	1	0
0 _{x1}	0 _{x0}	1 _{x1}	1	1
0	0	1	1	0
0	1	1	0	0

Image

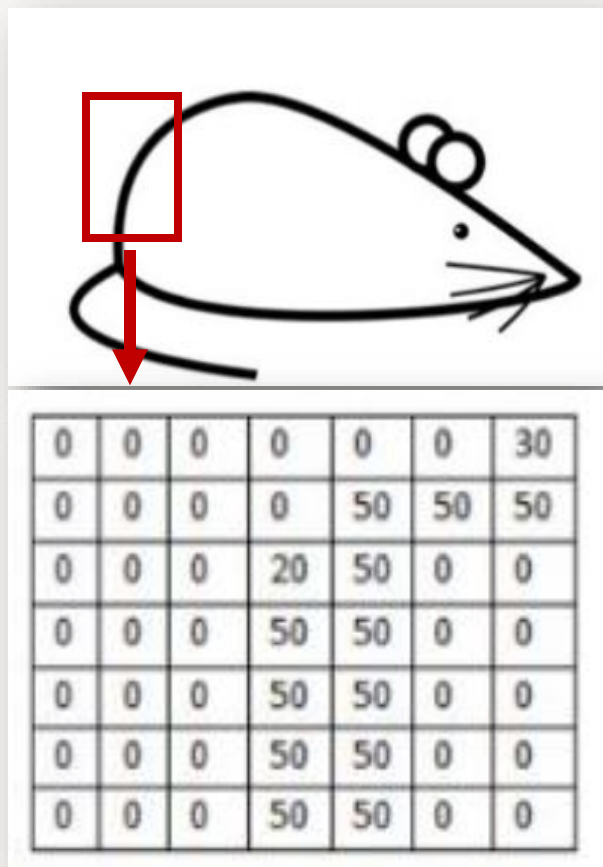
4		

Convolved
Feature

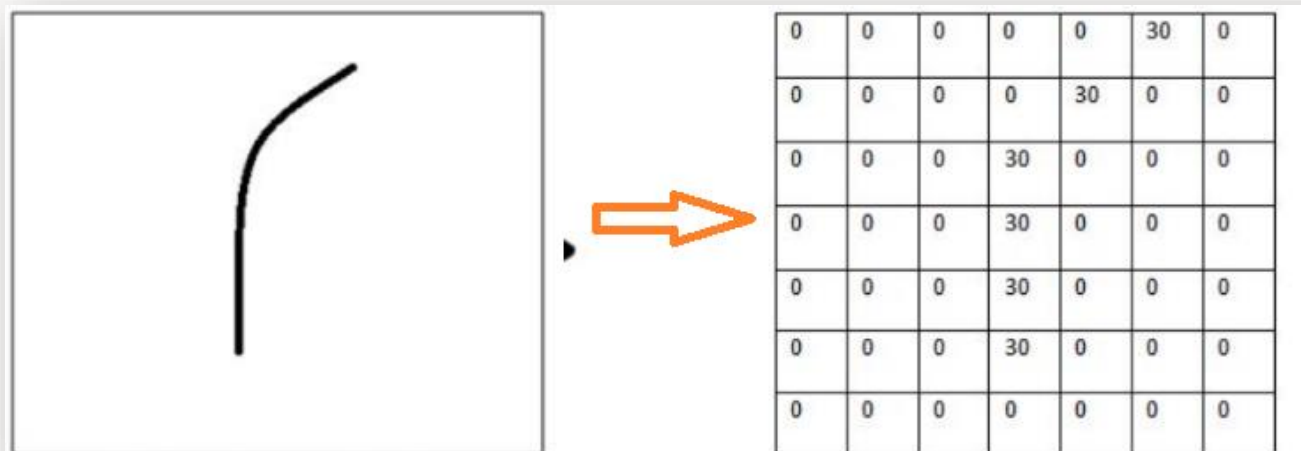
- 从图像的左上方开始一步步滑动至图像右下方为止，按位相乘并相加，得到一张feature map。
- 卷积过程中权值共享。
- 一个3*3的卷积核对应3*3个weights，这些就是你需要训练的参数（也可以有偏置bias）。

CNN 卷积神经网络

为什么这样卷积就可以提取特征呢？**每一个卷积核都可以表达一种特定的图像特点**



如果想要提取尾巴的特征，我们可以用这样的卷积核..

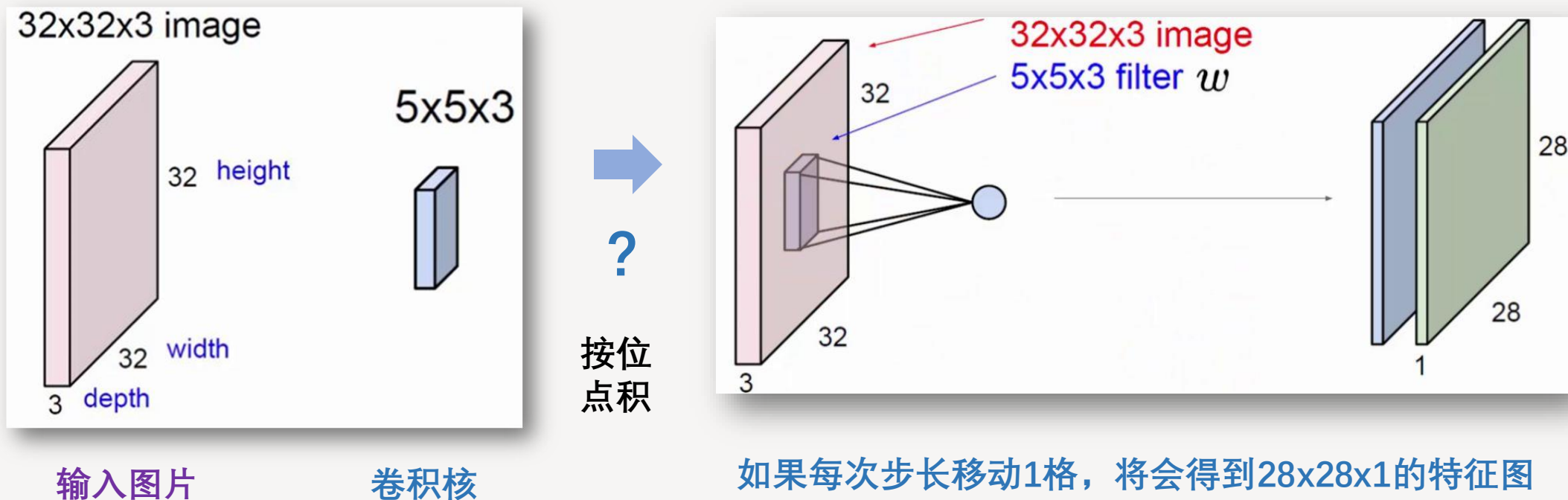


用这个卷积核对图片进行卷积计算后，
会发现期望识别的尾巴处计算出来的值非常大。

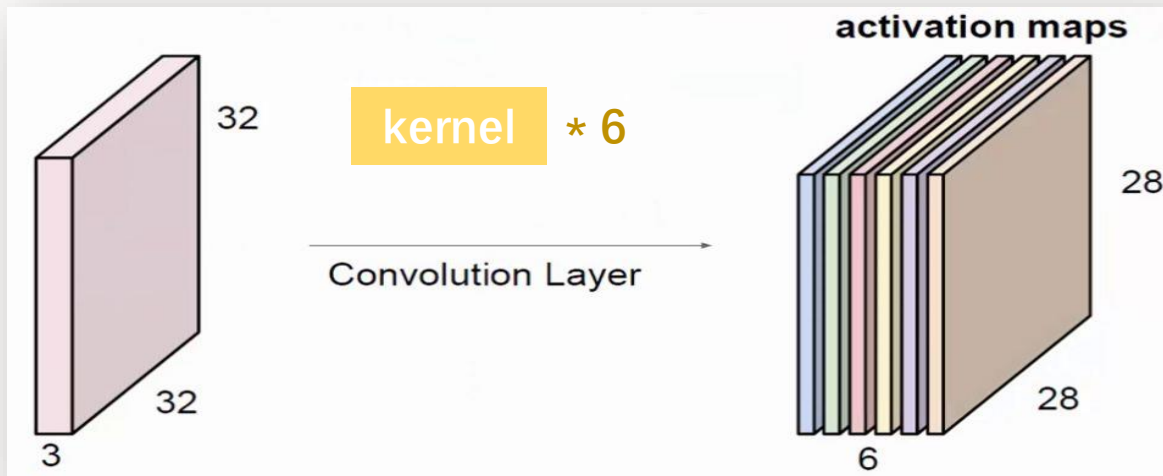
CNN 卷积神经网络

卷积层（Convolutional Layer）运算的目的是提取输入特征
我们往往运用多个卷积核，来产生多张Feature Map

以彩色图像（RGB三个通道）为例，为了提取每个通道特征，卷积核同样也要有深度。



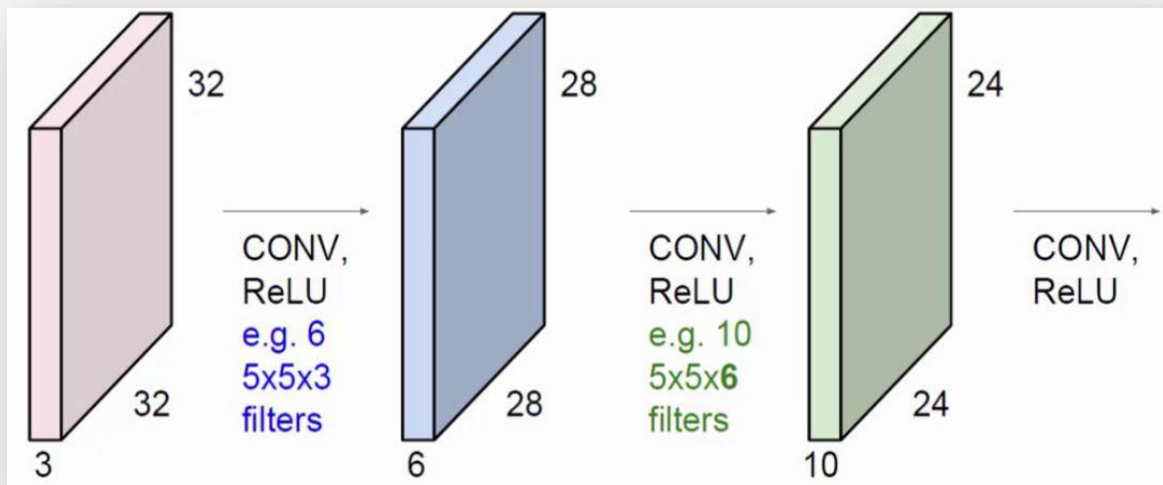
CNN 卷积神经网络



output : $28 \times 28 \times 6$

- 使用6个不同的kernel进行卷积，得到6张不同feature map。
- 将6层叠在一起，作为下一次的输入特征图。

对同一输入的一系列卷积运算定义为一个卷积层



多层卷积运算

每一次卷积的kernel深度与上一层输出的Feature Map深度相等。

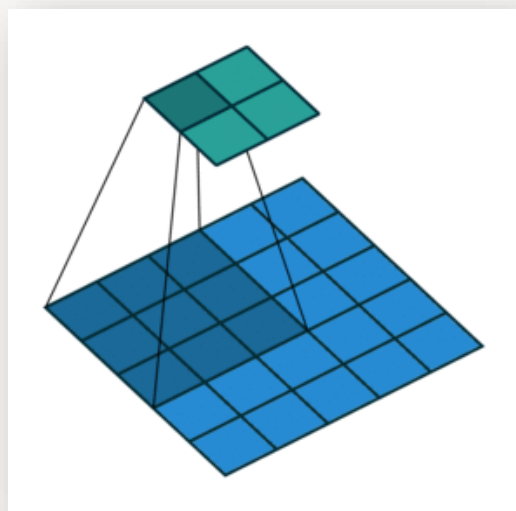
- 使用了6个5x5x3的kernel进行卷积，得到蓝色方块输出。
- 再使用10个5x5x6的kernel进行卷积，得到绿色方块输出。

CNN 卷积神经网络

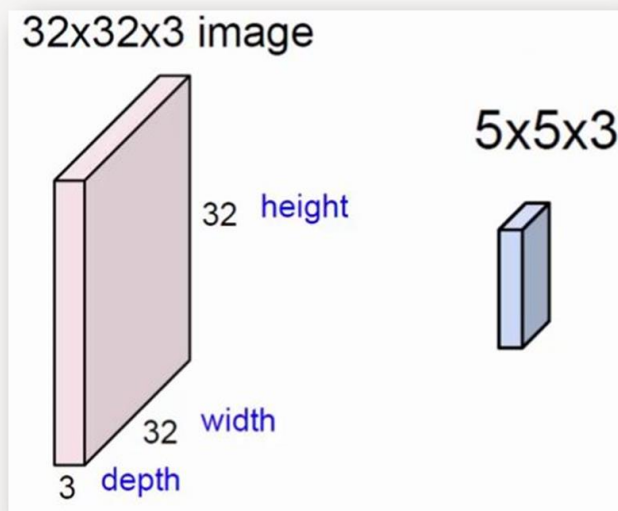
关于卷积层 (Convolutional Layer)

② 步幅 (Stride)：滑动的步长。它控制卷积核在输入图像上以怎样的步幅移动。

- Stride越小，卷积重叠部分较多，提取的特征多，但时间效率会降低。
- Stride越大，重叠区域越少，Feature Map变小，可能漏掉图像上的信息。



3*3的kernel, stride=2



输入图片

卷积核

如果步幅为3，padding=0
输出大小为？

10x10x1

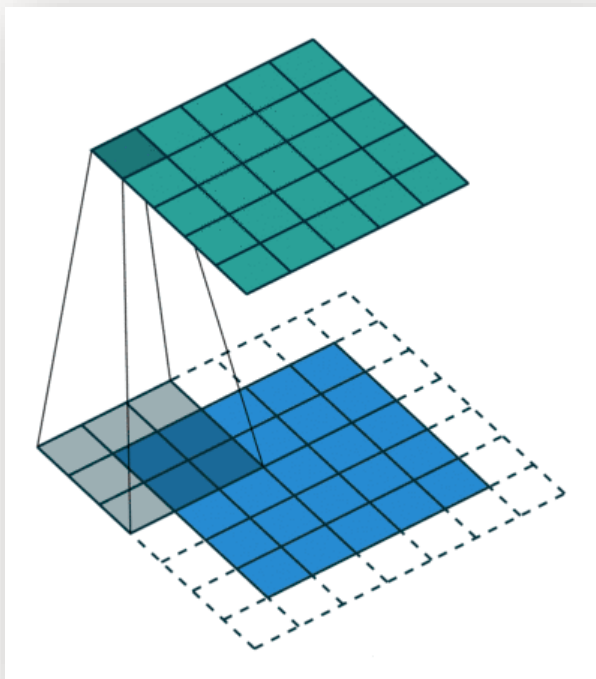
输出长、宽度为：

$$\frac{width - filter + 2pad}{stride} + 1$$

CNN 卷积神经网络

关于卷积层 (Convolutional Layer)

- ③ **填充 (padding)**：周围单元补0，扩大图片的输入尺寸，影响图片输出尺寸。
实际操作中，可根据自己的需要设置不同的padding值。



padding的用途

- 保持边界信息：如果不加padding，输入图片中间的像素点会被扫描多次，而最边缘的像素点只会被卷积一次，边缘信息容易被忽略。加入padding之后，实际处理过程中会**从新的边界**进行操作。
- 利用padding对输入尺寸有差异图片**进行补齐**，使得输入图片尺寸保持一致。
- 加入padding，可以**改变卷积层的输出size**，保持一致or变大扩展。

CNN 卷积神经网络

关于卷积层 (Convolutional Layer)

Stride = 1, padding='SAME', 输入输出size一致

- ③ 填充 (padding)：随着卷积次数的增多，特征图输出尺寸会越来越小，我们通来padding来控制输出特征图的大小。

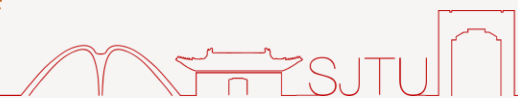
还有一些需要padding的情况，如：5*5的输入图，设置2*2的kernel，stride=2 时，如果不做padding补0，边缘数据无法被卷积，会造成数据丢失。

$$\begin{array}{|c|c|} \hline 1 & 1 \\ \hline 1 & 1 \\ \hline \end{array} \begin{array}{|c|c|} \hline 1 & 1 \\ \hline 1 & 1 \\ \hline \end{array} \begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \end{array} + \begin{array}{|c|c|} \hline -1 & 0 \\ \hline 0 & -1 \\ \hline \end{array} = \begin{array}{cc} -2 & -2 \\ -2 & -2 \end{array}$$

MindSpore: pad_mode='valid'

$$\begin{array}{|c|c|} \hline 1 & 1 \\ \hline 1 & 1 \\ \hline \end{array} \begin{array}{|c|c|} \hline 1 & 1 \\ \hline 1 & 1 \\ \hline \end{array} \begin{array}{c} 1 \\ 1 \\ 1 \\ 0 \end{array} + \begin{array}{|c|c|} \hline -1 & 0 \\ \hline 0 & -1 \\ \hline \end{array} = \begin{array}{ccc} -2 & -2 & -1 \\ -2 & -2 & -1 \\ -1 & -1 & -1 \end{array}$$

MindSpore: pad_mode='same'



以下是用MindSpore框架搭建卷积神经网络的一部分：

已知：

输入图片尺寸为：32*32*3

卷积核大小为5*5*3

卷积核数量为64

stride = 1

Pad_mode = " same"

请问，输出的feature map大小是： [填空1]

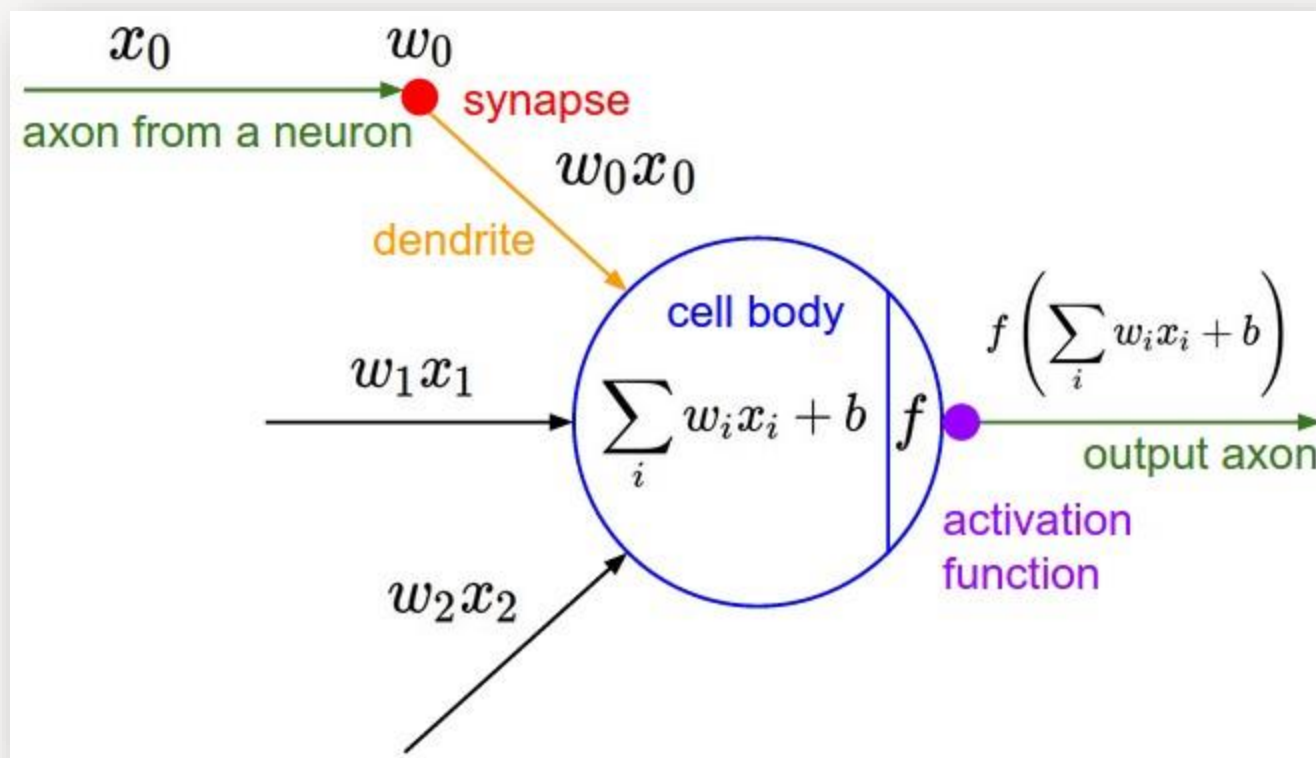
正常使用填空题需3.0以上版本雨课堂

作答

CNN 卷积神经网络

激活函数：在卷积操作之后用非线性函数进行激活。

激活函数（Activation function）一般使用线性整流（ReLU）函数



卷积运算

权重与输入的点积，加上偏置



激活函数进行非线性变化



与全连接神经网络是一样的

CNN 卷积神经网络

池化层 (Pooling Layer)： 特征压缩（无新参数）

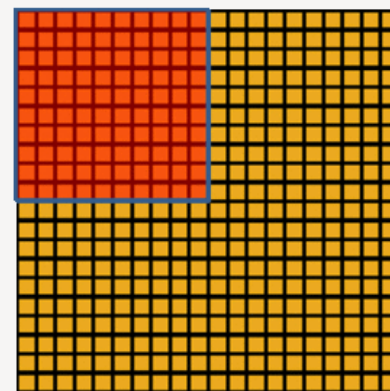
做法：将特征切成几个区域，取其最大值或平均值，得到新的、维度较小的特征。

Why pooling?

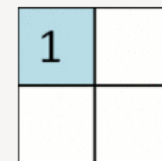
在卷积获得了特征之后，
直接利用这些特征去分类，
计算量将会非常大

因此，适当减少参数去描述较大的图像，对图像进行压缩（计算的一块区域的均值或最大值）

不仅可以降低维度，同时还会改善过拟合的结果
这种聚合的操作就是池化 (pooling)



Convolved
feature

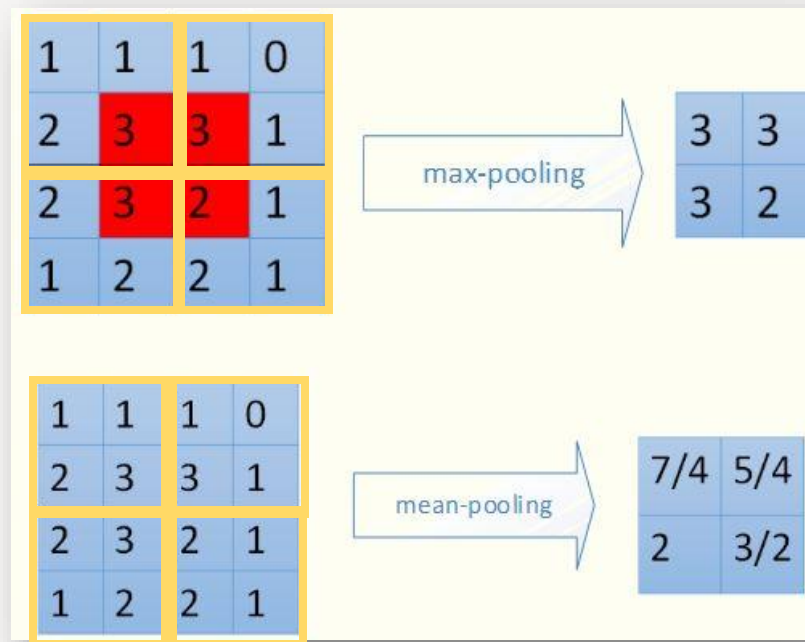
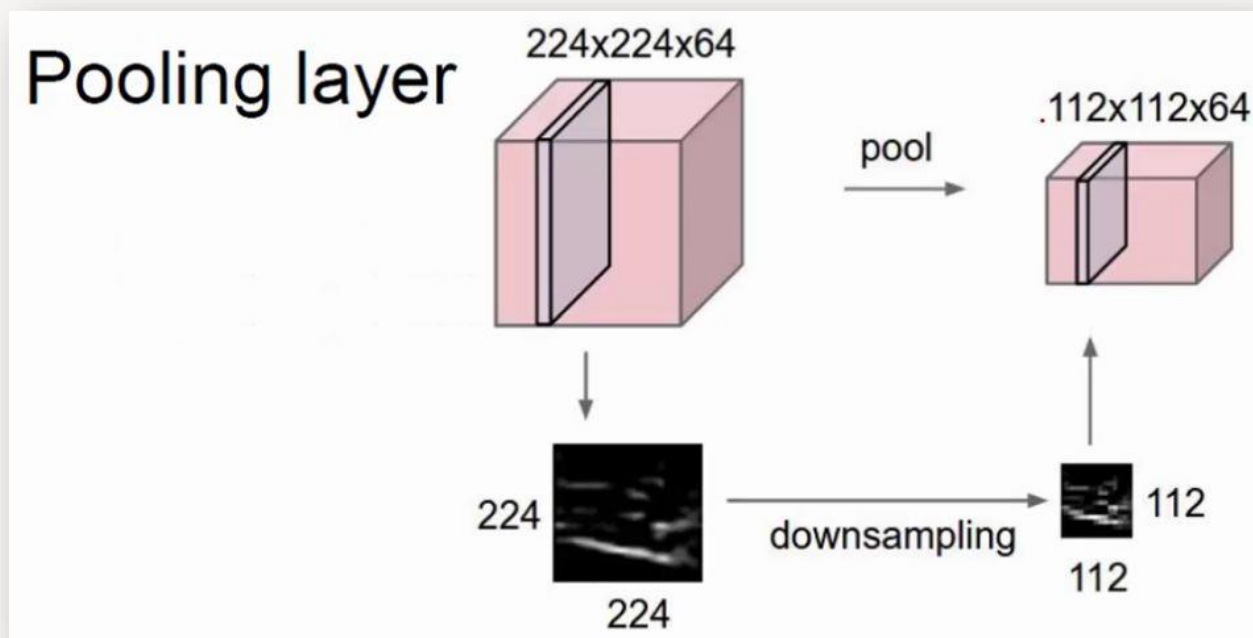


Pooled
feature

CNN 卷积神经网络

■ 池化层 (Pooling Layer) 非常容易理解，并且没有任何新增参数。

将Feature Map划分成几个区域，取其最大值/均值，得到新的、维度较小的特征。



池化层往往只改变尺寸size，不改变深度depth\channel。

池化层后面需要激活吗？



需要



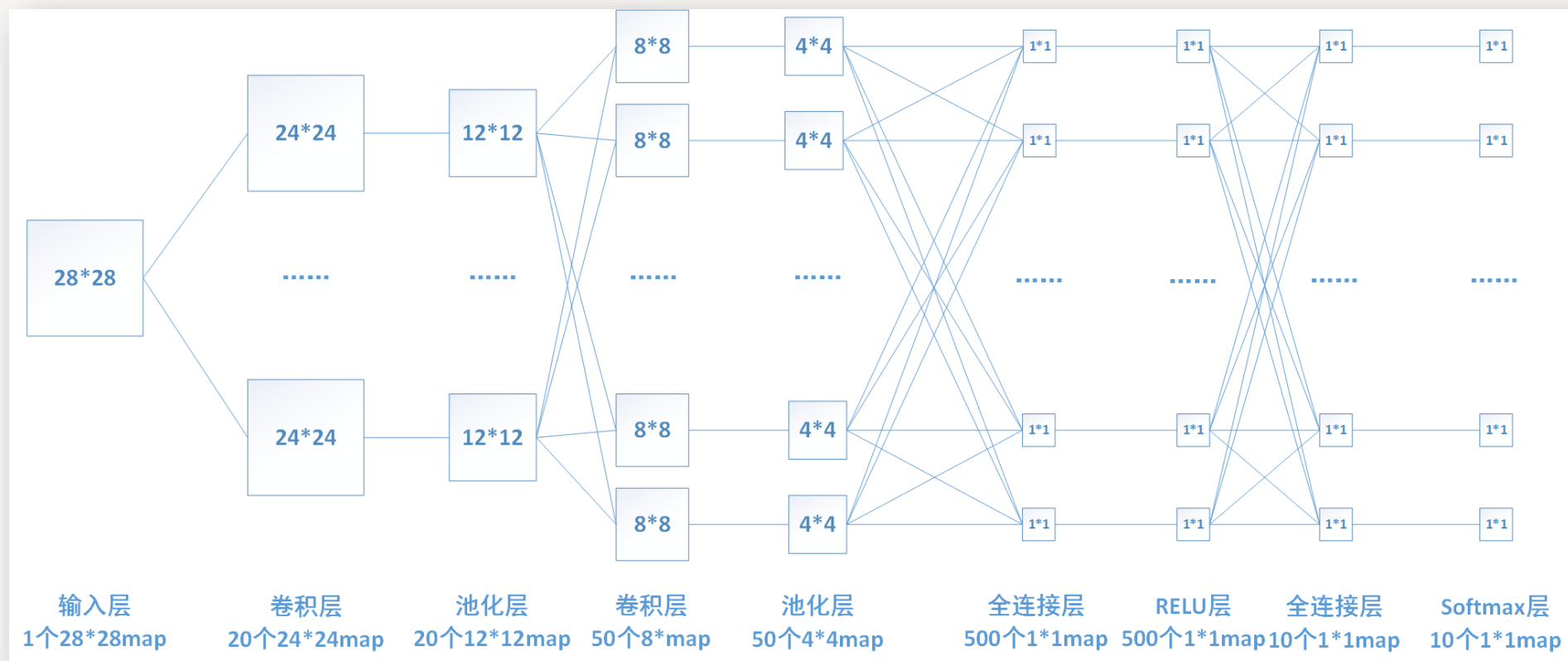
不需要

提交

CNN 卷积神经网络

全连接层 (Fully-connected Layer) 放在卷积和池化之后

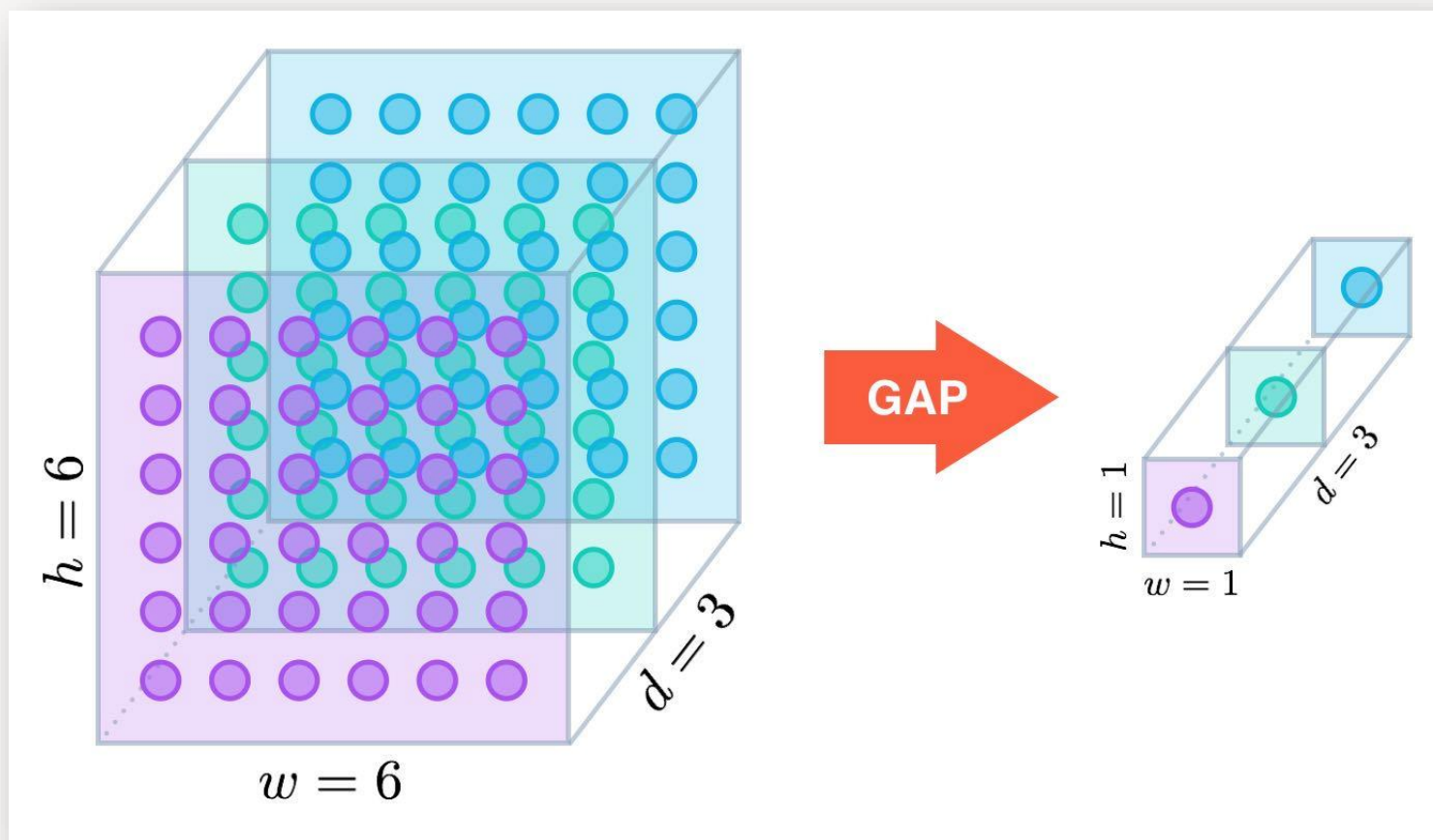
特征提取与降维之后，用全连接层进行最后的分类。



关键在于每一层的
参数理解

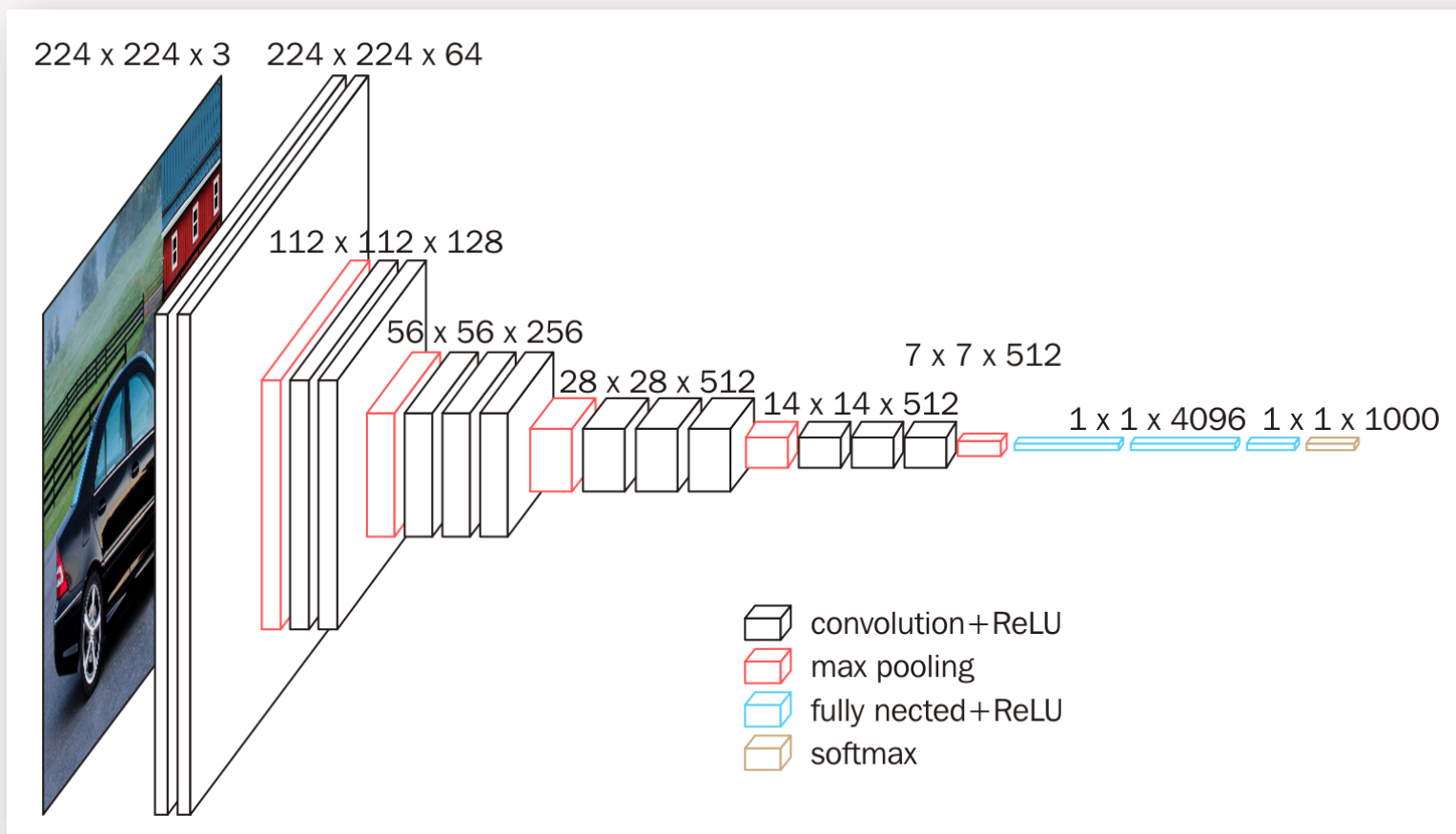
CNN 卷积神经网络

某些pipeline中，最后的FC会被替换为全局池化层Global Average Pooling (GAP)



经典卷积神经网络——VGG16

图为VGG16的网络结构，共16层（不包括池化和softmax层），所有的卷积核都使用3*3的大小，池化都使用大小为2*2，步长为2的最大池化，卷积层深度依次为64 -> 128 -> 256 -> 512 -> 512。



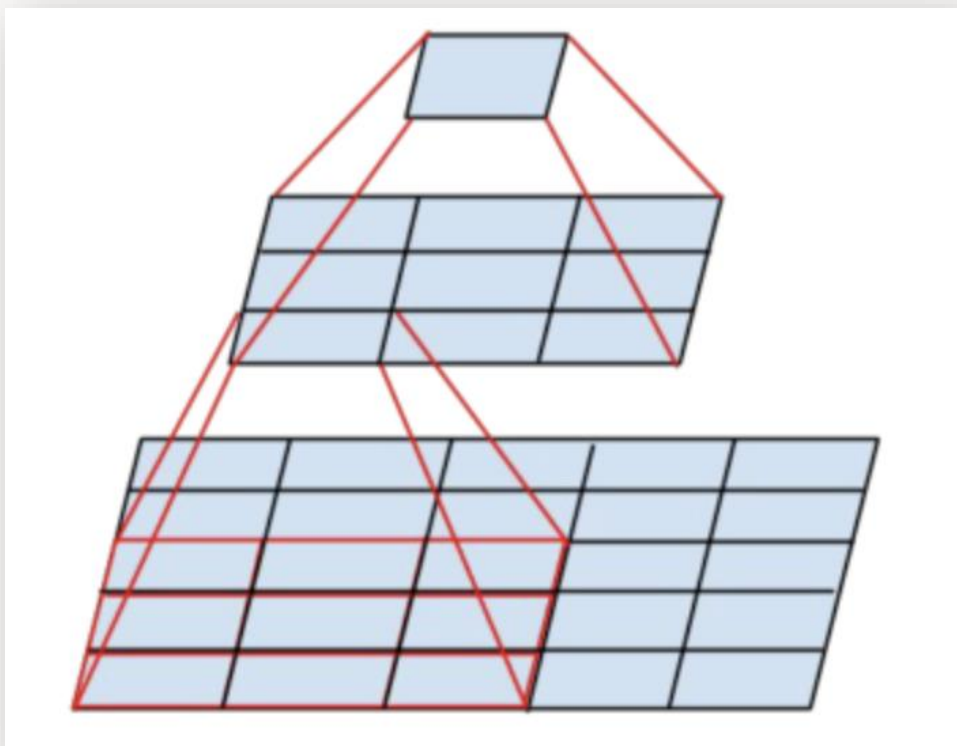
2	Conv-64	
	Conv-64	
	Max-pooling	
2	Conv-128	
	Conv-128	
	Max-pooling	
3	Conv-256	最后三层
	Conv-256	全连接
	Conv-256	FC-4096
	Max-pooling	FC-4096
3	Conv-512	FC-1000
	Conv-512	Softmax
	Conv-512	
3	Conv-512	
	Conv-512	
	Conv-512	
	Max-pooling	

经典卷积神经网络——VGG16

为什么VGG16里所有的卷积核都使用 3×3 的大小？

——几个小滤波器卷积层的组合比一个大滤波器卷积层好

早期的神经网络采用较大的卷积核，如 5×5 ， 7×7 ，这样参数很多，层次少，表达能力弱。



感受野：两个 3×3 卷积核 = 一个 5×5 卷积核

参数：2个 3×3 的卷积核共18个参数
1个 5×5 的卷积核有25个参数

感受野：三个 3×3 卷积核 = 一个 7×7 卷积核

参数：3个 3×3 的卷积核共27个参数
1个 7×7 的卷积核有49个参数

使用 3×3 卷积核节省参数，增加网络层数和深度，层之间的激活函数增加了网络的非线性表达能力

• Thanks •

学生创新中心：肖雄子彦



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



学生创新中心
Student Innovation Center