



卷积神经网络

@八斗学院--王小天(Michael)

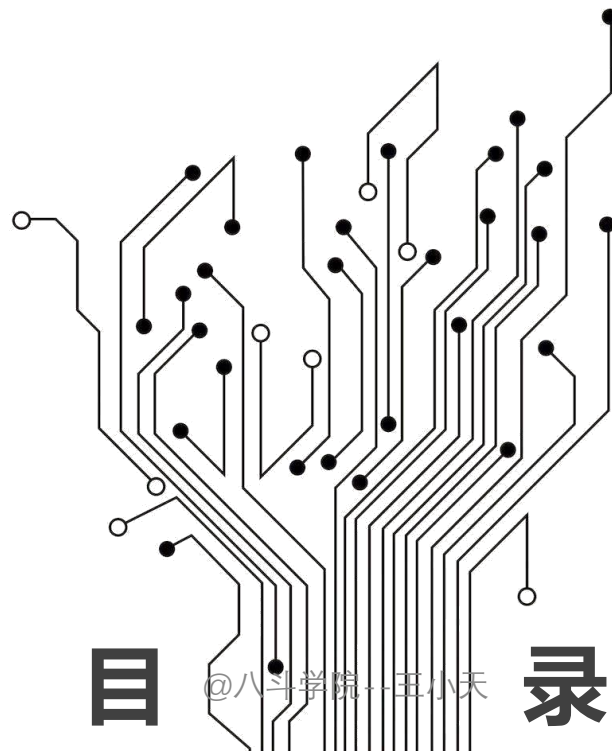
2022/1/23

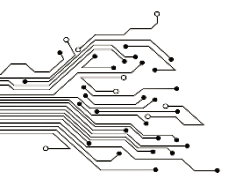
@八斗学院--王小天



---八斗人工智能，盗版必究---

1. 卷积神经网络简介
2. 卷积层、池化层
3. 卷积神经网络
4. cifar-10预测实例
5. Alexnet





卷积神经网络

---八斗人工智能，盗版必究---

CNN就是著名的卷积神经网络，是一种前馈神经网络。

CNN不同于传统的神经网络只有线性连接，CNN包括卷积（convolution）操作、汇合（pooling）操作和非线性激活函数映射（即线性连接）等等。

经典的CNN网络有Alex-Net、VGG-Nets、Resnet等。



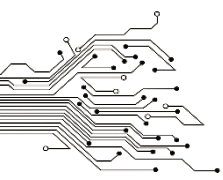
卷积神经网络

---八斗人工智能，盗版必究---

深度学习在计算机图像识别上的应用非常成功。

利用深度学习，我们能够对图片进行高精度识别，实现这一功能的，主要依靠神经网络中的一种分支，名为卷积网络。

卷积网络与我们前面实现的网络不同之处在于，它可以直接接受多维向量，而我们以前实现的网络只能接收一维向量。



卷积--3通道卷积

---八斗人工智能，盗版必究---

Input Volume (+pad 1) (7x7x3)

$x[:, :, 0]$

0	0	0	0	0	0	0
0	1	0	2	2	2	0
0	0	0	0	0	0	0
0	2	0	2	2	2	0
0	1	0	0	0	0	0
0	1	0	0	2	1	0
0	0	0	0	0	0	0

$x[:, :, 1]$

0	0	0	0	0	0	0
0	1	0	1	0	1	0
0	0	0	0	0	1	0
0	0	0	2	0	0	0
0	2	0	0	0	0	0
0	0	0	1	0	0	0
0	0	0	0	0	0	0

$x[:, :, 2]$

0	0	0	0	0	0	0
0	0	2	2	0	0	0
0	0	0	0	0	1	0
0	1	1	2	0	2	0
0	2	0	0	0	0	0
0	0	1	1	0	1	0
0	0	0	0	0	0	0

Filter W0 (3x3x3)

$w0[:, :, 0]$

1	0	0
-1	0	0
0	-1	1

$w0[:, :, 1]$

-1	-1	0
-1	-1	1
1	0	0

$w0[:, :, 2]$

1	1	0
0	-1	1
1	1	1

Bias b0 (1x1x1)

$b0[:, :, 0]$

1

Filter W1 (3x3x3)

$w1[:, :, 0]$

1	-1	0
-1	0	0
0	1	-1

$w1[:, :, 1]$

0	1	-1
0	0	0
0	1	1

$w1[:, :, 2]$

0	-1	0
1	1	-1
-1	0	1

Bias b1 (1x1x1)

$b1[:, :, 0]$

0

Output Volume (3x3x2)

$o[:, :, 0]$

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

$o[:, :, 1]$

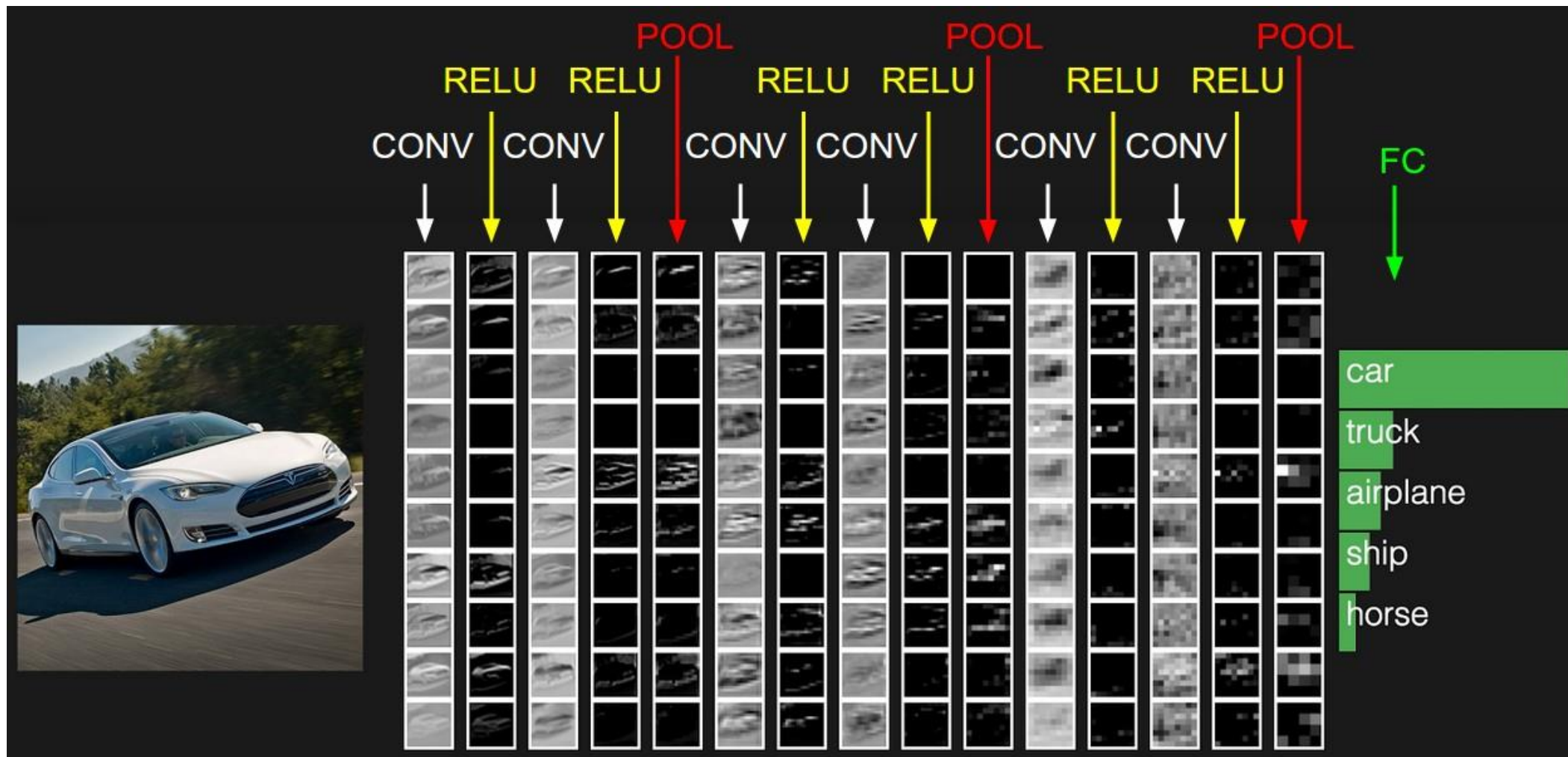
-2	4	-1
3	3	0
-2	2	-1

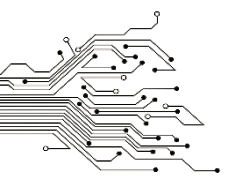
$$\begin{aligned} o(2,2,0) &= \sum x[:, :, 0] \times w[:, :, 0] + \sum x[:, :, 1] \times w[:, :, 1] + \sum x[:, :, 2] \times w[:, :, 2] + b_0 \\ &= 0 \times 1 + 0 \times 0 + 0 \times 0 + 2 \times (-1) + 1 \times 0 + 0 \times 0 + 0 \times 0 + 0 \times (-1) + 0 \times 1 \\ &\quad + 0 \times (-1) + 0 \times (-1) + 0 \times 0 + 0 \times (-1) + 0 \times (-1) + 0 \times 1 + 0 \times 1 + 0 \times 0 + 0 \times 1 \\ &\quad + 1 \\ &= -2 \end{aligned}$$



卷积神经网络 (CNN)

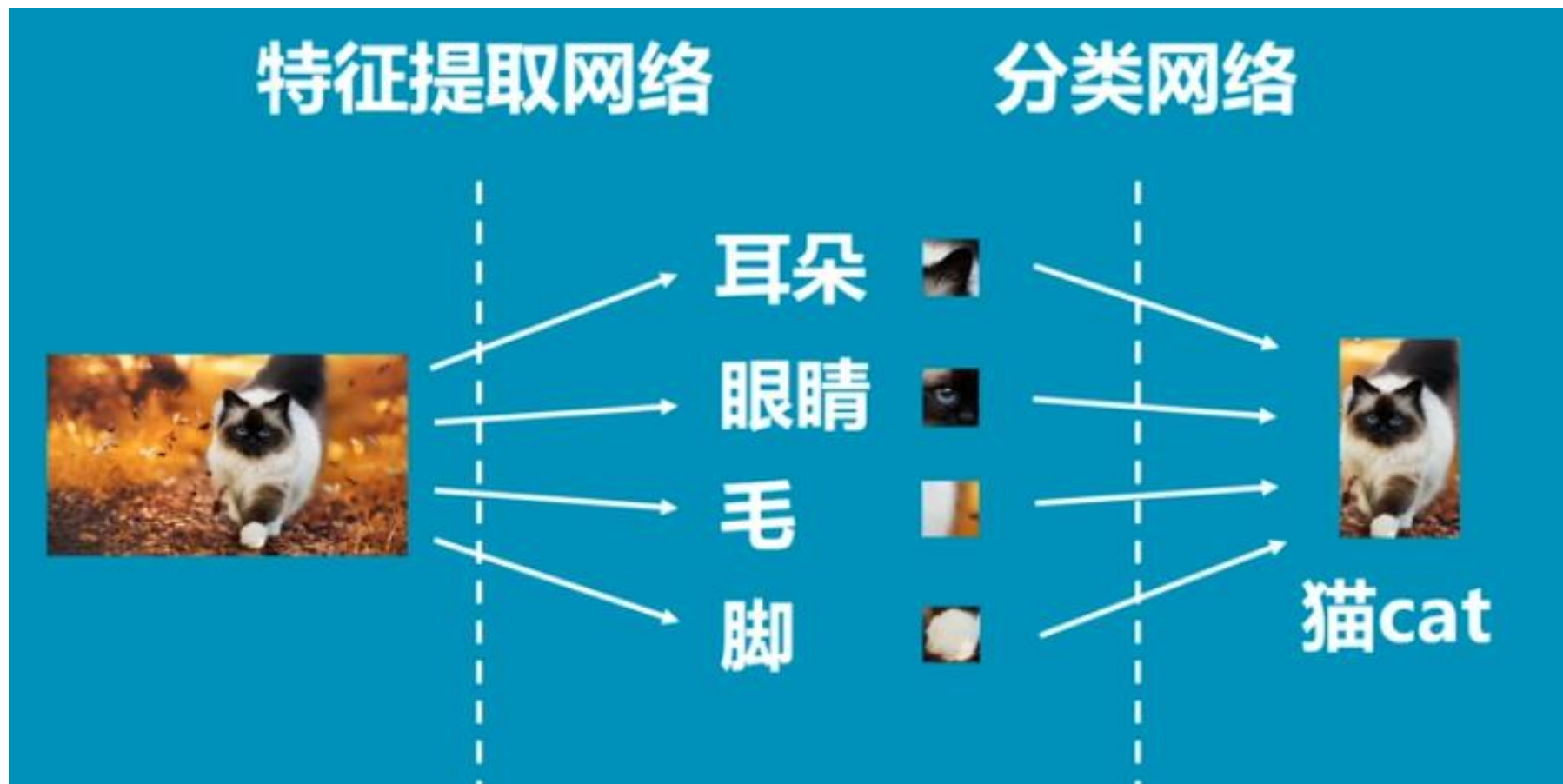
---八斗人工智能，盗版必究---





卷积神经网络 (CNN)

---八斗人工智能，盗版必究---

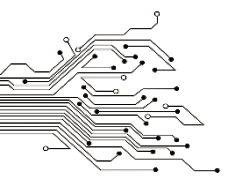




卷积

---八斗人工智能，盗版必究---

- 卷积操作，其实是把一张大图片分解成好多个小部分，然后依次对这些小部分进行识别。
- 通常我们会把一张图片分解成多个 3×3 或 5×5 的“小片”，然后分别识别这些小片段，最后把识别的结果集合在一起输出给下一层网络。
- 这种做法在图象识别中很有效。因为它能对不同区域进行识别，假设识别的图片是猫脸，那么我们就可以把猫脸分解成耳朵，嘴巴，眼睛，胡子等多个部位去各自识别，然后再把各个部分的识别结果综合起来作为对猫脸的识别。

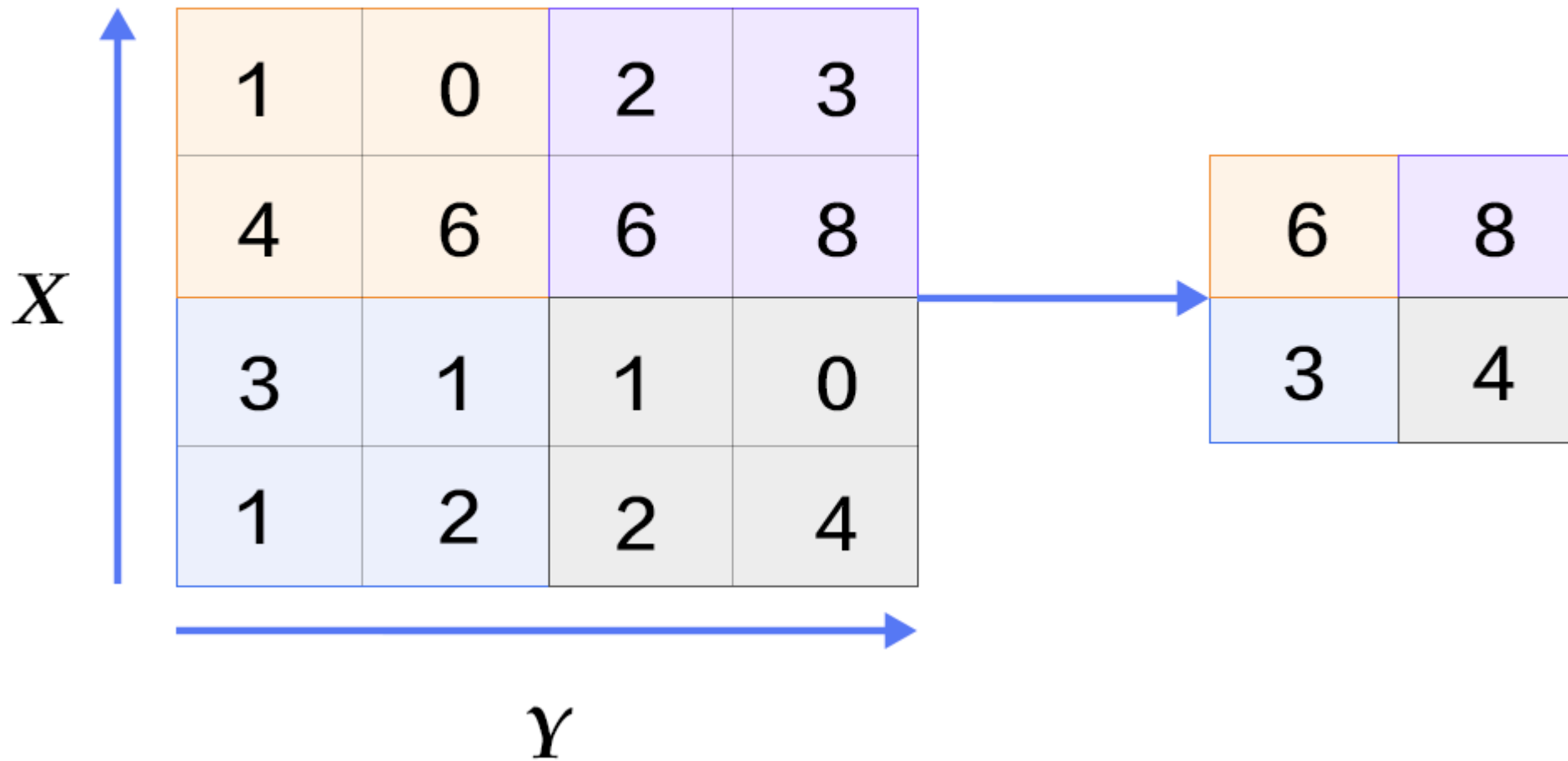


池化层-pool

---八斗人工智能，盗版必究---

Max pool

输入数据



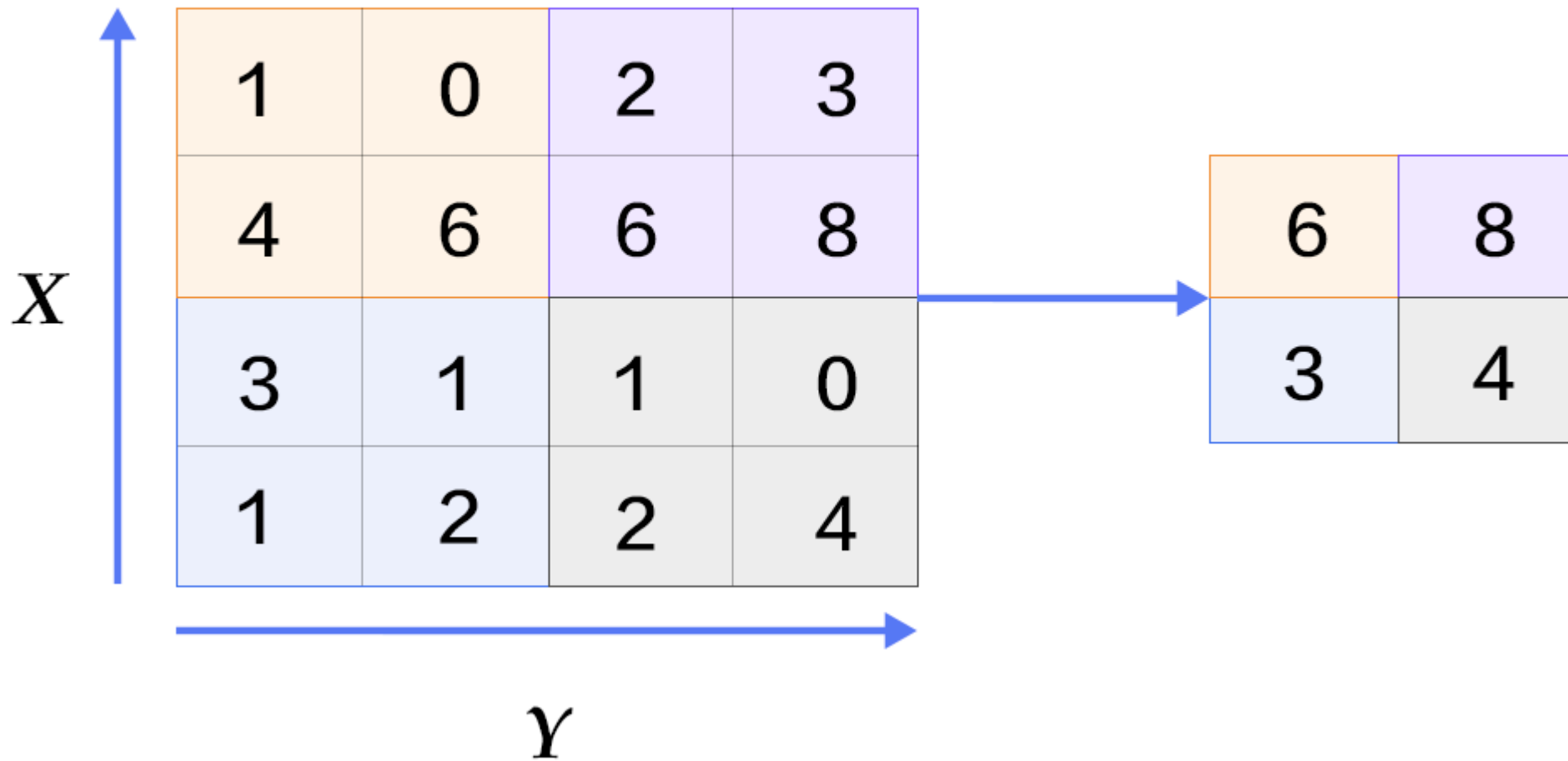


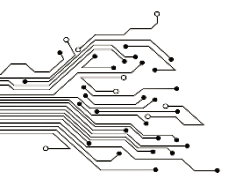
池化层-pool

---八斗人工智能，盗版必究---

Max pool

输入数据



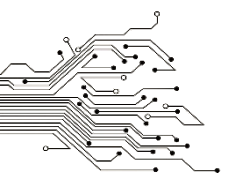


池化层-pool

---八斗人工智能，盗版必究---

池化的作用：

卷积操作产生了太多的数据，如果没有max pooling对这些数据进行压缩，那么网络的运算量将会非常巨大，而且数据参数过于冗余就非常容易导致过度拟合。



卷积神经网络

---八斗人工智能，盗版必究---

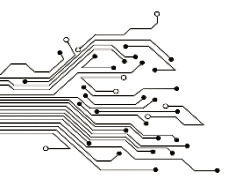
当我们的图片（黑白图片厚度为1，彩色图片厚度为3）输入到神经网络后，我们会通过卷积神经网络将图片的长和宽进行压缩，然后把厚度增加。最后就变成了一个长宽很小，厚度很高的像素块。然后结果放入普通的神经网络（全连接）中处理，最后链接一个分类器（比如softmax），从而分辨出图片是什么。



卷积核

图片的采样器也可以叫做共享权值，用来在图片上采集信息。卷积核有自己的长宽，也可以定义自己的步长stride，每跨多少步进行一次抽离信息，跨的步长越多就越容易丢失图片信息。然后对抽取的信息进行像素的加权求和得到Feature Map 增加了采集结果的厚度。

总而言之 卷积是用来不断的提取特征，每提取一个特征就会增加一个feature map，所以采集后的图片厚度不断变厚



卷积神经网络

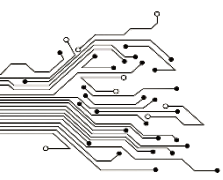
---八斗人工智能，盗版必究---

我们学过高斯滤波、sobel滤波等等设定好卷积核的滤波方法。

这时我们不禁要想，如果不是由人来设计一个滤波器，而是从一个随机滤波器开始，根据某种目标、用某种方法去逐渐调整它，直到它接近我们想要的样子，可行么？

这就是卷积神经网络（Convolutional Neural Network, CNN）的思想了。

可调整的滤波器是CNN的“卷积”那部分；如何调整滤波器则是CNN的“神经网络”那部分（训练）。

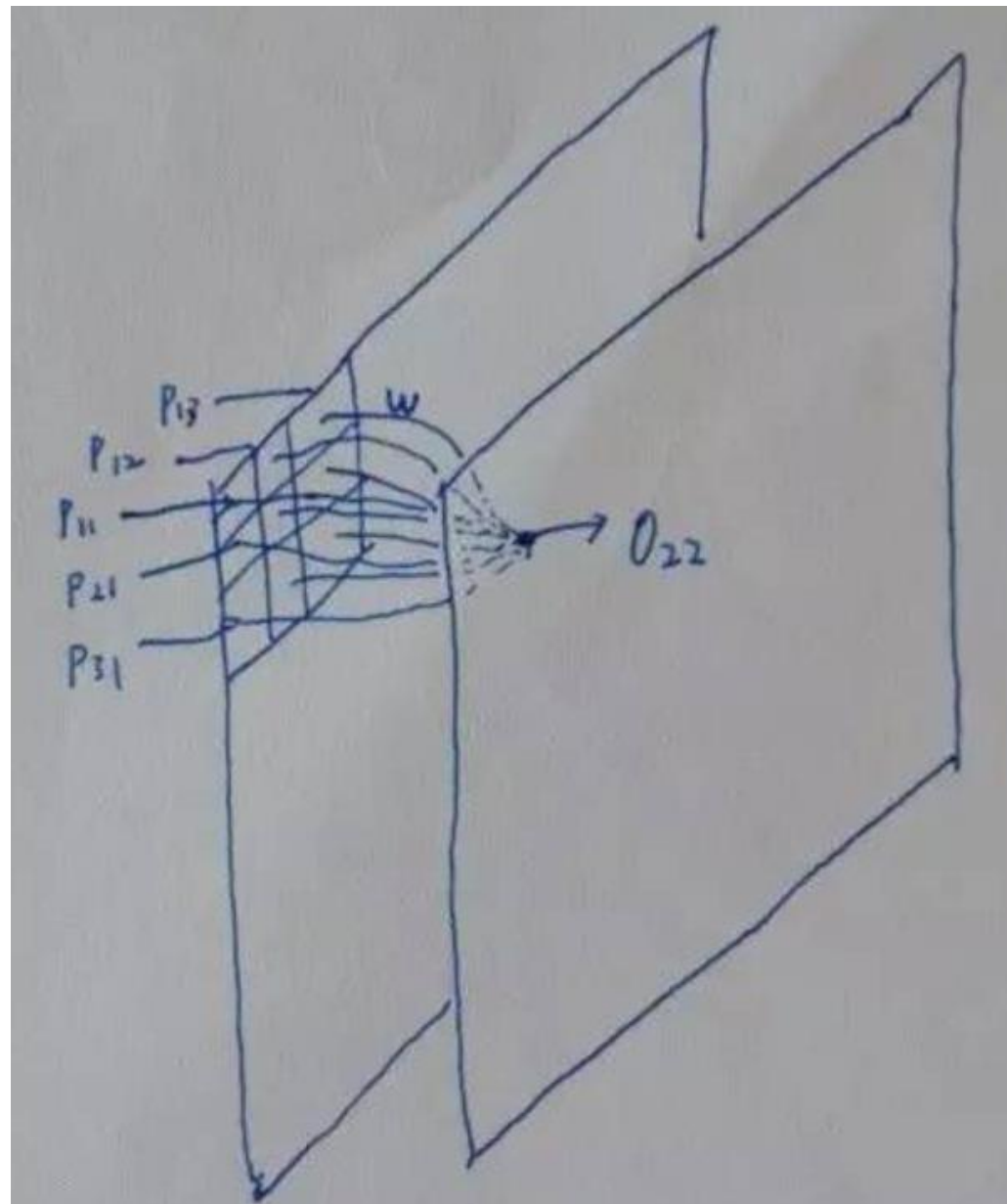


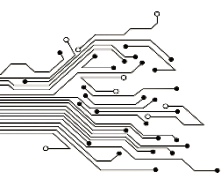
卷积神经网络

---八斗人工智能，盗版必究---

把卷积滤波器和神经网络两个思想结合起来。
卷积滤波器无非就是一套权值。而神经网络也可以有
(除全连接外的) 其它拓扑结构。

- 左边的平面包含 $n \times n$ 个格子，每个格子中是一个 $[0, 255]$ 的整数值，它就是输入图像，也是这个神经网络的输入。
- 右边的平面也是 $n \times n$ 个格子，每个格子是一个神经元。每个神经元根据二维位置关系连接到输入上它周围 3×3 范围内的值。
- 每个连接有一个权值 w 。
- 所有神经元都如此连接 (图中只画了一个，出了输入图像边缘的连接就认为连接到常数 0, *pad*)。
- 右边层的神经元的输出就是该神经网络的输出。

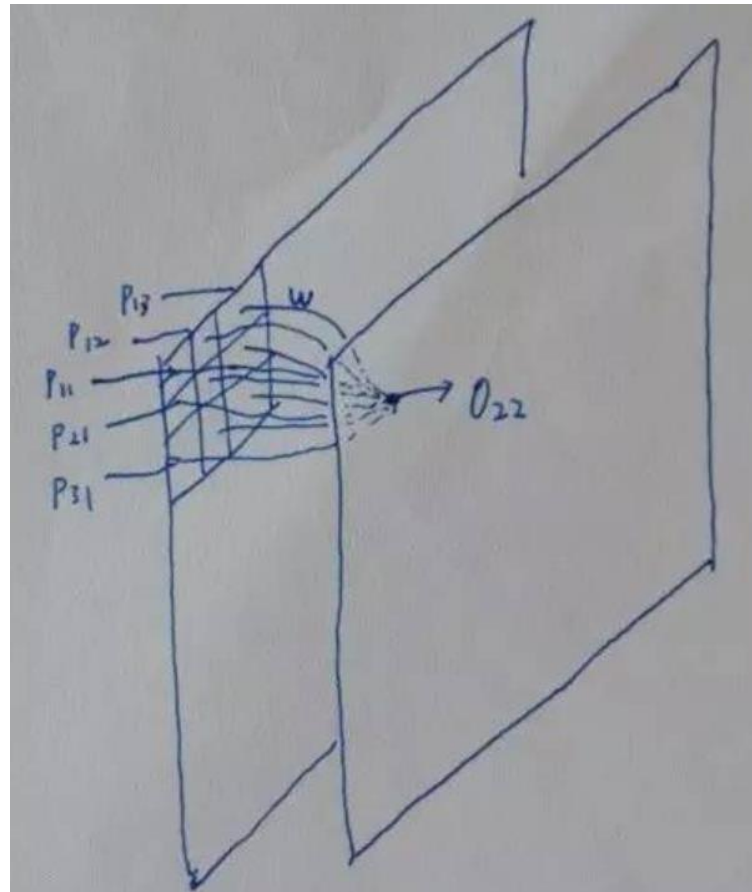




卷积神经网络

这个网络有两点与全连接神经网络不同：

1. 它不是全连接的：右层的神经元并非连接上全部输入，而是只连接了一部分。这里的一部分就是输入图像的一个局部区域。我们常听说 CNN 能够把握图像局部特征、alphaGO 从棋局局部状态提取信息等等，就是这个意思。这样一来权值少了很多，因为连接少了。
2. 权值其实还更少，因为每一个神经元的9个权值都是和其他神经元共享的。全部 $n*n$ 个神经元都用这共同的一组9个权值。那么**这个神经网络其实一共只有9个参数需要调整。**



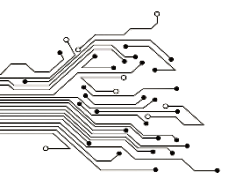


卷积神经网络

---八斗人工智能，盗版必究---

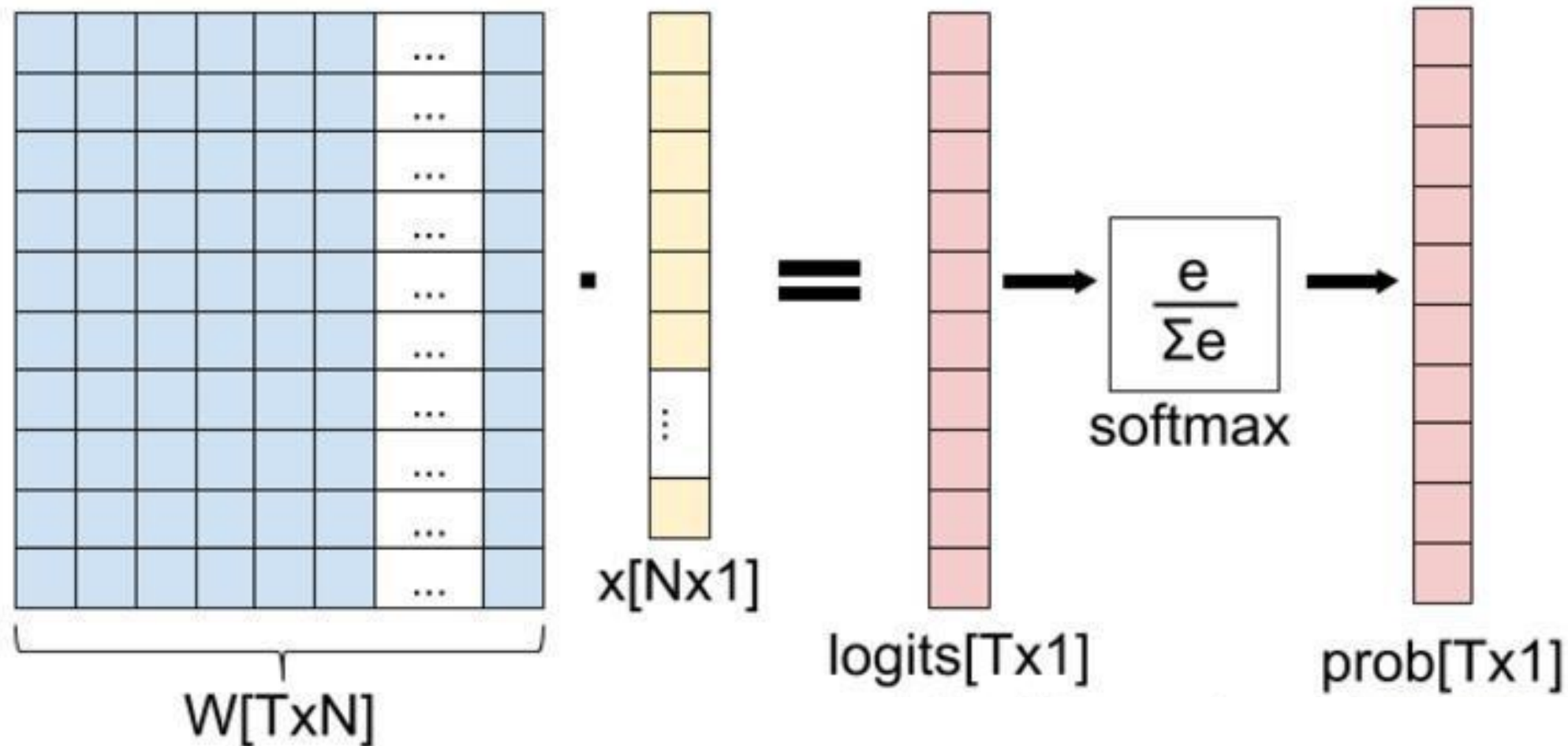
综上，
这个神经网络不就是一个卷积滤波器么？只不过卷积核的参数未定，需要我们去训练——它是一个“可训练滤波器”。

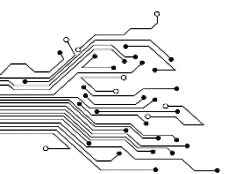
这个神经网络就已经是一个拓扑结构特别简单的 CNN 了。



卷积神经网络

---八斗人工智能，盗版必究---





卷积神经网络实战--Cifar训练集

---八斗人工智能，盗版必究---

Cifar-10数据集包含10类共60000张32*32的彩色图片， 每类6000张图。包括50000张训练图片和10000张测试图片

Here are the classes in the dataset, as well as 10 random images from each:

airplane



automobile



bird



cat



deer



dog



frog



horse



ship



truck

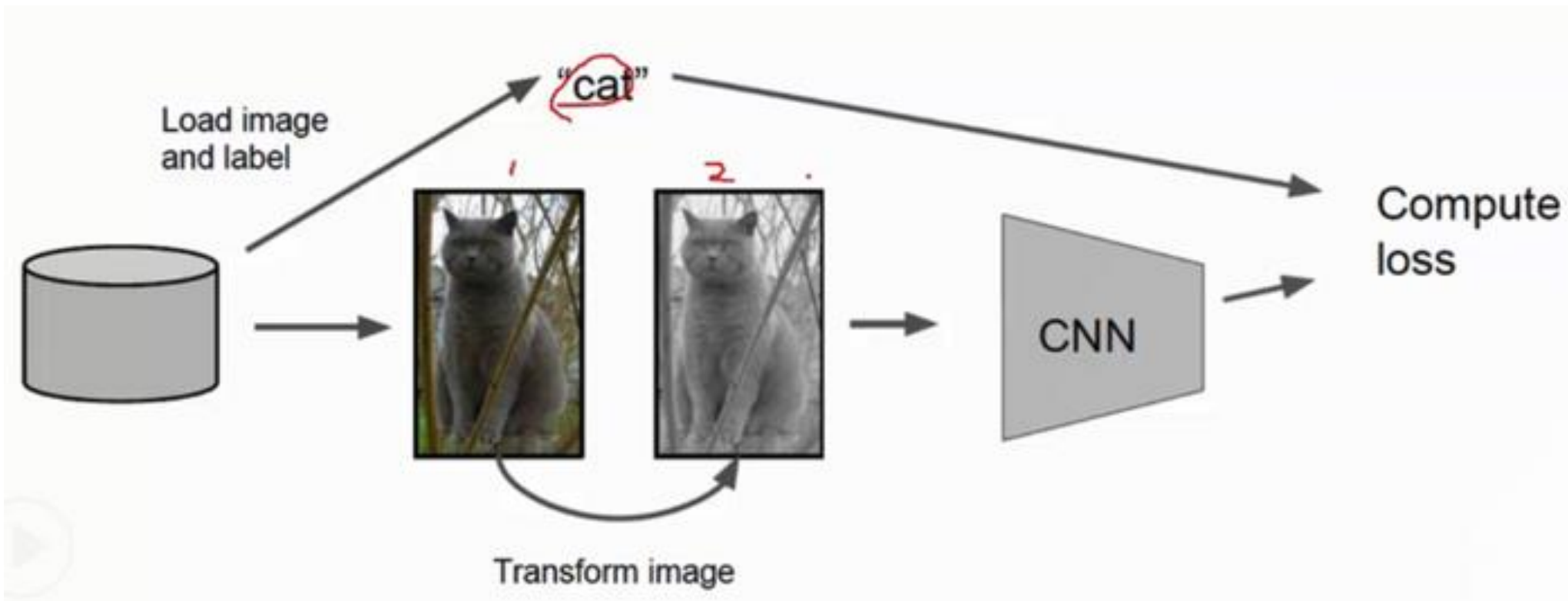




数据预处理--图像增强

由于深度学习对数据集的大小有一定的要求，若原始的数据集比较小，无法很好地满足网络模型的训练，从而影响模型的性能，而图像增强是对原始图像进行一定的处理以扩充数据集，能够在一定程度上提升模型的性能。

图像增强表示的是，在原始图像的基础上，对数据进行一定的改变，增加了数据样本的数量，但是数据的标签值并不发生改变。



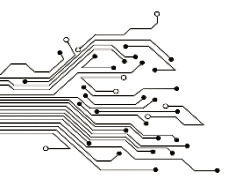


数据预处理--图像增强

这里讲的图像增强指的是突出图像中感兴趣区域及特征。

图像增强分为两种：

1. 增强“自我”：通过一定手段将感兴趣区域增强，直至从图像中脱颖而出的那种，也是正常思维下常用的方法。
2. 削弱“别人”：是增强“自我”的反方法，指的是通过一定手段将不感兴趣区域削弱，直至感兴趣区域脱颖而出。



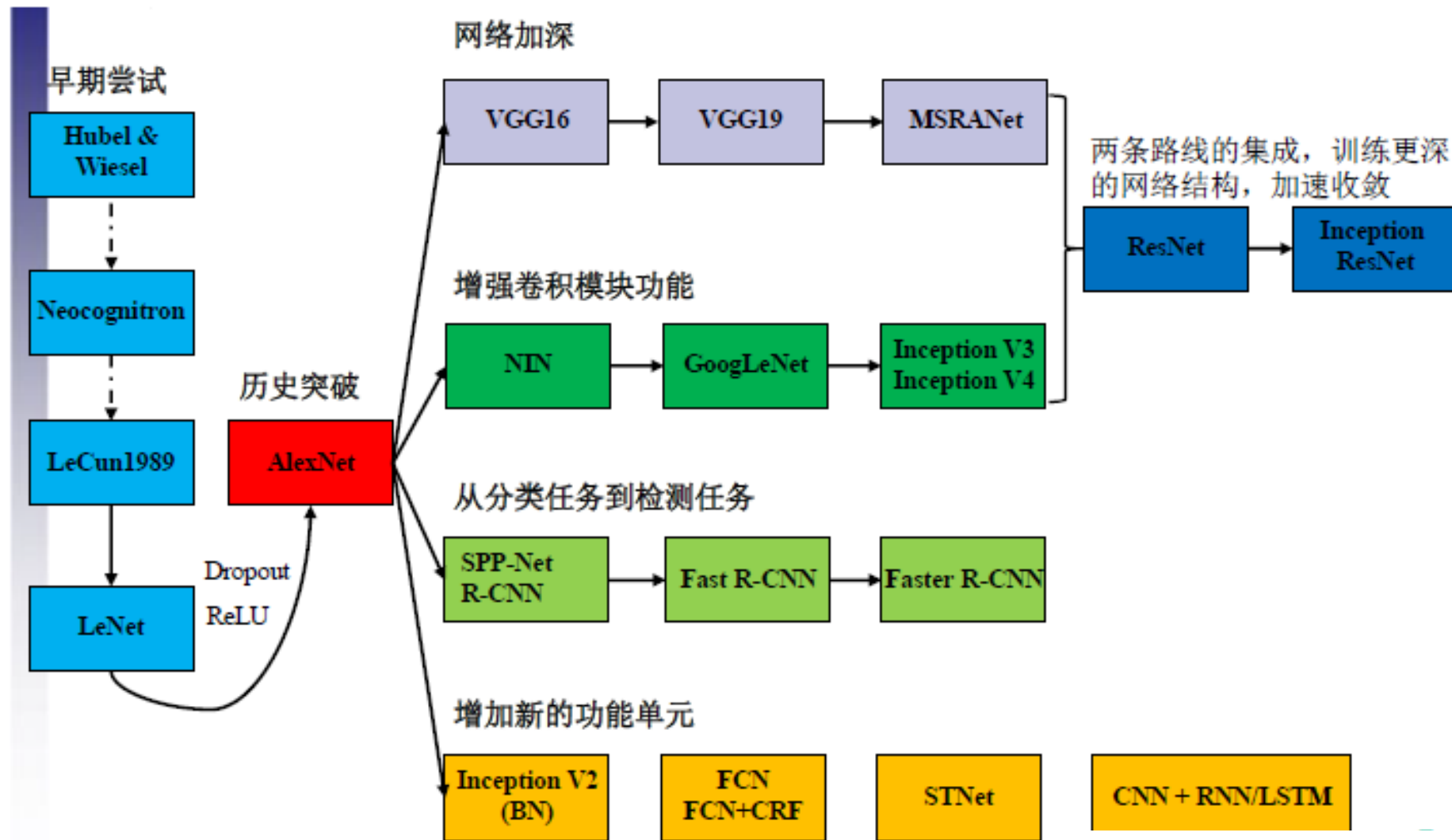
图像增强常用方法（包括但不限于）：

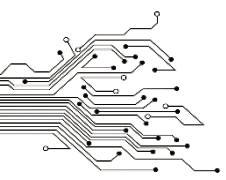
1. 翻转、平移、旋转、缩放
2. 分离单个r、g、b三个颜色通道
3. 添加噪声
4. 直方图均衡化
5. Gamma变换
6. 反转图像的灰度
7. 增加图像的对比度
8. 缩放图像的灰度
9. 均值滤波
10. 中值滤波
11. 高斯滤波



常见的卷积神经网络

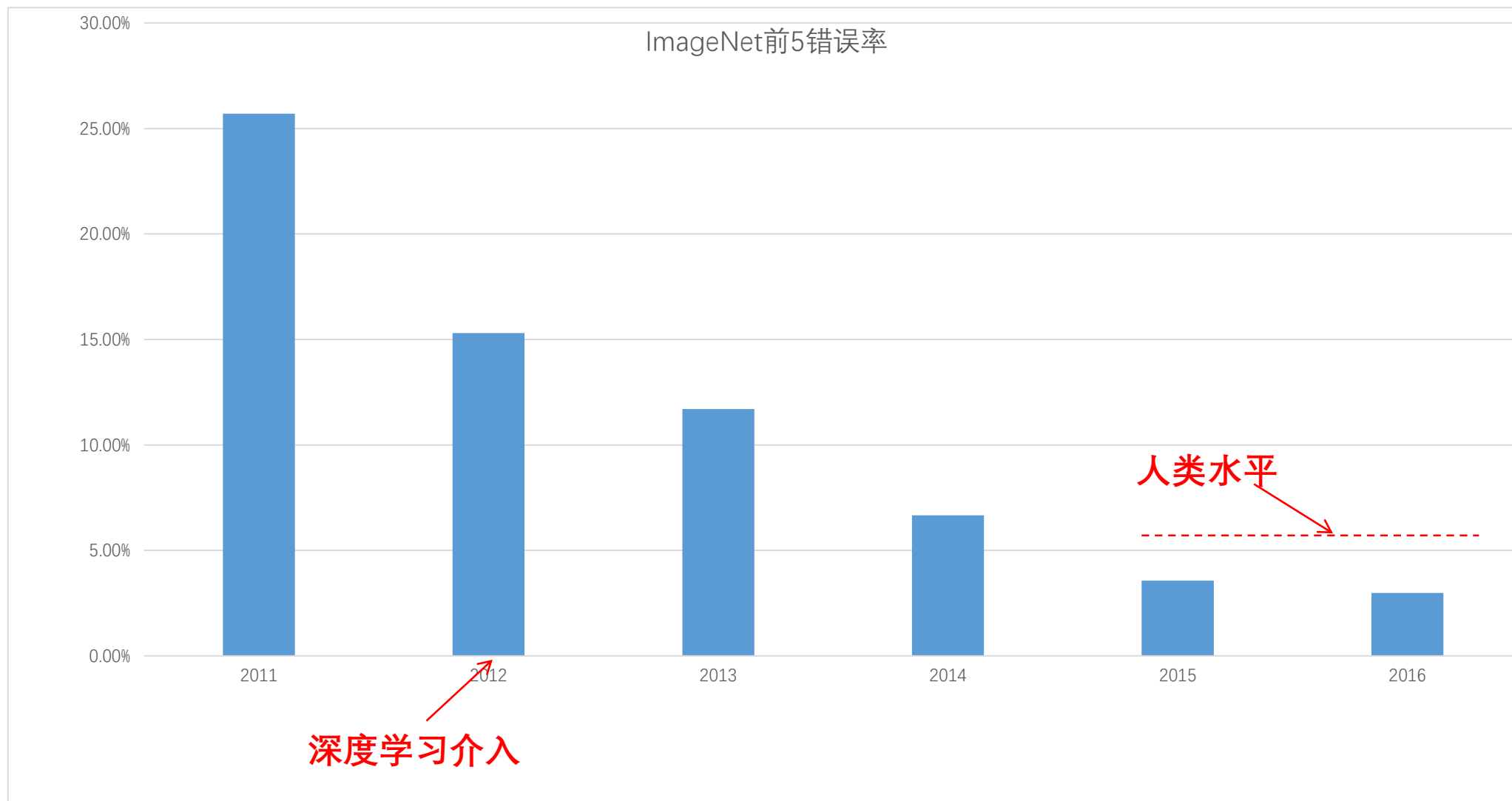
---八斗人工智能，盗版必究---





深度学习为何崛起

---八斗人工智能，盗版必究---

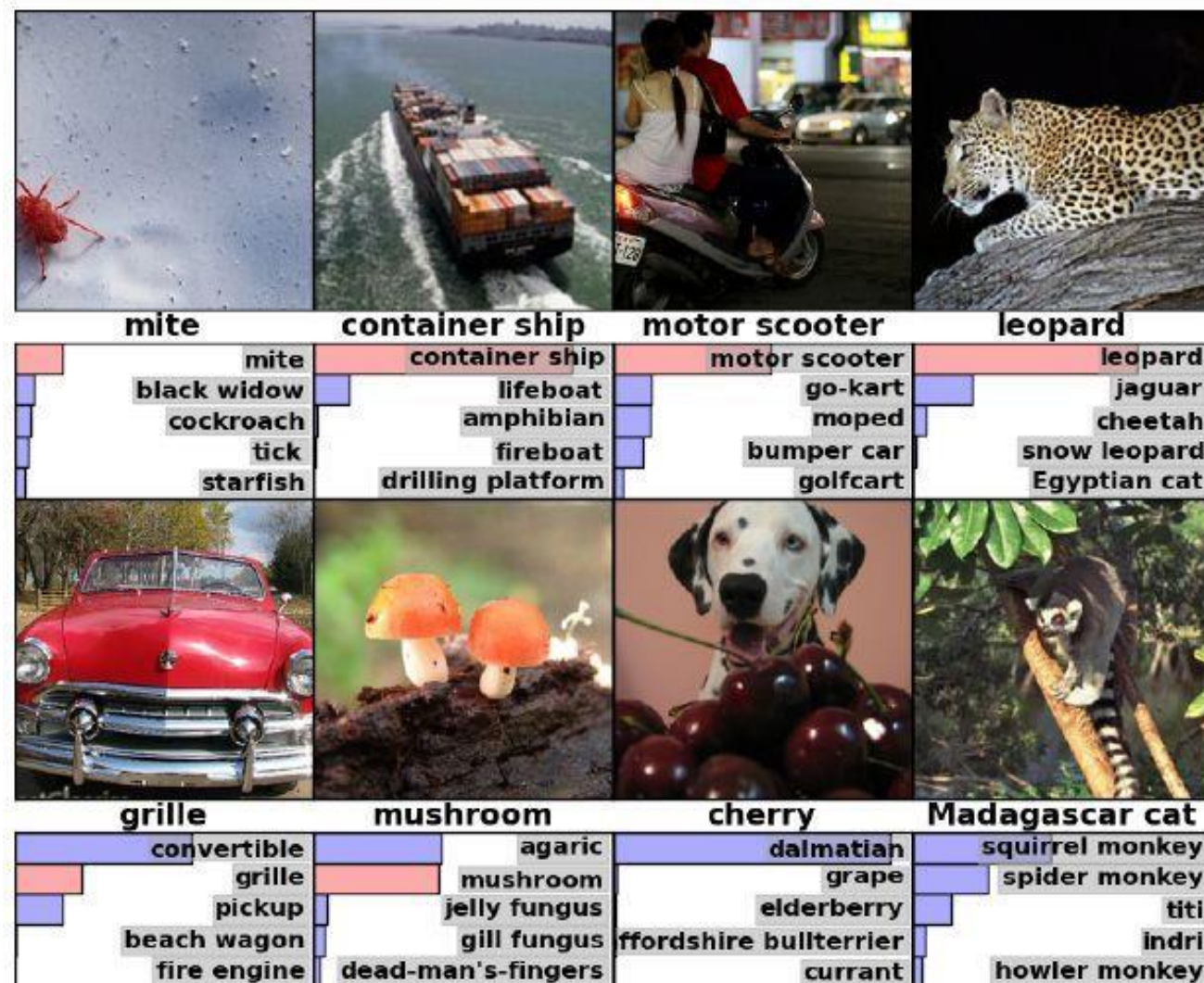




Alexnet

---八斗人工智能，盗版必究---

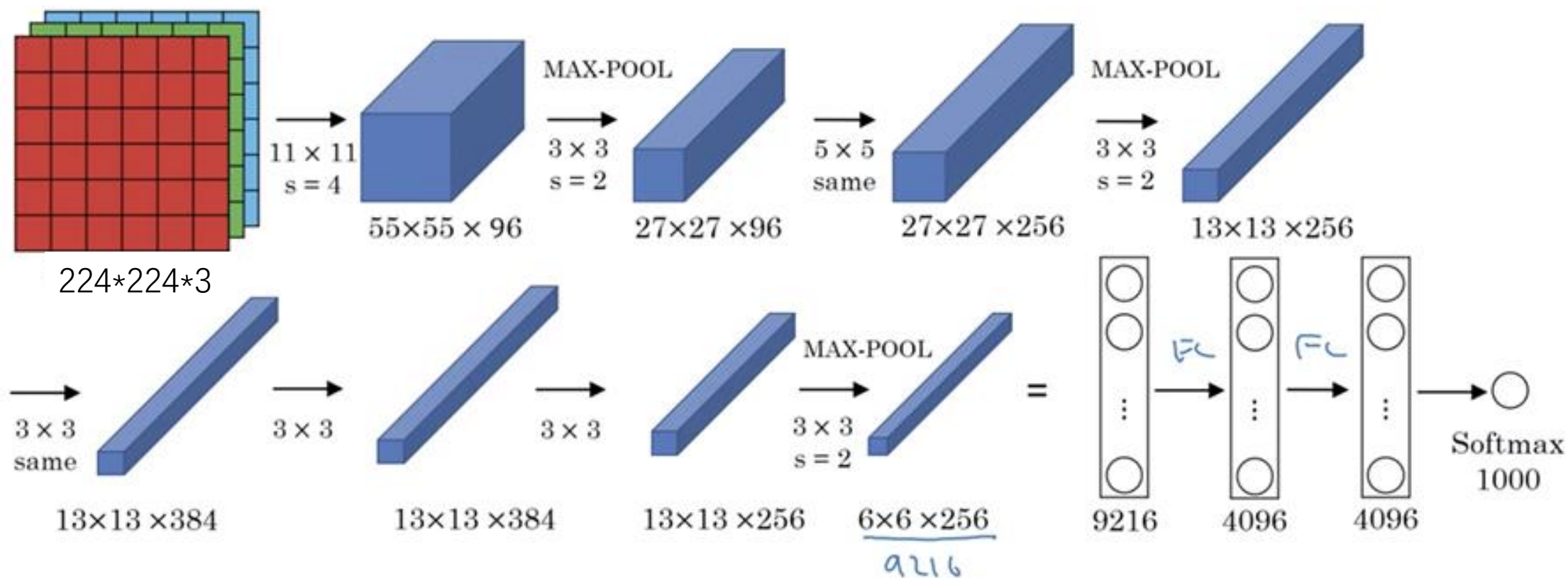
AlexNet是2012年ImageNet竞赛冠军获得者Hinton和他的学生Alex Krizhevsky设计的。也是在那年之后，更多的更深的神经网络被提出。





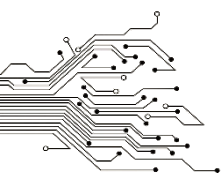
Alexnet

---八斗人工智能，盗版必究---





- 1、一张原始图片被resize到(224,224,3);
- 2、使用步长为4x4，大小为11的卷积核对图像进行卷积，输出的特征层为96层，输出的shape为(55,55,96);
- 3、使用步长为2的最大池化层进行池化，此时输出的shape为(27,27,96)
- 4、使用步长为1x1，大小为5的卷积核对图像进行卷积，输出的特征层为256层，输出的shape为(27,27,256);
- 5、使用步长为2的最大池化层进行池化，此时输出的shape为(13,13,256);
- 6、使用步长为1x1，大小为3的卷积核对图像进行卷积，输出的特征层为384层，输出的shape为(13,13,384);
- 7、使用步长为1x1，大小为3的卷积核对图像进行卷积，输出的特征层为384层，输出的shape为(13,13,384);
- 8、使用步长为1x1，大小为3的卷积核对图像进行卷积，输出的特征层为256层，输出的shape为(13,13,256);
- 9、使用步长为2的最大池化层进行池化，此时输出的shape为(6,6,256);
- 10、两个全连接层，最后输出为1000类



Alexnet

---八斗人工智能，盗版必究---

第一层

第一层输入数据为原始图像的 $224 \times 224 \times 3$ 的图像，这个图像被 $11 \times 11 \times 3$ （3代表深度，例如RGB的3通道）的卷积核进行卷积运算，卷积核每次对原始图像的卷积都会生成一个新的像素。

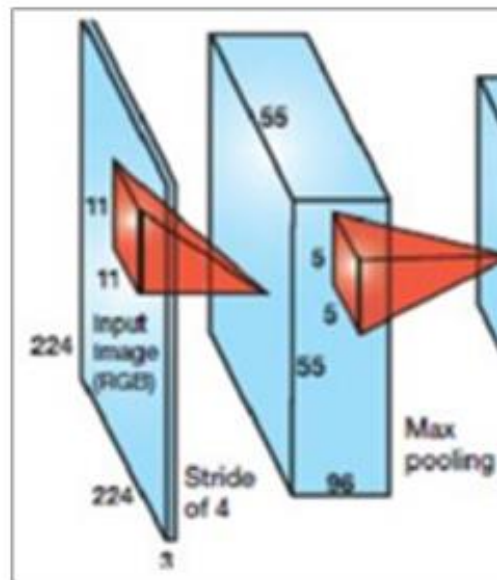
卷积核的步长为4个像素，朝着横向和纵向这两个方向进行卷积。

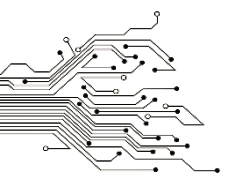
由此，会生成新的像素；

第一层有96个卷积核，所以就会形成 $55 \times 55 \times 96$ 个像素层。

pool池化层：这些像素层还需要经过pool运算（池化运算）的处理，池化运算的尺度由预先设定为 3×3 ，运算的步长为2，则池化后的图像的尺寸为：

$(55-3) / 2 + 1 = 27$ 。即经过池化处理过的规模为 $27 \times 27 \times 96$ 。





---八斗人工智能，盗版必究---

