



人工智能基础与进阶

深度网络

梁子睿
上海交通大学

目录 content



第一节

从特征工程到深度神经网络

第二节

神经元模型

第三节

深度神经网络的结构

第四节

深度学习与经典机器学习的优势

第五节

深度学习的应用

引言



1950

1980

2010

上海交通大学人工
智能创新教育实验室

上海交通大学人工
智能创新教育实验室

上海交通大学人工
智能创新教育实验室

第一节 从特征工程到深度神经网络

特征工程

什么是特征？

一事物异于其他事物的特点；

人或事物所具有的特别或特殊之处。

一个客体或一组客体特性的抽象结果。特征是用来描述概念的。



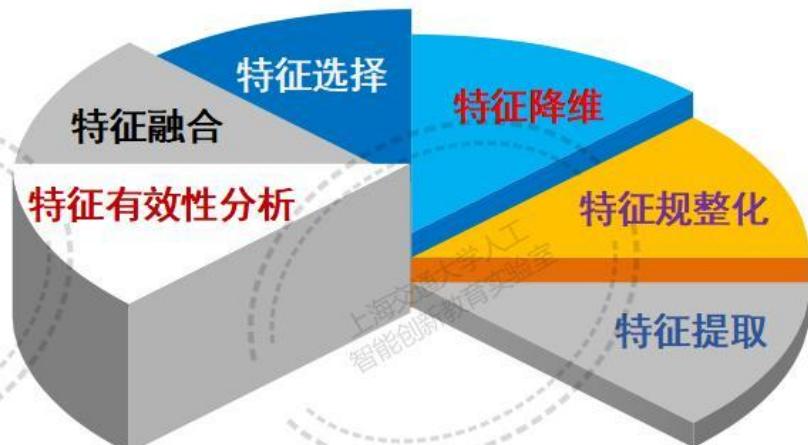
数据和特征决定了机器学习的上限，而模型和算法只是逼近这个上限而已。

特征工程

特征是机器学习系统的原材料。

特征工程是最大限度地从原始数据中提取特征以供算法和模型使用。

特征工程是利用数据领域的相关知识来创建能够使机器学习算法达到最佳性能的特征的过程。



特征工程

特征是机器学习系统的原材料。

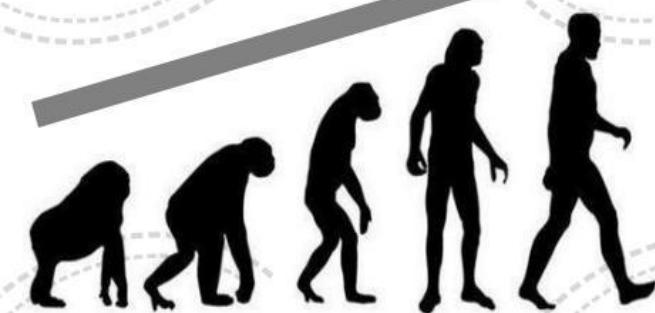
特征工程是最大限度地从原始数据中提取特征以供算法和模型使用。

特征工程是利用数据领域的相关知识来创建能够使机器学习算法达到最佳性能的特征的过程。



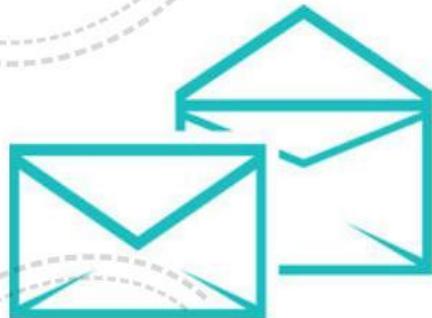
例如，在设计垃圾邮件过滤系统的应用中，首先要收集大量的用户邮件和对应标记，然后通过分析这些数据，提取一些与垃圾邮件相关的特征，如邮件是否包含“交友”、“发票”、“免费促销”等一些与垃圾邮件相关的关键词，再使用线性逻辑回归来训练得到模型。

特征工程



由于特征工程提取到的信息对后续学习算法具有很强的指导性，而不同时间设计的特征提取方法也会有所差异。因此，为了从原始数据尽可能多地获取信息，特征工程是一个长期的过程，需要不断的提出新的特征。

特征工程



例如：

- 通过分析被错误分类的邮件，可能会发现含有大量彩色文字和图片的邮件大概率是垃圾邮件，于是需要新加入一个样式特征。
- 根据经验，一个长期使用中文的人收到的俄语邮件不是正常邮件，可以直接过滤掉，于是又加入字符编码这一新特征。

由此可以看出，特征工程建立在对问题的不断深入和对数据的不断寻求之上。

特征工程



特征工程

上海交通大学人工
智能创新教育实验室



通常情况下，人根据数据能抽象出来的特征种类非常有限。

一方面，人自身能够概括出来的概念特征非常有限；常用的如大小、颜色、形状、纹理等

另一方面，即使人们通过数学工具也很难将一些特征表达出来。例如，行人、城市街道、乡村道路等。

人们一直在努力提高特征提取的能力。

表示学习

解决这个问题的途径之一是使用机器学习来发掘表示本身，而不仅仅把表示映射到输出。这种方法称为表示学习。

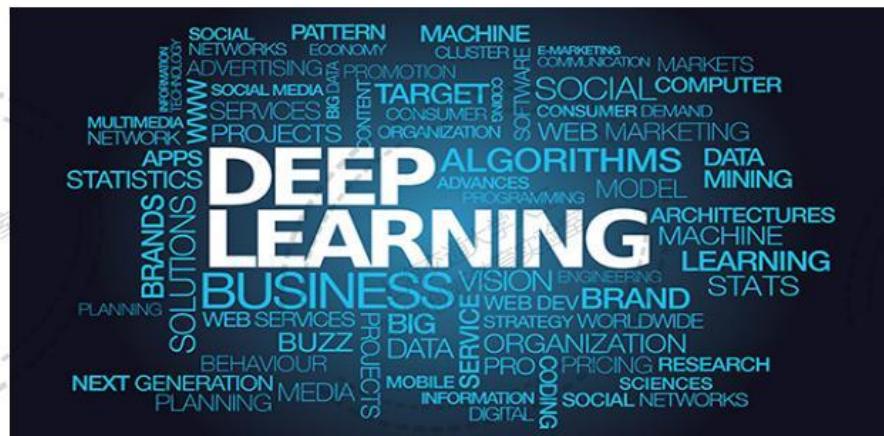


传统机器学习的不足：

- 对于复杂任务需要几小时到几个月，而且耗费大量的人工时间和精力。
- 在许多现实的人工智能应用中，比如，在一张包含红色汽车的图片中，其单个像素在夜间可能会非常接近黑色，汽车轮廓的形状在不同视角下会有很大的变化；复杂的场景很难用确定的特征表述来表达。

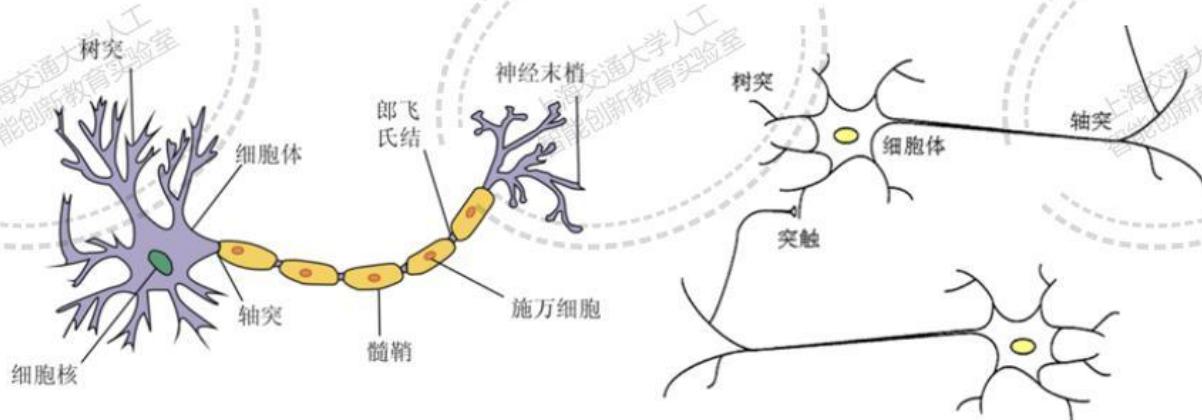
深度学习

- 深度学习将所需的复杂映射分解为一系列嵌套的简单映射（每个由模型的不同层描述）来解决这一难题。
 - 可以通过其他较简单的表示来表达复杂表示，有效地解决了表示学习中的核心问题。



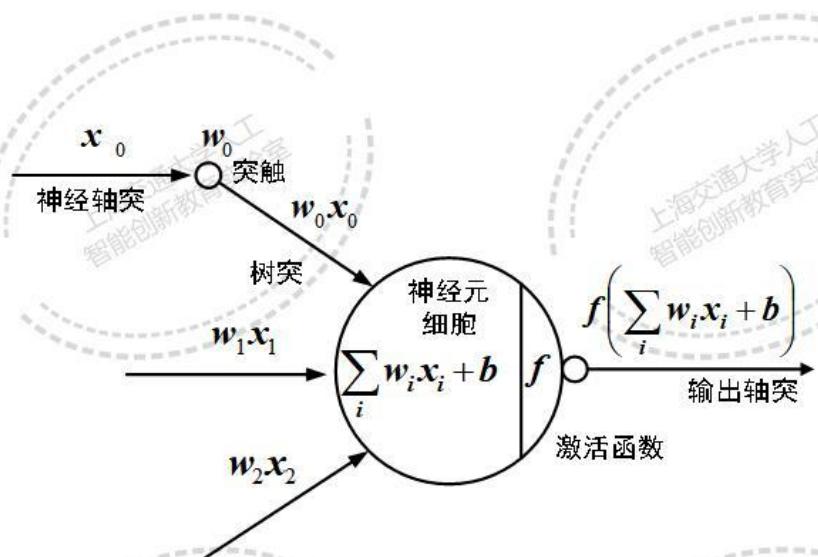
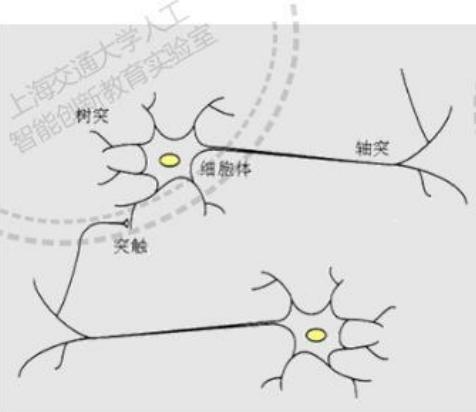
第二节 神经元模型

神经元模型



在生物神经网络中，每个神经元都通过轴突和树突与其它神经元相连，当它“兴奋”时，就会向相连的其它神经元发送化学物质，从而改变这些神经元内的电位；如果某神经元的电位超过了一个“阈值”，那么它就会被激活，即“兴奋”起来，向其他神经元发送化学物质。

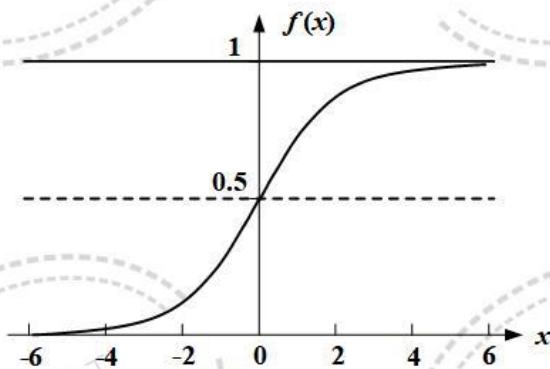
神经元模型



1943年，McCulloch and Pitts将上述情形抽象为图所示的简单模型，这就是“M-P神经元模型”。在这个模型中，神经元接收到来自n个其它神经元传递过来的输入信号，这些输入信号通过带权重的连接进行传递，神经元接收到的总输入值通过“激活函数”处理以产生神经元的输出。

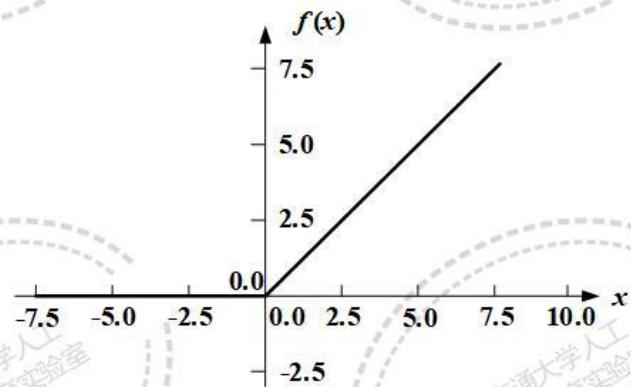
激活函数

(1) Sigmoid函数



$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}$$

(2) ReLU函数



$$f(x) = \max(x, 0)$$

激活函数

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}}$$

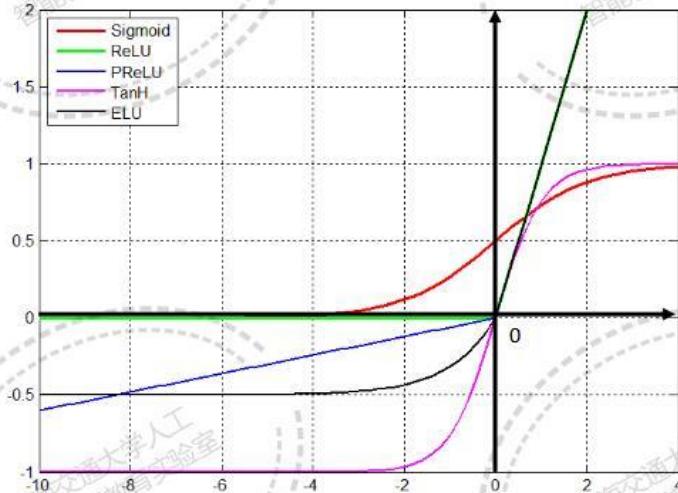
$$\text{ReLU}(x) = \max(x, 0)$$

$$\text{PReLU}(x) = \begin{cases} x, & x > 0 \\ \alpha x, & x \leq 0 \end{cases}$$

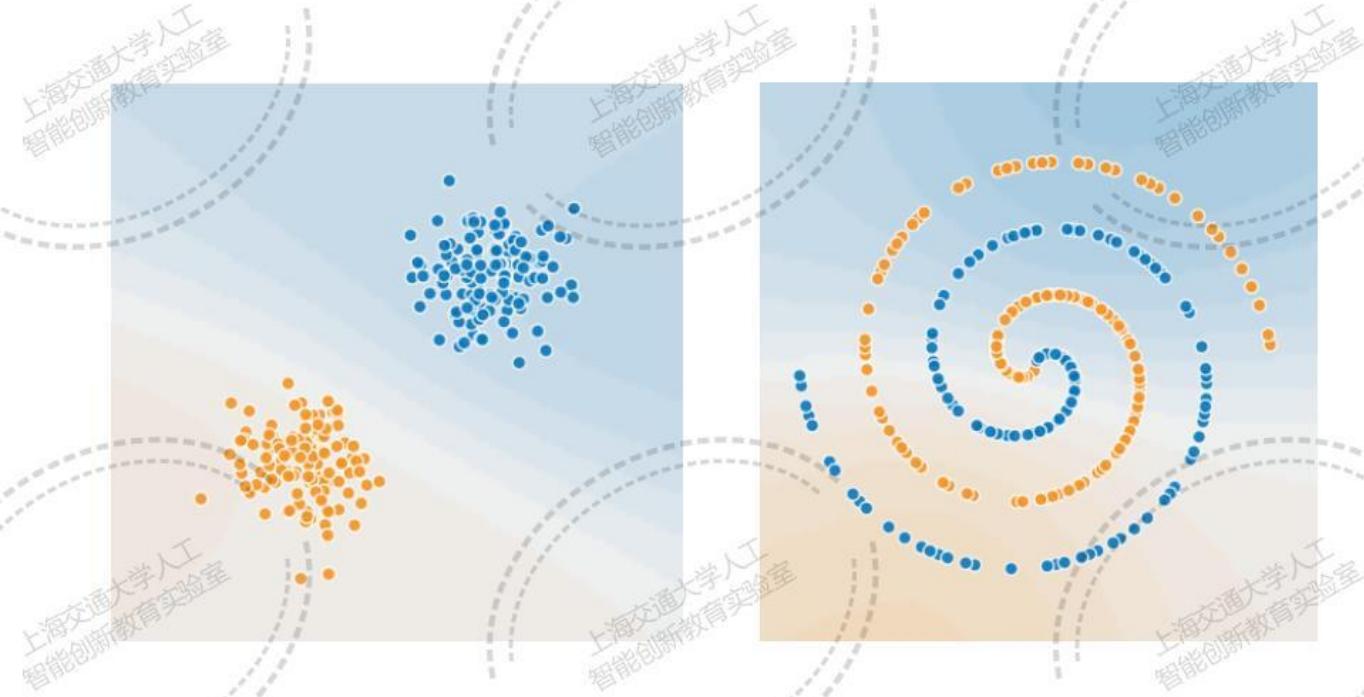
$$\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$\text{ELU}(x) = \begin{cases} x, & x > 0 \\ \alpha(e^x - 1), & x \leq 0 \end{cases}$$

$$\text{CReLU}(x) = [\text{ReLU}(x), \text{ReLU}(-x)]$$



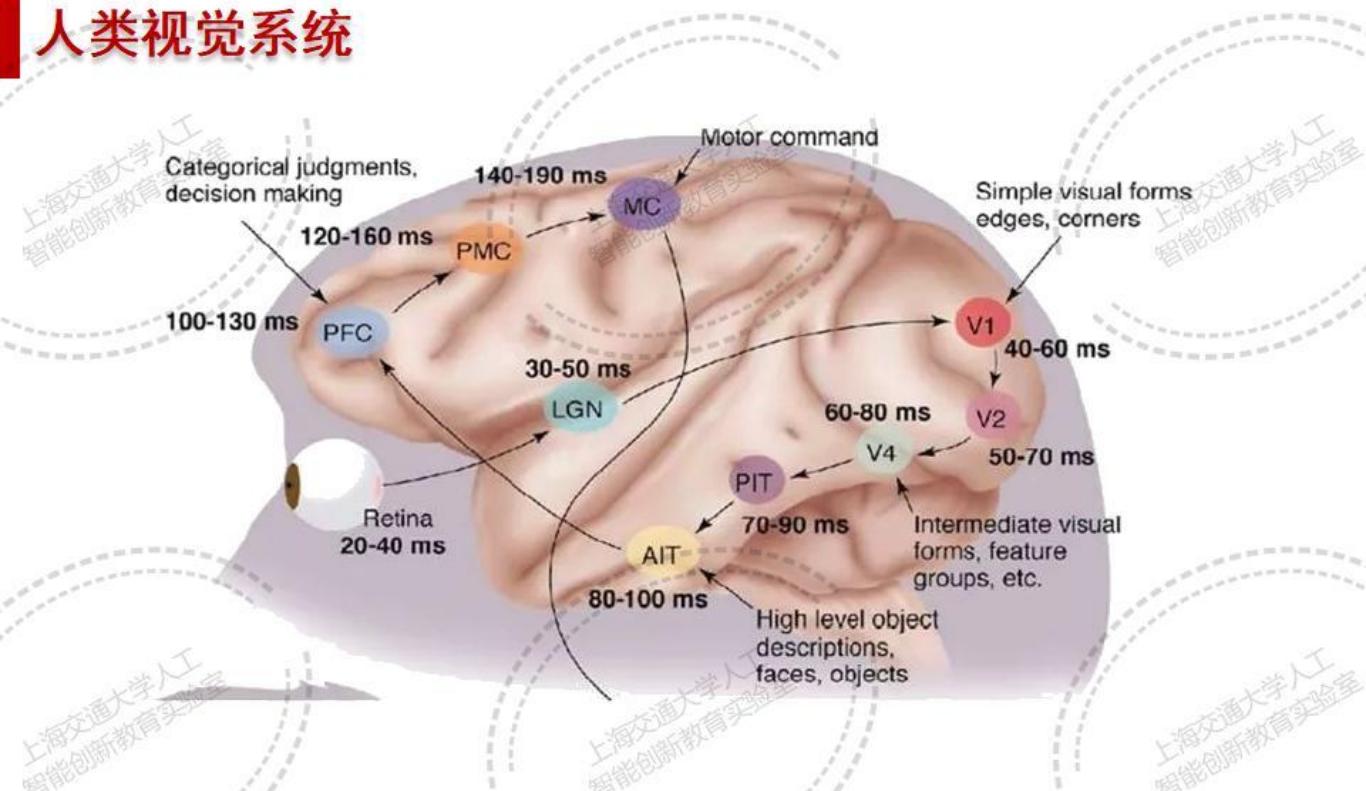
激活函数





第三节 深度神经网络结构

人类视觉系统



美国神经生物学家David Hubel和Torsten Wiesel (1981年诺贝尔医学获得者) 的发现：人类负责视觉系统的信息处理的可视皮层是分级的。

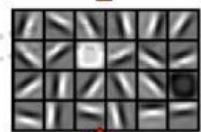
人类视觉系统



对象模型



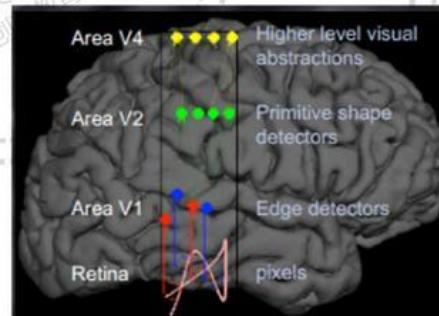
对象组件



边沿

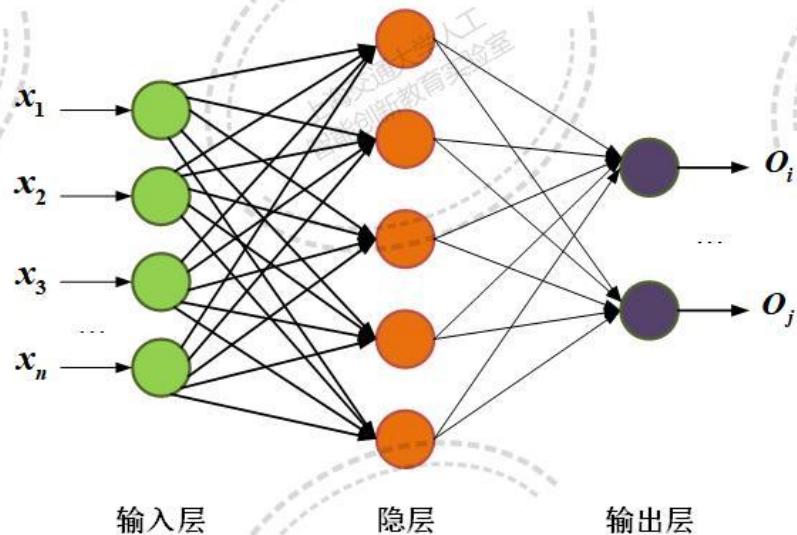


像素



这个发现激发了人们对于神经系统的进一步思考：神经系统-中枢-大脑的工作过程，或许是一个不断迭代、不断抽象的过程。从原始信号，做低级抽象，逐渐向高级抽象迭代。而人类的逻辑思维，经常使用高度抽象的概念。

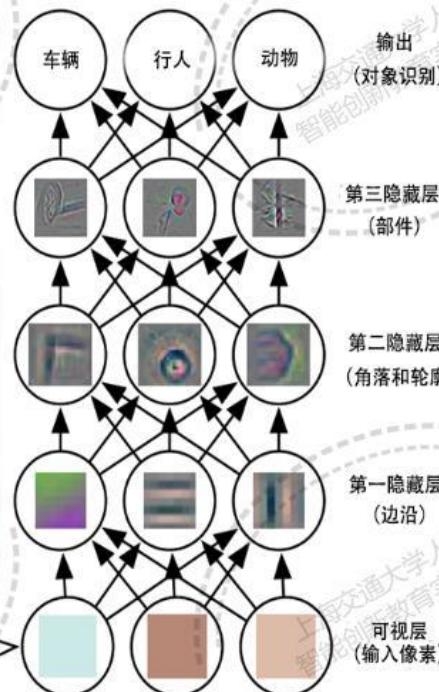
神经网络结构



一个深度神经网络通常由多个顺序连接的层（layer）组成。这种层次化的特征提取过程可以累加，赋予神经网络强大的特征提取能力。经过很多层的变换之后，神经网络就可以将原始图像变换为高层次的抽象的特征。

深度学习

上海交通大学人工
智能创新教育实验室



根据图像描述中包含的对象部分，可
以识别图像中存在的对象

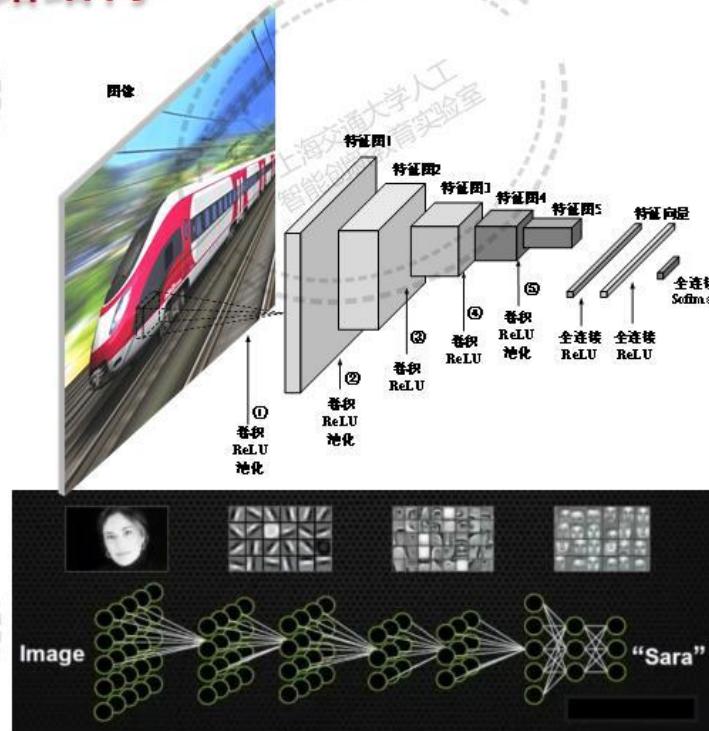
可以找到轮廓和角的特定集合来检测
特定对象的整个部分

可以容易地搜索可识别为角和扩展轮
廓的边集合

可以轻易地通过比较相邻像素的亮度
来识别边缘

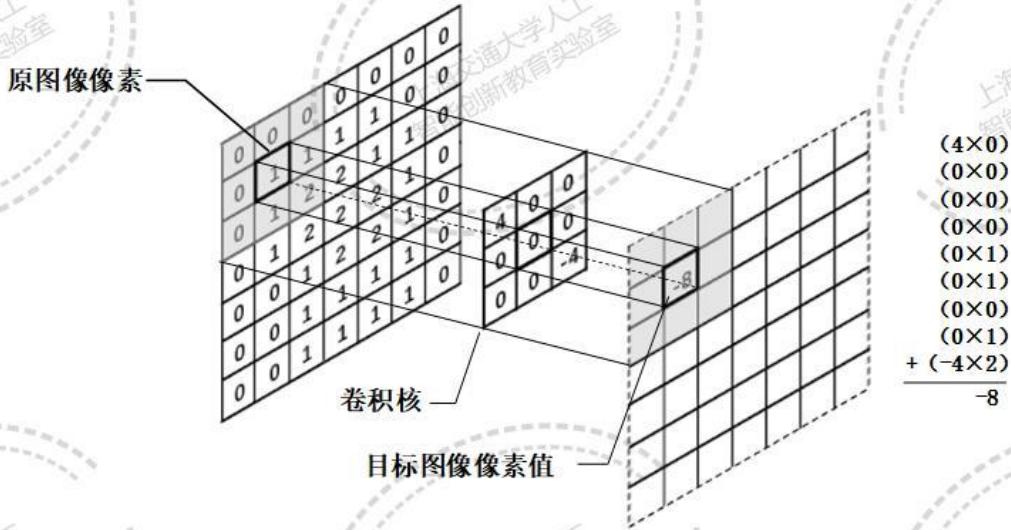
输出

深度神经网络结构



在2012年，Alex Krizhevsky发表了AlexNet。这个神经网络的主体部分由五个卷积层和三个全连接层组成。五个卷积层位于网络的最前端，依次对图像进行变换以提取特征。

卷积层



卷积核通常是一个正方形的网格结构，该区域上每个方格都有一个权重值。使用卷积进行计算时，需要将卷积核的中心放置在要计算的像素上，依次计算卷积核中每个元素和其覆盖的图像像素值的乘积并求和，得到的结构就是该位置的新像素值。

卷积层

卷积计算步骤：

1. 用这个卷积核去覆盖原始图片；
2. 覆盖一块跟卷积核一样大的区域之后，对应元素相乘，然后求和；
3. 计算一个区域之后，就向其他区域挪动（假设步长是1），继续计算；
4. 直到把原图片的每一个角落都覆盖到为止。

计算下面的卷积：

10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0

*

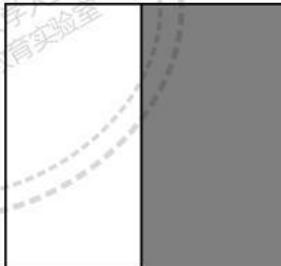
1	0	-1
1	0	-1
1	0	-1

卷积层

边缘提取：

10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0

原图像素

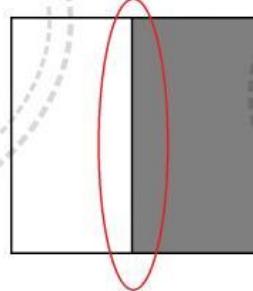


$$\begin{matrix} * & \begin{matrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{matrix} & = \end{matrix}$$

0	30	30	0
0	30	30	0
0	30	30	0
0	30	30	0

卷积核

激活值



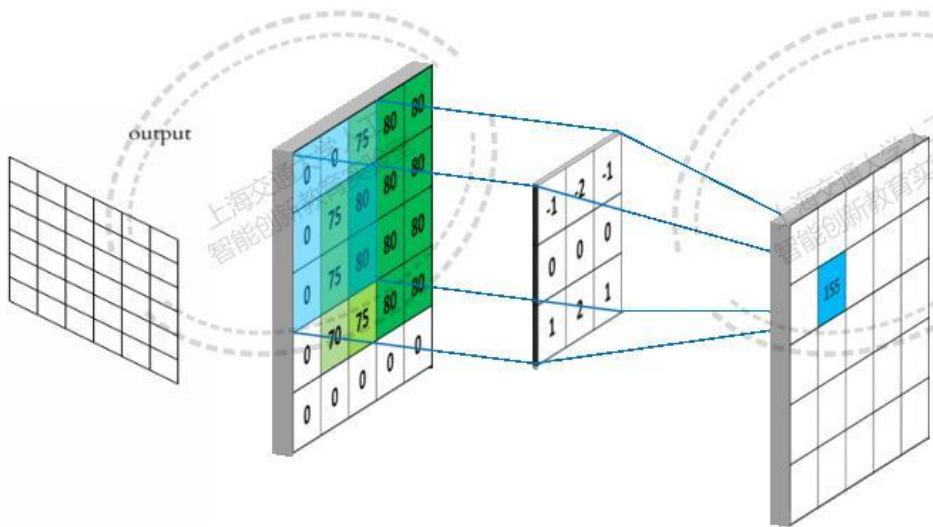
卷积层

上海交通大学人工
智能创新教育实验室

0	-1	0
-1	5	0
0	-1	0

7	6	5	5	6	7
6	4	3	3	4	6
5	3	2	2	3	5
5	3	2	2	3	5
6	4	3	2	3	5
7	6	5	3	4	6
6	5	5	5	6	7

input



上海交通大学人工
智能创新教育实验室

卷积层

10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0

6*6

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

*

=

0	30	30	0
0	30	30	0
0	30	30	0
0	30	30	0

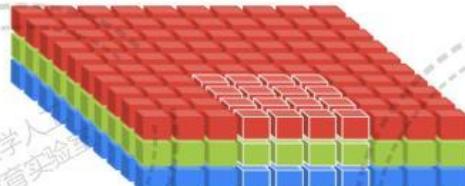
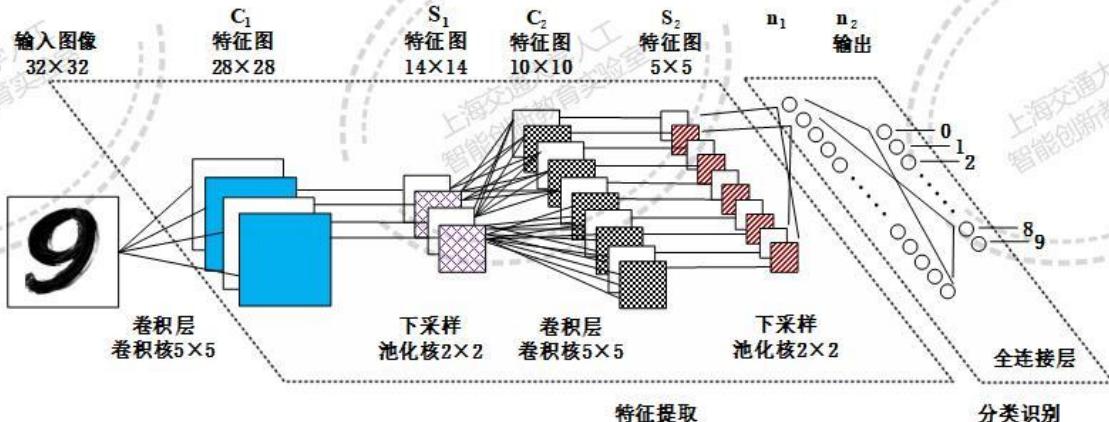
4*4

Height: 原图高度
Width: 原图宽度
Kernel size: 卷积核大小
Stride: 卷积步长
Padding: 边缘填补

3*3

$$\left\{ \left\lceil \frac{H + 2P - K}{S} \right\rceil + 1 \right\} \times \left\{ \left\lceil \frac{W + 2P - K}{S} \right\rceil + 1 \right\}$$

卷积层



实践操作

<https://www.cs.ryerson.ca/~aharley/vis/conv/>
<https://playground.tensorflow.org/>
<https://poloclub.github.io/cnn-explainer/>

池化层

池化层的主要目的是通过降采样的方式，在不影响图像质量的情况下，压缩图片、减少参数。简单来说，假设特征图像大小为 3×3 ，池化层采用MaxPooling，大小为 2×2 ，步长为1，那么图片的尺寸就会从 3×3 变为 2×2 。

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

-1	-1	0
1	1	0
1	0	0

卷积后的特征图



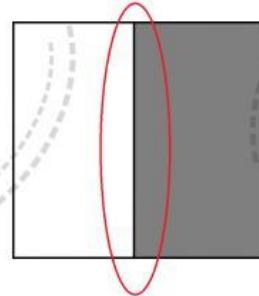
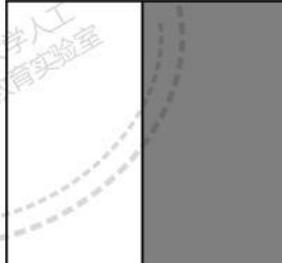
最大池化后的特征图

深度学习学什么？——卷积核

10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0

$$\begin{matrix} * & \begin{matrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{matrix} & = & \begin{matrix} 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \end{matrix} \end{matrix}$$

原图像素



深度学习怎么学？——代价函数

随机初始化一个卷积核：

1	0	3
0	2	-1
2	1	-2

计算卷积：

10	10	10	0	0	0	0
10	10	10	0	0	0	0
10	10	10	0	0	0	0
10	10	10	0	0	0	0
10	10	10	0	0	0	0
10	10	10	0	0	0	0

$$\begin{matrix} * & \begin{matrix} 1 & 0 & 3 \\ 0 & 2 & -1 \\ 2 & 1 & -2 \end{matrix} \end{matrix} = \begin{matrix} 60 & 60 & 30 & 0 \\ 60 & 60 & 30 & 0 \\ 60 & 60 & 30 & 0 \\ 60 & 60 & 30 & 0 \end{matrix}$$

计算代价：cost =

$$(\begin{matrix} 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \end{matrix} - \begin{matrix} 60 & 60 & 30 & 0 \\ 60 & 60 & 30 & 0 \\ 60 & 60 & 30 & 0 \\ 60 & 60 & 30 & 0 \end{matrix})^2$$

最终目标：最小化代价cost

手段：反向传播、梯度下降法等优化方法

深度学习怎么学？——反向传播

理想值: Y , 实际值: $f(x_1, x_2) = w_1x_1 + w_2x_2$

代价(损失)函数: $L = (Y - f(x_1, x_2))^2$

代价函数: 本次检查卫生本班被扣的分数

理想值: 卫生满分

实际值: 我们目前卫生状况的得分

x_1 : 教室卫生状况

x_2 : 走廊卫生状况

目标: 最小化代价函数 (使得下次卫生检查我们被扣分最少)

反向传播: 从被扣的分数出发, 反思哪块区域还做得不够好

L 分别对 x_1, x_2 求偏导:

即通过被扣分数情况, 判断教室、走廊的卫生状况分别对于扣分多少的影响

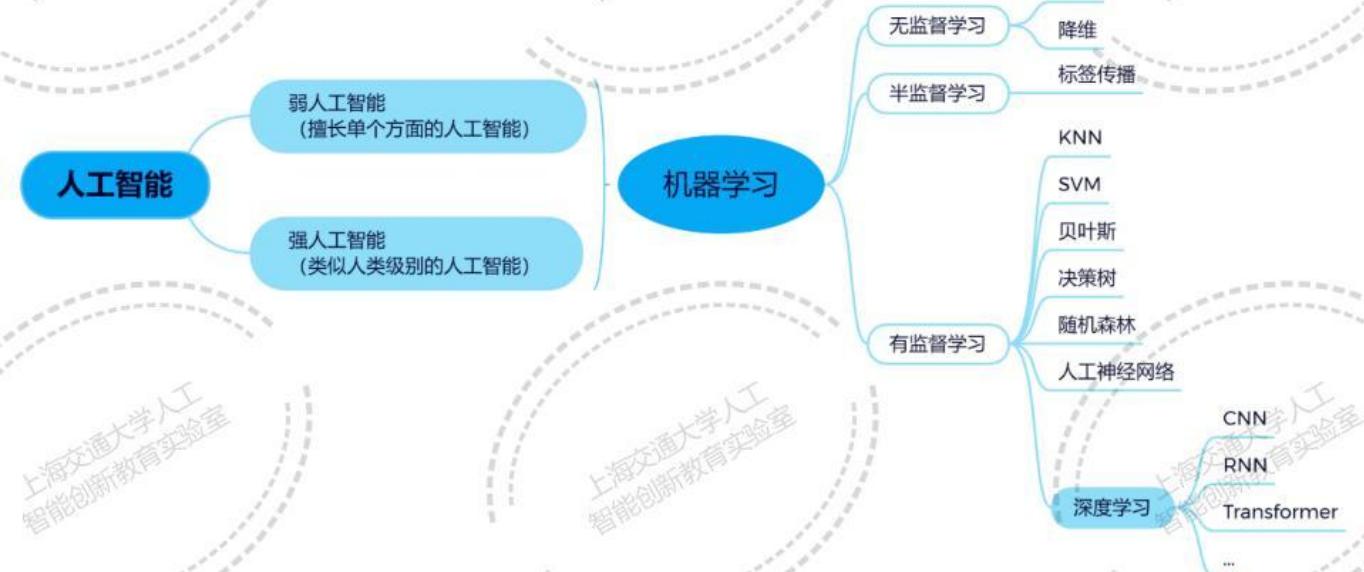
更新权重 w_1, w_2 :

即更新我们对教室、走廊的卫生状况的重视程度

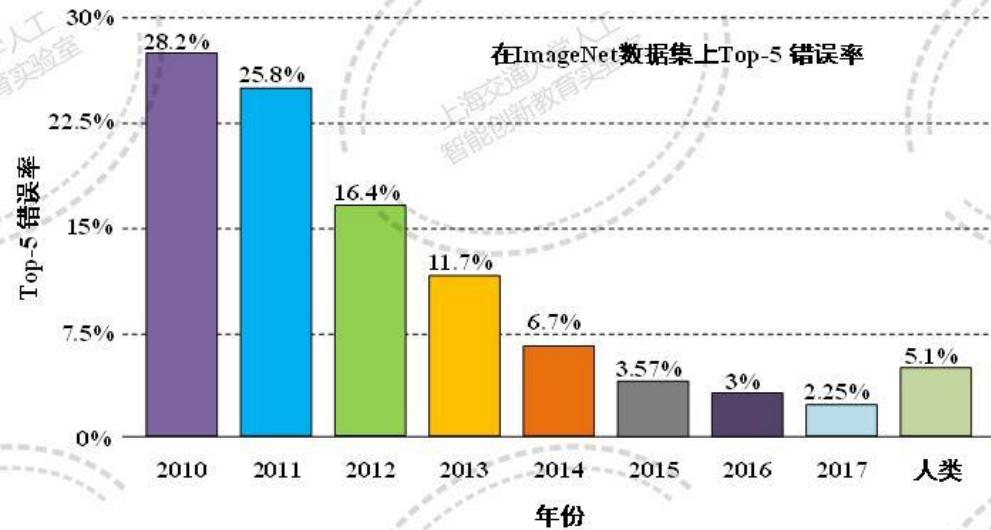
第四节

深度学习与经典机器学习的优势

机器学习？深度学习？

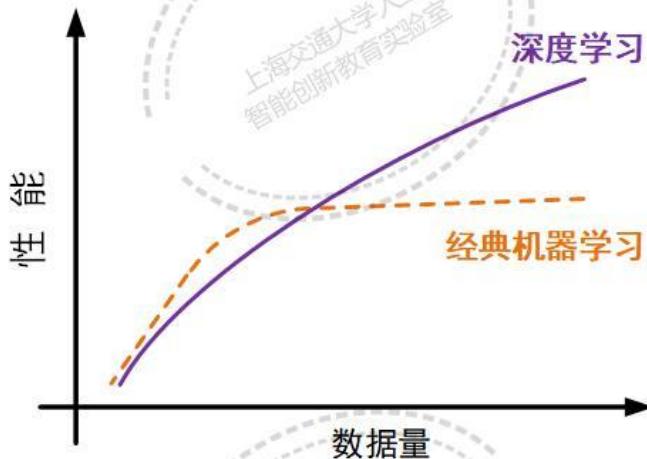


深度学习的优势



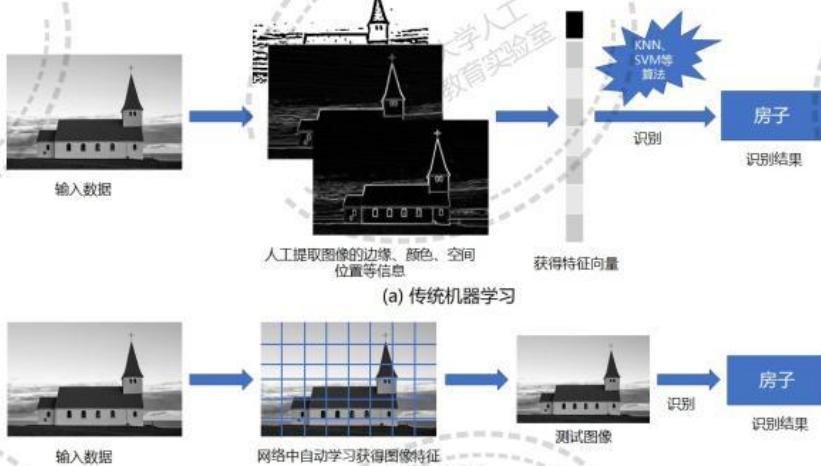
一、**优异的表现**。深度网络已经实现了远远超过传统ML方法的精确度，包括语音、自然语言、视觉和玩游戏等诸多领域。在许多任务中，经典ML甚至无法竞争。例如，上图显示了ImageNet数据集上不同方法的图像分类准确性，蓝色表示经典ML方法，紫色表示深度卷积神经网络(CNN)方法。

深度学习的优势



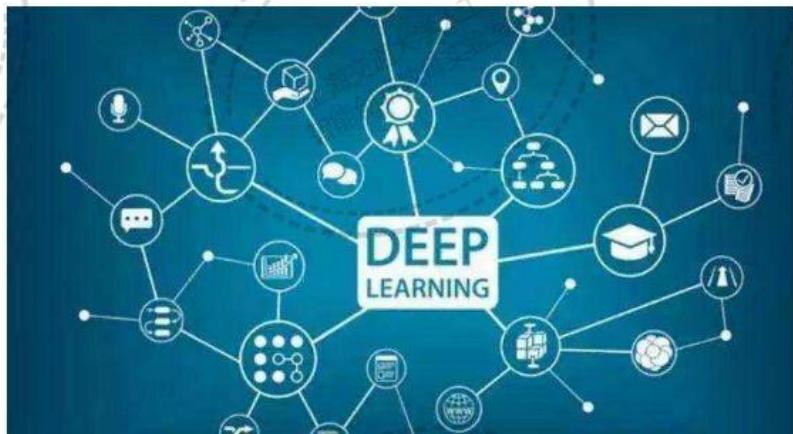
二、**使用数据进行有效缩放**。与传统ML算法相比，深度网络使用更多的数据可以更好地扩展。上面的图表是一个简单而有效的例子。很多时候，通过深层网络来提高准确性的最佳建议就是使用更多的数据。使用经典的ML算法，这种快速简单的修复方法甚至几乎没有效果，并且通常需要更复杂的方法来提高准确性。

深度学习的优势



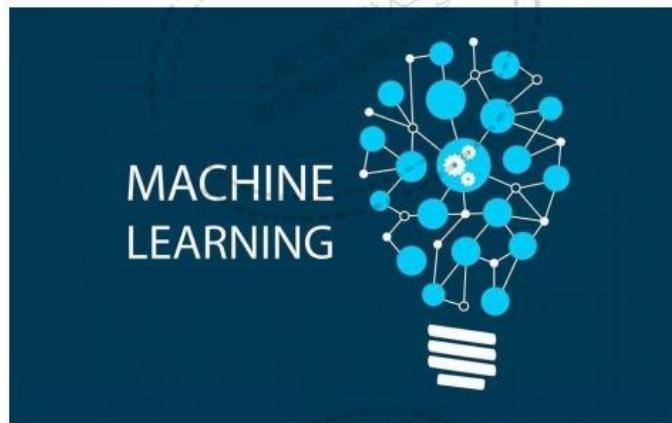
三、不需要特征工程: 经典的ML算法通常需要复杂的特征工程。首先在数据集上执行深度探索性数据分析，然后做一个简单的降低维数的处理，最后，必须仔细选择最佳功能以传递给ML算法。当使用深度网络时，不需要这样做，因为只需将数据直接传递到网络，通常就可以实现良好的性能。这完全消除了整个过程的大型和具有挑战性的特征工程阶段。

深度学习的优势



四、适应性强，易于转换。与传统的ML算法相比，深度学习技术可以更容易地适应不同的领域和应用。例如，在计算机视觉中，预先训练的图像分类网络通常用作对象检测和分割网络的特征提取前端。将这些预先训练的网络用作前端，可以减轻整个模型的训练，并且通常有助于在更短的时间内实现更高的性能。

经典机器学习的优势



一、对小数据更好。为了实现高性能，深层网络需要非常大的数据集。对于许多应用来说，这样的大数据集并不容易获得，并且花费昂贵且耗时。对于较小的数据集，传统的ML算法通常优于深度网络。

经典机器学习的优势



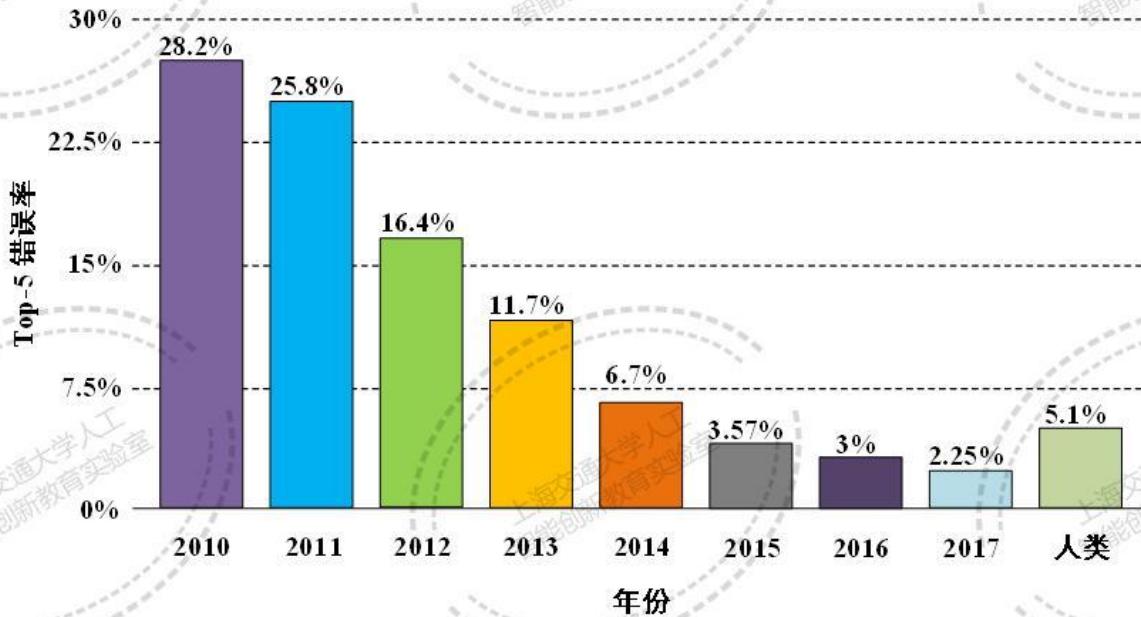
二、更容易理解。由于传统ML中涉及直接特征工程，这些算法很容易解释和理解。此外，调整超参数并更改模型设计更加简单，因为我们对数据和底层算法都有了更全面的了解。另一方面，深层网络是“黑匣子”型，即使现在研究人员也不能完全了解深层网络的“内部”。由于缺乏理论基础，超参数和网络设计也是一个相当大的挑战。



第五节 深度学习的应用

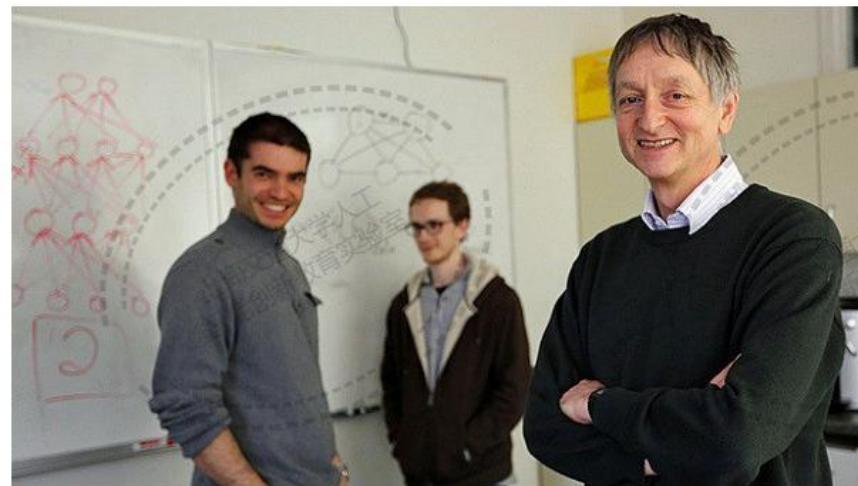
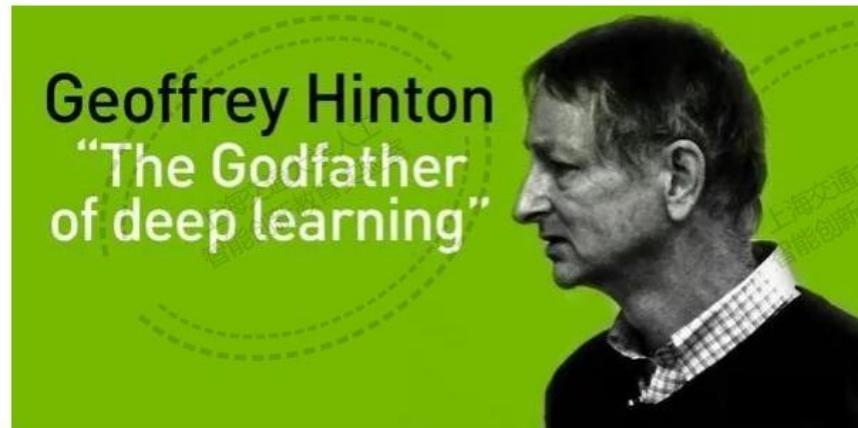


ILSVRC, 全称是ImageNet Large-Scale Visual Recognition Challenge



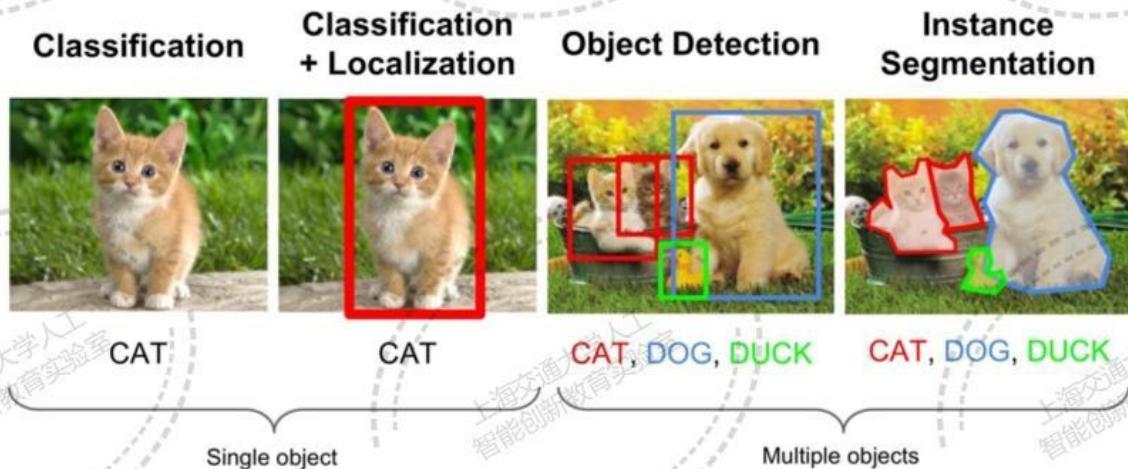
Alexnet

上海交通大学人工
智能创新教育实验室

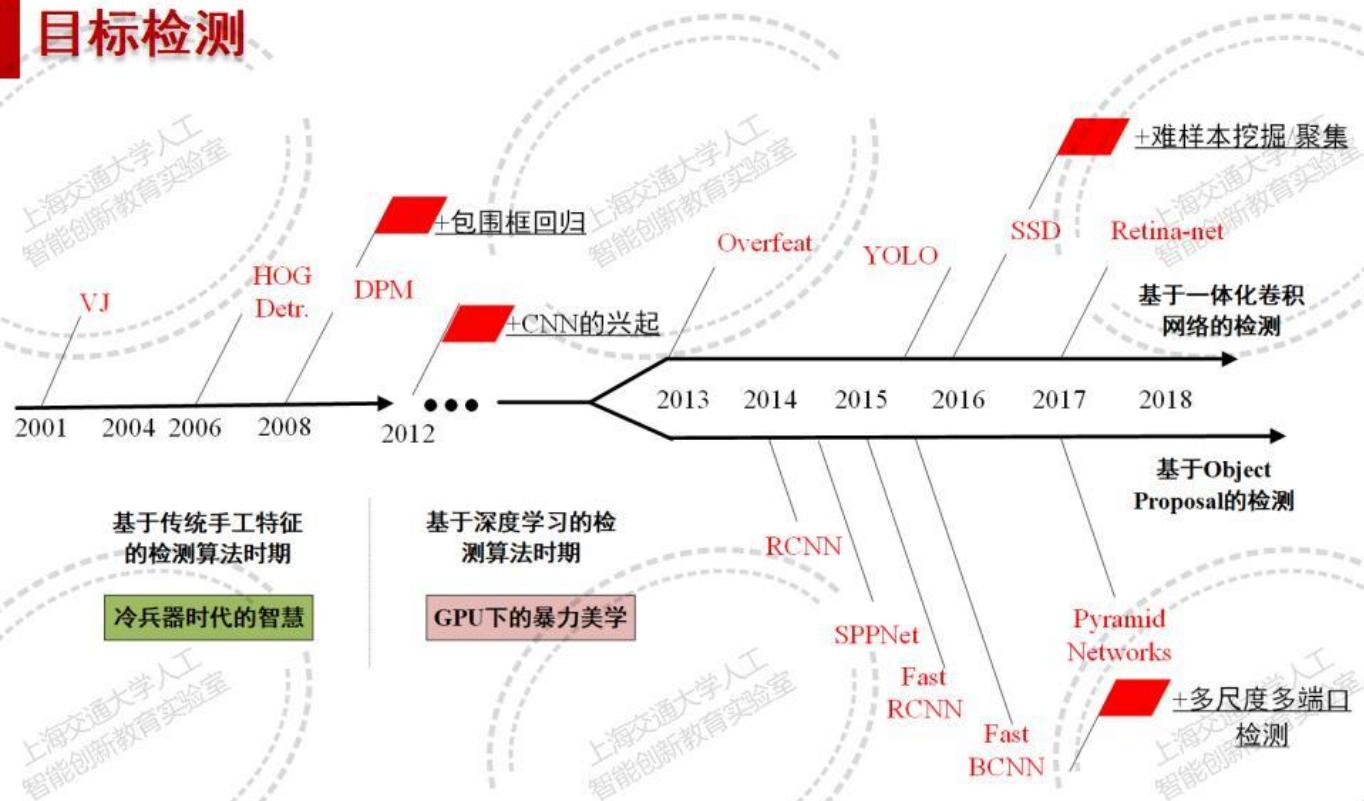


目标检测

目标检测即找出图像中所有感兴趣的物体，包含物体定位和物体分类两个子任务，同时确定物体的类别和位置。



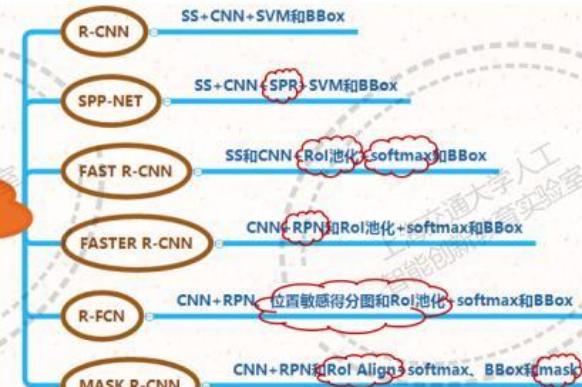
目标检测



目标检测

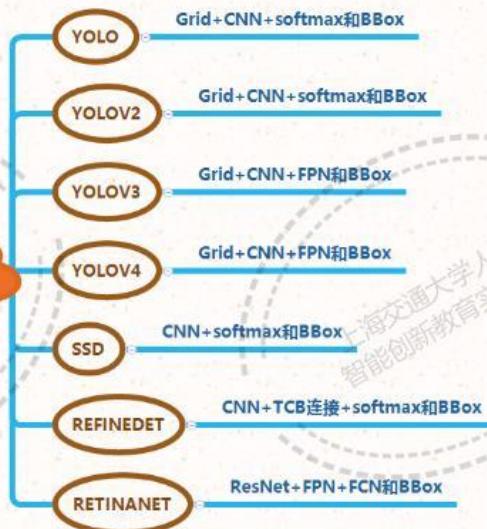
两阶段 (two stage)：先生成候选框（可能包含物体的区域），然后再对每个候选框进行分类（也会修正位置）。这类算法相对较慢，因为它需要多次运行检测和分类流程，但精度相对较高

两阶段

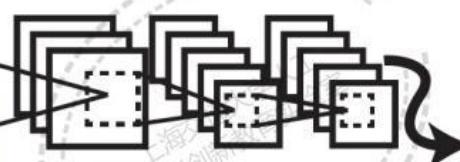
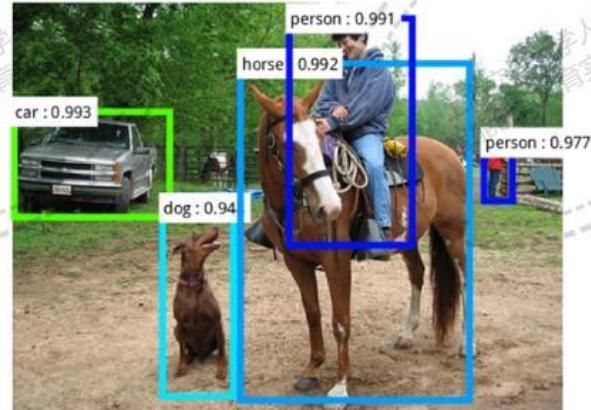


一阶段 (one stage)：特点是一步到位，仅仅需要送入网络一次就可以预测出所有的边界框，速度相对较快，非常适合移动端

一阶段



目标检测

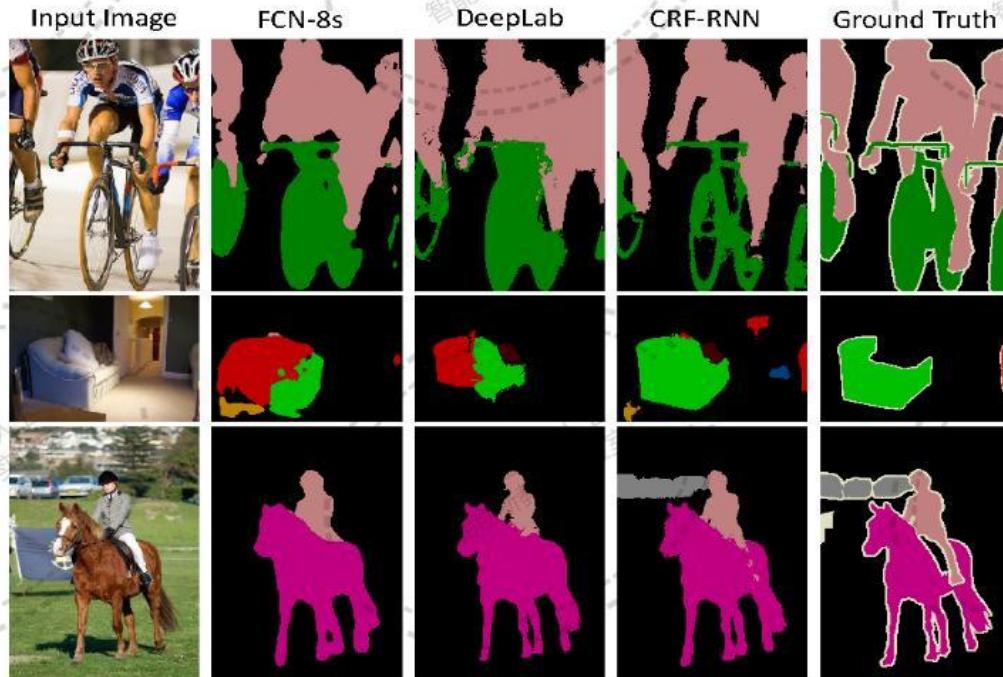


1. Resize image.
2. Run convolutional network.
3. Non-max suppression.

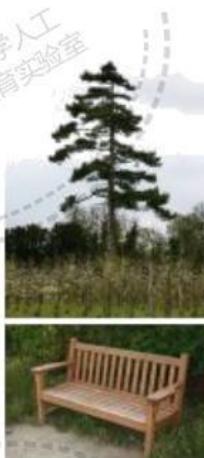


图像分割

图像分割是指在像素水平上对图像进行分类。目的是把图像中各种不同物体给用不同颜色分割出来。



图像分割



图像分割

数字图像 (1921)

数字图像处理 (~1960)

像素值

第一代分割：阈值化算法 (~1960)

Otsu 法：最大类间方差阈值化。 (IEEE Trans. Sys., Man., Cyber, 1979)

微处理器、大规模集成电路出现 (~1970)

低层特征

第二代分割：基于边缘、区域或图论的分割 (~1970)

边缘检测、分水岭、KMeans、Graph Cut、MeanShift



卷积神经网络的崛起 (2012)

高层语义

全卷积神经网络 (2014)

第三代分割：语义分割 & 实例分割

FCNs, ParseNet,
PSPNet, UNet, SegNet,
Mask RCNN



图像标注 (Image to Text)

图像标注是一项引人注目的研究领域，它的研究目的是给出一张图像 (Image)，让机器用一段文本 (Text) 描述它。

一个人在尘土飞扬的
路面上骑摩托车



两只狗在草地上玩耍



一群年轻人在玩飞碟



两个曲棍球球员正在争夺一个冰球



图像生成 (Text to Image)

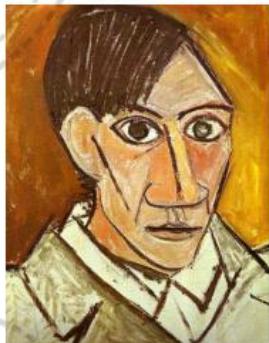
根据给定的文本条件 (Text) 准确的生成一张精度足够高的图像 (Image)

《哈利·波特》的赫敏本该长这样? DALL-E2「还原」小说人物形象



图像风格变换

图像风格变换是指参考某种风格的图片，让另一张图片转换成该风格的过程。



图像风格变换



(a) Input



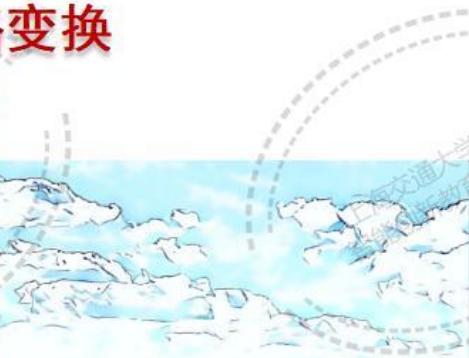
(b) Reference style



(c) Our result



图像风格变换



图像风格变换

上海交通大学人工
智能创新教育实验室



上海交通大学人工
智能创新教育实验室





AI

稍后继续

THANKS FOR YOUR ATTENTION

上海交通大学人工智能
智能创新教育实验室