

原 Deep Learning (深度学习) 学习笔记整理系列之 (一)

分类: Deep Learning 机器学习 机器视觉 神经网络

2013-04-08 23:35 2525人阅读

Deep Learning (深度学习) 学习笔记整理系列

zouxy09@qq.com

<http://blog.csdn.net/zouxy09>

作者: **Zouxy**

version 1.0 2013-04-08

声明:

- 1) 该Deep Learning的学习系列是整理自网上很大牛和机器学习专家所无私奉献的资料的。具体引用的资料请看参考文献。具体的版本声明也参考原文献。
- 2) 本文仅供学术交流,非商用。所以每一部分具体的参考资料并没有详细对应。如果某部分不小心侵犯了大家的利益,还望海涵,并联系博主删除。
- 3) 本人才疏学浅,整理总结的时候难免出错,还望各位前辈不吝指正,谢谢。
- 4) 阅读本文需要机器学习、计算机视觉、神经网络等等基础(如果没有也没关系了,没有就看看,能不能看懂,呵呵)。
- 5) 此属于第一版本,若有错误,还需继续修正与增删。还望大家多多指点。大家都共享一点点,一起为祖国科研的推进添砖加瓦(呵呵,好高尚的目标啊)。请联系:
zouxy09@qq.com

目录:

一、概述

二、背景

三、人脑视觉机理

四、关于特征

4.1、特征表示的粒度

4.2、初级（浅层）特征表示

4.3、结构性特征表示

4.4、需要有多少个特征？

五、Deep Learning的基本思想

六、浅层学习（Shallow Learning）和深度学习（Deep Learning）

七、Deep learning与Neural Network

八、Deep learning训练过程

8.1、传统神经网络的训练方法

8.2、deep learning训练过程

九、Deep Learning的常用模型或者方法

9.1、AutoEncoder自动编码器

9.2、Sparse Coding稀疏编码

9.3、Restricted Boltzmann Machine(RBM)限制波尔兹曼机

9.4、Deep Belief Networks深信度网络

9.5、Convolutional Neural Networks卷积神经网络

十、总结与展望

十一、参考文献和Deep Learning学习资源

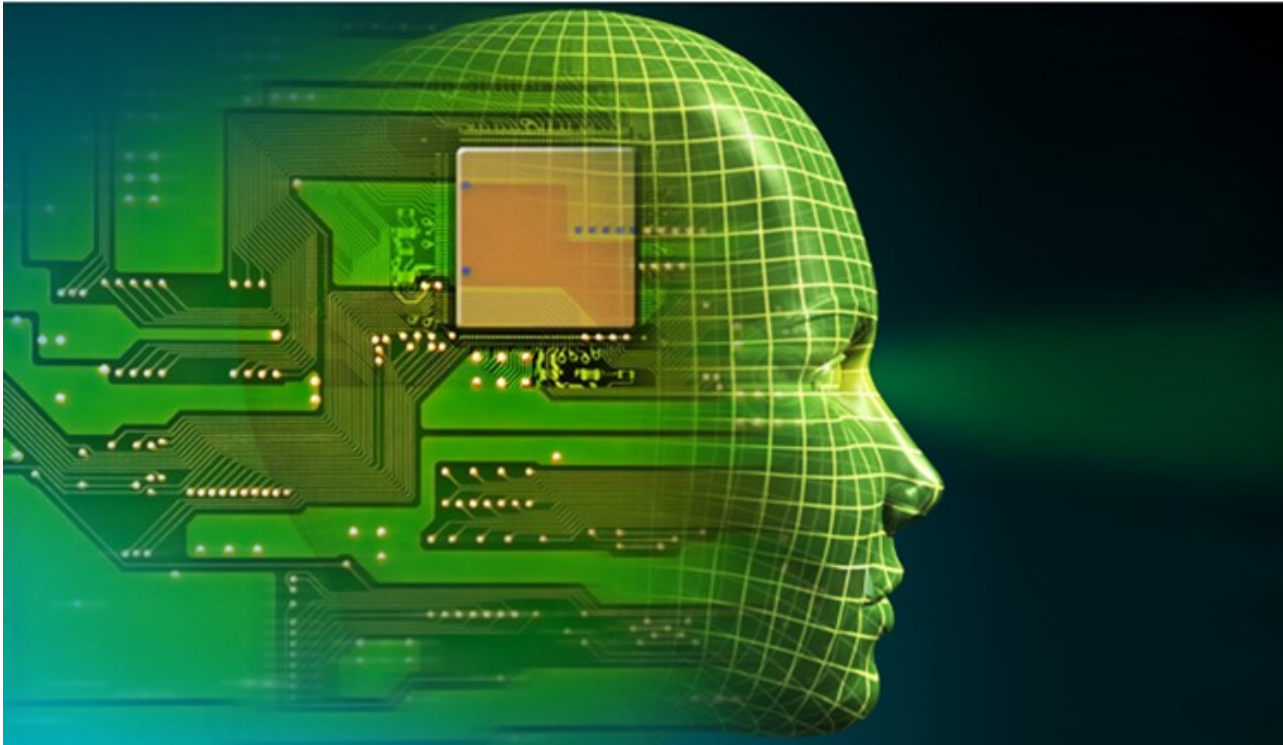
一、概述

Artificial Intelligence，也就是人工智能，就像长生不老和星际漫游一样，是人类最美好的梦想之一。虽然计算机技术已经取得了长足的进步，但是到目前为止，还没有一台电脑能产生“自我”的意识。是的，在人类和大量现成数据的帮助下，电脑可以表现的十分强大，但是离开了这两者，它甚至都不能分辨一个喵星人和一个汪星人。

图灵（图灵，大家都知道吧。计算机和人工智能的鼻祖，分别对应于其著名的“图灵机”和“图灵测试”）在 1950 年的论文里，提出图灵试验的设想，即，隔墙对话，你将不知道与你谈话的，是人还是电脑。这无疑给计算机，尤其是人工智能，预设了一

个很高的期望值。但是半个世纪过去了，人工智能的进展，远远没有达到图灵试验的标准。这不仅让多年翘首以待的人们，心灰意冷，认为人工智能是忽悠，相关领域是“伪科学”。

但是自 2006 年以来，机器学习领域，取得了突破性的进展。图灵试验，至少不是那么可望而不可及了。至于技术手段，不仅仅依赖于云计算对大数据的并行处理能力，而且依赖于算法。这个算法就是，**Deep Learning**。借助于 **Deep Learning** 算法，人类终于找到了如何处理“抽象概念”这个亘古难题的方法。



2012年6月，《纽约时报》披露了Google Brain项目，吸引了公众的广泛关注。这个项目是由著名的斯坦福大学的机器学习教授**Andrew Ng**和在大规模计算机系统方面的世界顶尖专家**Jeff Dean**共同主导，用16000个CPU Core的并行计算平台训练一种称为“深度神经网络”（DNN, **Deep Neural Networks**）的机器学习模型（内部共有10亿个节点。这一网络自然是不能跟人类的神经网络相提并论的。要知道，人脑中可是有150多亿个神经元，互相连接的节点也就是突触数更是如银河沙数。曾经有人估算过，如果将一个人的大脑中所有神经细胞的轴突和树突依次连接起来，并拉成一根直线，可从地球连到月亮，再从月亮返回地球），在语音识别和图像识别等领域获得了巨大的成功。

项目负责人之一**Andrew**称：“我们没有像通常做的那样自己框定边界，而是直接把海量数据投放到算法中，让数据自己说话，系统会自动从数据中学习。”另外一名负责人**Jeff**则说：“我们在训练的时候从来不会告诉机器说：‘这是一只猫。’系统其实是自己发明或者领悟了“猫”的概念。”



2012年11月，微软在中国天津的一次活动上公开演示了一个全自动的同声传译系统，讲演者用英文演讲，后台的计算机一气呵成自动完成语音识别、英中机器翻译和中文语音合成，效果非常流畅。据报道，后面支撑的关键技术也是DNN，或者深度学习（DL，DeepLearning）。

2013年1月，在百度年会上，创始人兼CEO李彦宏高调宣布要成立百度研究院，其中第一个成立的就是“深度学习研究所”（IDL，Institute of Deep Learning）。



为什么拥有大数据的互联网公司争相投入大量资源研发深度学习技术。听起来感觉deeplearning很牛那样。那什么是deep learning？为什么有deep learning？它是怎么来的？又能干什么呢？目前存在哪些困难呢？这些问题的简答都需要慢慢来。咱们先来了解下机器学习（人工智能的核心）的背景。

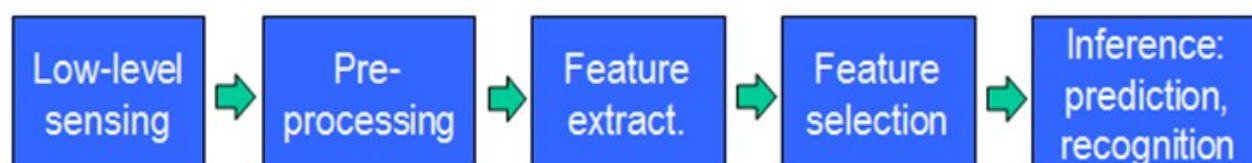
二、背景

机器学习（Machine Learning）是一门专门研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能的学科。机器能否像人类一样能具有学习能力呢？1959年美国的塞缪尔(Samuel)设计了一个下棋程序，这个程序具有学习能力，它可以在不断的对弈中改善自己的棋艺。4年后，这个程序战胜了设计者本人。又过了3年，这个程序战胜了美国一个保持8年之久的常胜不败的冠军。这个程序向人们展示了机器学习的能力，提出了许多令人深思的社会问题与哲学问题（呵呵，人工智能正常的轨道没有很大的发展，这些什么哲学伦理啊倒发展的挺快。什么未来机器越来越像人，人越来越像机器啊。什么机器会反人类啊，ATM是开第一枪的啊等等。人类的思维无穷啊）。

机器学习虽然发展了几十年，但还是存在很多没有良好解决的问题：



例如图像识别、语音识别、自然语言理解、天气预测、基因表达、内容推荐等等。目前我们通过机器学习去解决这些问题的思路都是这样的（以视觉感知为例子）：

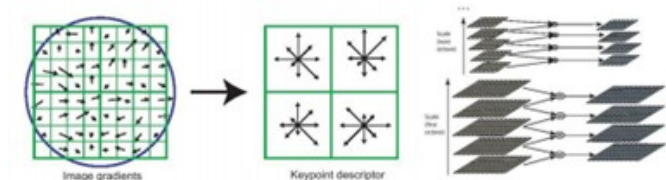


从开始的通过传感器（例如CMOS）来获得数据。然后经过预处理、特征提取、特征选择，再到推理、预测或者识别。最后一个部分，也就是机器学习的部分，绝大部分的工作是在这方面做的，也存在很多的paper和研究。

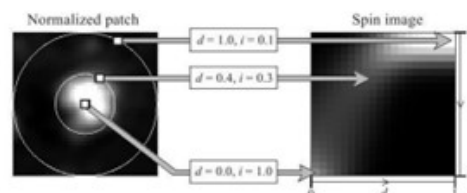
而中间的三部分，概括起来就是特征表达。良好的特征表达，对最终算法的准确性起了非常关键的作用，而且系统主要的计算和测试工作都耗在这一大部分。但，这块实际中一般都是人工完成的。靠人工提取特征。



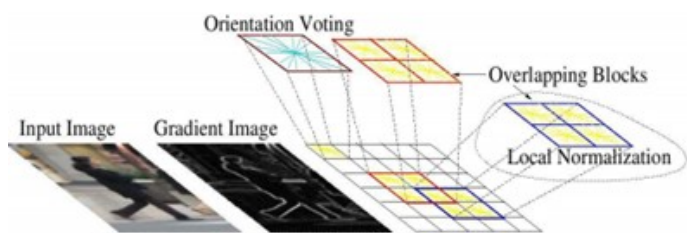
截止现在，也出现了不少NB的特征（好的特征应具有不变性（大小、尺度和旋转等）和可区分性）：例如Sift的出现，是局部图像特征描述子研究领域一项里程碑式的工作。由于SIFT对尺度、旋转以及一定视角和光照变化等图像变化都具有不变性，并且SIFT具有很强的可区分性，的确让很多问题的解决变为可能。但它也不是万能的。



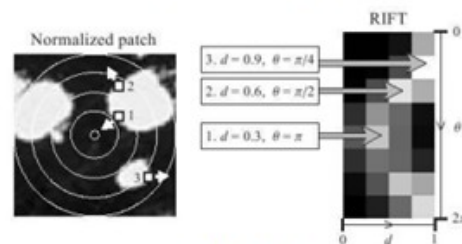
SIFT



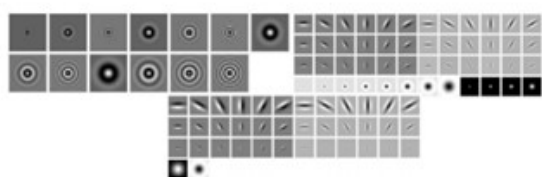
Spin image



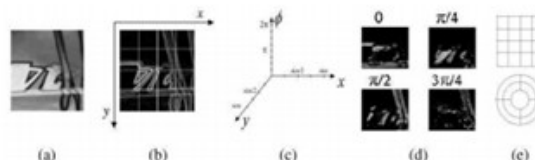
HoG



RIFT



Textons



GLOH

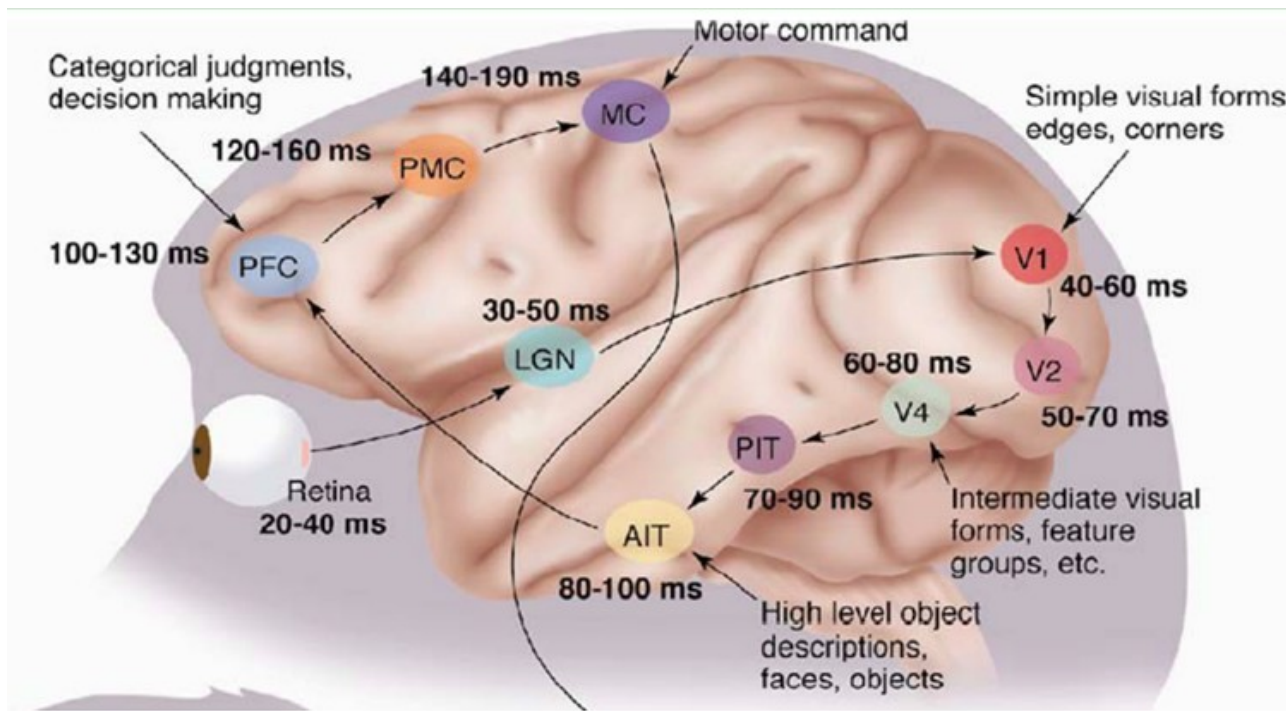
然而，手工地选取特征是一件非常费力、启发式（需要专业知识）的方法，能不能选取好很大程度上靠经验和运气，而且它的调节需要大量的时间。既然手工选取特征不太好，那么能不能自动地学习一些特征呢？答案是能！**Deep Learning**就是用来干这个事情的，看它的一个别名**Unsupervised Feature Learning**，就可以顾名思义了，**Unsupervised**的意思就是不要人参与特征的选取过程。

那它是怎么学习的呢？怎么知道哪些特征好哪些不好呢？我们说机器学习是一门专门研究计算机怎样模拟或实现人类的学习行为的学科。好，那我们人的视觉系统是怎么工作的呢？为什么在茫茫人海，芸芸众生，滚滚红尘中我们都可以找到另一个她（因为，你存在我深深的脑海里，我的梦里 我的心里 我的歌声里.....）。人脑那么**NB**，我们能不能参考人脑，模拟人脑呢？（好像和人脑扯上点关系的特征啊，算法啊，都不错，但不知道是不是人为强加的，为了使自己的作品变得神圣和高雅。）

近几十年以来，认知神经科学、生物学等等学科的发展，让我们对自己这个神秘的而又神奇的大脑不再那么的陌生。也给人工智能的发展推波助澜。

三、人脑视觉机理

1981 年的诺贝尔医学奖，颁发给了 **David Hubel**（出生于加拿大的美国神经生物学家）和**Torsten Wiesel**，以及 **Roger Sperry**。前两位的主要贡献，是“发现了视觉系统的信息处理”：可视皮层是分级的：



我们看看他们做了什么。1958 年，David Hubel 和 Torsten Wiesel 在 John Hopkins University，研究瞳孔区域与大脑皮层神经元的对应关系。他们在猫的后脑头骨上，开了一个3 毫米的小洞，向洞里插入电极，测量神经元的活跃程度。

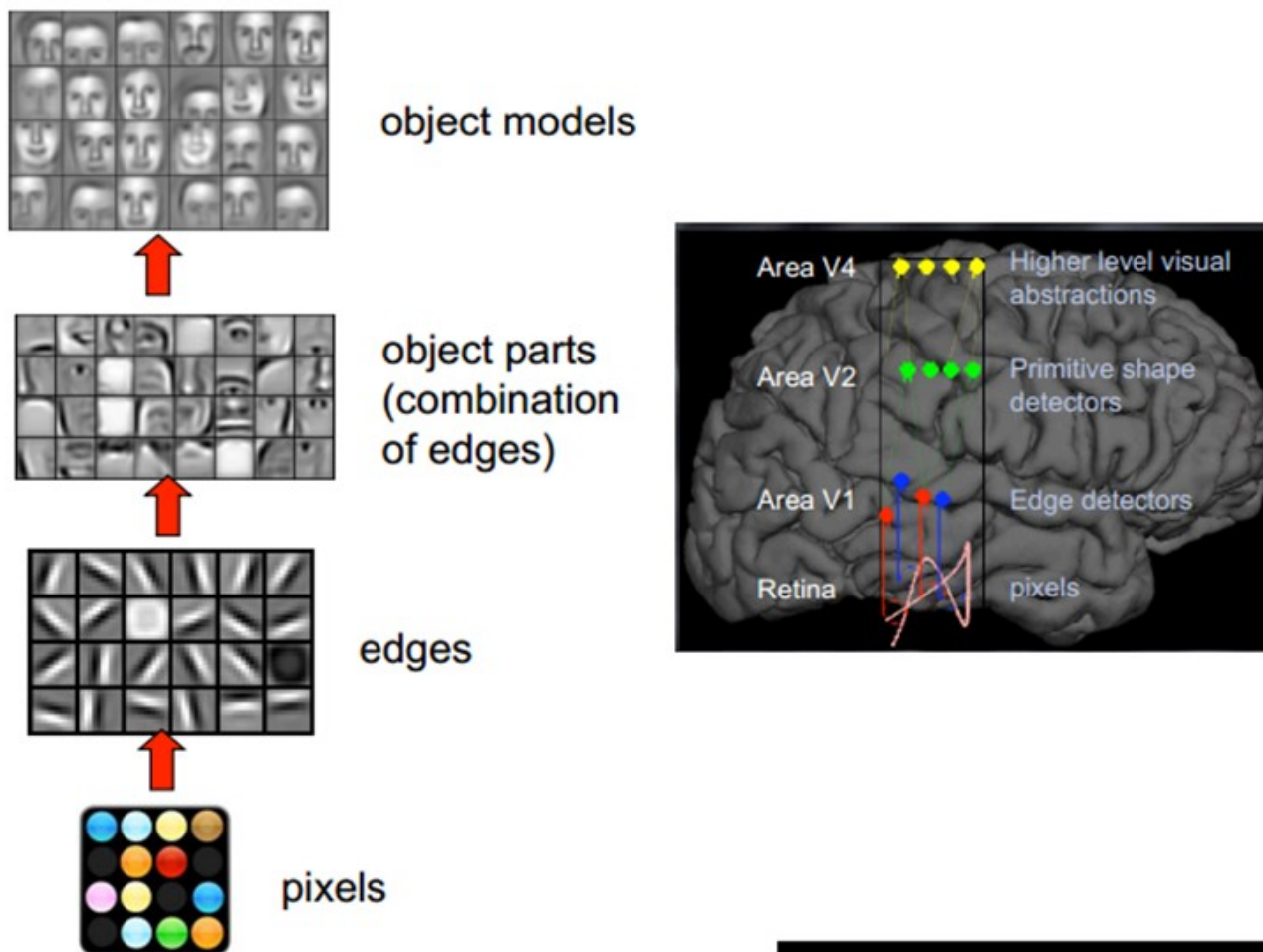
然后，他们在小猫的眼前，展现各种形状、各种亮度的物体。并且，在展现每一件物体时，还改变物体放置的位置和角度。他们期望通过这个办法，让小猫瞳孔感受不同类型、不同强弱的刺激。

之所以做这个试验，目的是去证明一个猜测。位于后脑皮层的不同视觉神经元，与瞳孔所受刺激之间，存在某种对应关系。一旦瞳孔受到某一种刺激，后脑皮层的某一部分神经元就会活跃。经历了很多天反复的枯燥的试验，同时牺牲了若干只可怜的小猫，David Hubel 和 Torsten Wiesel 发现了一种被称为“方向选择性细胞（Orientation Selective Cell）”的神经元细胞。当瞳孔发现了眼前的物体的边缘，而且这个边缘指向某个方向时，这种神经元细胞就会活跃。

这个发现激发了人们对于神经系统的进一步思考。神经-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。

这里的关键词有两个，一个是抽象，一个是迭代。从原始信号，做低级抽象，逐渐向高级抽象迭代。人类的逻辑思维，经常使用高度抽象的概念。

例如，从原始信号摄入开始（瞳孔摄入像素 Pixels），接着做初步处理（大脑皮层某些细胞发现边缘和方向），然后抽象（大脑判定，眼前的物体的形状，是圆形的），然后进一步抽象（大脑进一步判定该物体是只气球）。



这个生理学的发现，促成了计算机人工智能，在四十年后的突破性发展。

总的来说，人的视觉系统的信息处理是分级的。从低级的V1区提取边缘特征，再到V2区的形状或者目标的部分等，再到更高层，整个目标、目标的行为等。也就是说高层的特征是低层特征的组合，从低层到高层的特征表示越来越抽象，越来越能表现语义或者意图。而抽象层面越高，存在的可能猜测就越少，就越利于分类。例如，单词集合和句子的对应是多对一的，句子和语义的对应又是多对一的，语义和意图的对应还是多对一的，这是个层级体系。

敏感的人注意到关键词了：分层。而Deep learning的deep是不是就表示我存在多少层，也就是多深呢？没错。那Deep learning是如何借鉴这个过程的呢？毕竟是归于计算机来处理，面对的一个问题就是怎么对这个过程建模？

因为我们要学习的是特征的表达，那么关于特征，或者说关于这个层级特征，我们需要了解地更深入点。所以在说Deep Learning之前，我们有必要再啰嗦下特征（呵呵，实际上是看到那么好的对特征的解释，不放在这里有点可惜，所以就塞到这儿了）。