



深度学习



覃雄派

提纲

- 深度学习
- 深度学习在文本、图像、视频、音频等方面的应用

深度学习

深度学习





深度学习

- **深度学习**

- 深度学习是21世纪初流行起来的机器学习方法，它依赖于更深层次的神神经网络
- 深度学习在图像识别、语音识别、自然语言处理、机器人等领域，获得了超过传统机器学习方法的性能
 - 在人脸识别(Face Recognition)比赛LFW和自然图像分类比赛ImageNet中，获得了超过人类的识别能力
 - 2016年，Google的Alpha GO围棋程序，击败了人类棋手李世石九段，再次显示了深度学习技术的强大威力

深度学习

- **深度学习**

- 深度学习能够流行起来的原因包括几个方面，包括大数据集的积累、计算机运算能力的提高、深度学习训练算法的改进、以及深度学习模型具有能够自主从数据上学习到有用的特征(Feature)等
- 深度学习流行起来的其中一个因素，是人们找到了提高深度神经网络模型训练效率的方法
 - 主要的贡献者是Toronto大学的Hinton教授，2006年，他在Science杂志上发表了深度学习的里程碑式的论文，该论文将深度学习的性能提升了一大截

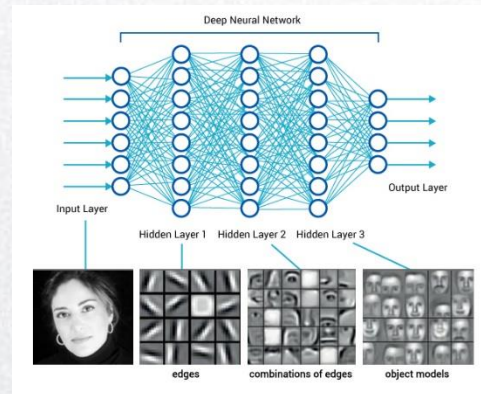


深度学习

- **深度神经网络：自动识别样本的特征**

- 深度神经网络可以自动识别样本的特征

- 这一点，使得深度学习在一些不知如何设计有效的特征的应用场合，比如图像识别和语音识别等，获得了很好的性能
 - 在神经网络中，浅层的神经元学习到初级的简单的特征(Primitive Feature)，馈入下一层神经网络层，深层的神经元在前一层神经元识别到的特征的基础上，学习到更加复杂的特征(Complex Features)
 - 这个过程在相邻的神经网络层间重复，各个神经网络层学习到不同抽象级别的特征，越是靠后的神经网络层，学习到更加抽象的特征，最后完成预定的识别任务，比如语音识别和图像识别
 - 比如，在图像识别中，第一个隐藏层学习到的是“边缘”的特征，第二个隐藏层学习到的是由“边缘”组成的“形状”的特征，第三个隐藏层学习到的是由“形状”组成的“图案”的特征，最后的隐藏层学习到的是由“图案”组成的“对象”的特征等



<https://arxiv.org/pdf/1311.2901.pdf>

http://vision.cse.psu.edu/people/chrisF/deep-learning/DL_pres.pdf

深度学习

- **深度学习的应用：Image Net**

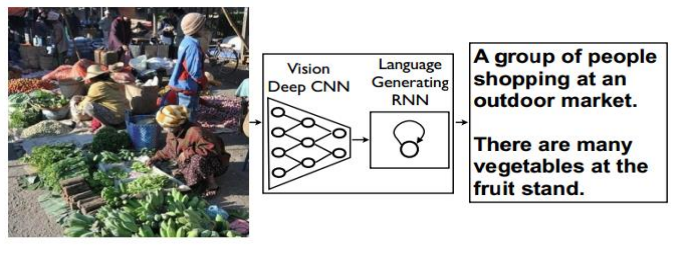
- 2012年，多伦多大学的Krizhevsky等构造了一个大型的卷积神经网络，该网络共有9层，65万个神经元，6千万个参数
- 网络的输入是图片，输出是1000个图片分类，表示不同的对象类型，比如美洲豹、救生艇等
- 他们使用大量的图片训练这个模型，最后在ImageNet图片分类方面，识别性能优于当时所有的其它分类器，错误率由25%降低为17%
- ImageNet 是 Stanford 大学 Li Feifei 教授创建的目前最大的图像识别 (Image Recognition) 数据库，目前共包含大约22000类，15M左右的标定图像
 - 其中，目前最常用的LSVRC-2010 contest包含1000类，1.2M图像



深度学习

- **深度学习的应用**

- 深度学习的应用非常广泛，包括图像/视频的识别、语音识别、自然语言处理等。
 - 在图像和视频应用方面，深度学习模型可以识别照片中的物体，对照片进行自动分类和搜索，比如Google Photo、百度识图、淘宝拍立淘等，都使用了深度学习模型
 - 深度学习模型应用于自动驾驶系统，对人员、车辆等路况信息进行识别和追踪，进而做出有效的应对
 - 深度学习模型还可以用于人脸识别，实现刷脸支付等功能，为人们的生活带来方便
 - 2014年，Google试图利用深度学习技术，从图像直接生成一段自然语言的描述
 - Google把两个深度神经网络结合起来(卷积神经网络CNN和循环神经网络RNN)，组成一个模型，完成这个任务。其中一个神经网络负责图像识别，另外一个神经网络负责语言生成。下图展示了他们使用的网络结构；这个网络结构，基于左边的图片，生成了右边的文字描述





深度学习

• 深度学习的应用

- 2014年，香港中文大学教授汤晓鸥、王晓刚及其研究团队，利用深度学习技术，研发的DeepID人脸识别技术，准确率达到99.15%，比人用肉眼识别更加精准
 - LFW是人脸识别领域使用最广泛的测试基准，人用肉眼在LFW上的识别率为97.52%
 - 该研究组早已在2011年，开始开展深度学习方法的研究，在2013年把LFW上的识别率提高到92.52%
- 2015年末，微软亚洲研究院，利用深度达152层的“深层残差网络”，参加ImageNet挑战赛；该网络的层数比以往任何成功使用的神经网络的层数多5倍以上，并且有效避免了梯度消失
 - 他们以绝对优势获得图像分类、图像定位以及图像检测全部三个主要项目的冠军
 - 此外，他们在另外一项图像识别挑战赛MSCOCO(Microsoft Common Objects in Context, 常见物体图像识别)中，同样获得冠军
 - 2014年ImageNet挑战赛获胜的系统错误率为6.6%，而2015年微软系统的错误率已经低至3.57%
 - 该团队在2015年1月，首次实现了超过人类视觉分类能力的突破，他们系统的错误率已降低至4.94%，在同样的实验中，人眼辨识的错误率大概为5.1%

• 深度学习的应用

- 在语音识别方面，深度学习也得到了广泛应用
 - 在深度学习技术的帮助下，计算机拥有强大的语音识别(speech recognition)能力，人机交互的模式将变得更加丰富
 - Google(Now)、Apple(Siri)、Microsoft(Cortana)、Baidu(Deep Speech)、科大讯飞等公司，都推出了其语音识别产品
 - 2012年12月，微软亚洲研究院展示了中英即时传译系统，集成了语音识别、机器翻译和语音合成技术，其错误率仅为7%(<http://tech2ipo.com/56452>)
 - 2012年，Google公司利用深度学习技术，大幅度改善了Android操作系统上的语音识别的精度
 - 使用深度学习技术，语音识别的成功率大大提高，尤其是在嘈杂的环境中，智能手机语音识别系统的错误率就下降到了25%，语音搜索结果也有了不小的改善
 - 2013年，多伦多大学的Alex Graves、Abdel-rahman Mohamed以及Geoffrey E. Hinton 等，使用双向LSTM/RNN打破了著名的TIMIT语音识别测试的记录
 - TIMIT语音库为研究中常用的语音库，适用于语音识别、说话人识别等语音信号处理



深度学习

- **深度学习的局限性**

- 深度学习技术并非万能的技术，人们不应对其过于迷信
- 它学习到的可能是数据中的相关关系，但不一定是因果关系
- 深度学习的能力很强，但是和我们预期的真正的人工智能相比，仍然缺乏必要的能力，比如逻辑推理的能力，集成抽象的知识的能力
 - 所以深度学习可以看作是实现人工智能的一种途径，而不是终极解决方案
- 在工程实践中，把深度学习技术和其它机器学习技术结合起来，比如贝叶斯推理和演绎推理等技术，互相取长补短，是一个有前途的策略
 - 比如，深度学习技术可以和迁移学习、增强学习(Reinforcement Learning)技术相结合
 - 所谓增强学习，是指计算机通过与环境交互，从中得到的奖赏和惩罚，进而自主学习(Self-Learning)更优的策略
 - 2016年3月份，击败李世石的围棋程序Alpha Go，使用了深度学习、增强学习、基于蒙特卡洛的搜索策略(MCST, Monte Carlo search tree)等方法，验证了深度学习技术和其它机器学习技术结合，可以使得计算机能够不断自主学习，从而获得高度优化的训练模型