

深度学习

原理与应用实践

张重生 著

精解深度学习原理，实战深度学习应用



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

深度学习

原理与应用实践

张重生 著

电子工业出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书全面、系统地介绍深度学习相关的技术，包括人工神经网络，卷积神经网络，深度学习平台及源代码分析，深度学习入门与进阶，深度学习高级实践。所有章节均附有源程序，所有实验读者均可重现，具有高度的可操作性和实用性。通过学习本书，研究人员、深度学习爱好者，能够在3个月内，系统掌握深度学习相关的理论和技术。

本书内容新颖、层次清晰，适合高校教师、研究人员、研究生、高年级本科生、深度学习爱好者使用。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目(CIP)数据

深度学习：原理与应用实践 / 张重生著. —北京：电子工业出版社，2016.12
ISBN 978-7-121-30413-2

I. ①深… II. ①张… III. ①学习系统 IV. ①TP273

中国版本图书馆CIP数据核字(2016)第280651号

责任编辑：赵娜

印刷：北京季蜂印刷有限公司

装订：北京季蜂印刷有限公司

出版发行：电子工业出版社

北京市海淀区万寿路173信箱 邮编 100036

开本：720×1000 1/16 印张：14.5 字数：259千字

版次：2016年12月第1版

印次：2017年4月第2次印刷

定 价：48.00元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010) 88254888，88258888。

质量投诉请发邮件至 zllts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：(010) 88254694。

序言 1



如今，深度学习是国际上非常活跃、非常多产的研究领域，它被广泛应用于计算机视觉、图像分析、语音识别和自然语言处理等诸多领域。在多个领域上，深度神经网络已大幅超越了已有算法的性能。

本书是深度学习领域的一本力作。它对深度神经网络尤其是卷积神经网络进行介绍，且注重深度学习的实际应用。而且，本书还对深度学习研发现状进行总结和阐述，包括对 Google 和 Facebook 的研究与总结。

本书通过示例的方式详解深度学习的具体应用，包括手写数字识别，物体识别，及以人为中心的计算（包括人脸识别、人脸表情识别、年龄估计、人脸关键点定位等）。

本书也介绍了深度学习 Caffe 和 Pylearn2 两个平台，并给出具体示例，介绍如何使用。

本书的所有实验均可重现，对初学者、研究生和工程师有重要参考价值，能够帮助读者掌握深度学习的实战技能。

我在访问河南大学期间与本书作者相识。作者和他的学生为本书的出版投入了 1 年多的时间，付出了巨大的心血。我相信，本书将会被中国科技界所认可。

Ioannis Pitas
IEEE Fellow
亚里士多德大学

序言 2



深度学习是当前最炙手可热的研究领域，可是缺乏一本系统深入的著作。今天，这本书来了！

本书是目前国内为数不多的深度学习原著之一，更是最接地气的、同时兼顾深度学习理论与应用开发实践的深度学习著作。本书面向深度学习领域的研究人员、工程师、爱好者，能够帮助人们把握国内外深度学习的研究现状和产业化趋势，快速理解深度学习的基础理论，牢固掌握深度学习应用开发的实用技术。无论是本书的深度学习理论介绍还是应用开发实例详解，作者都采用实例驱动的方式，使得抽象的理论变得容易理解、难懂的图像处理与模式识别问题变得具体、落地。尤其难能可贵的是，本书还对深度学习平台 Caffe 的源代码进行了精彩解析。

本书的 8 个应用及对应实验都经过作者的反复验证，只需按照书中的相关步骤进行实验，读者就可以重现相关的实验结果。通过本书的学习，读者不但能掌握深度学习相关的基础理论，更能切实掌握深度学习的应用开发技能。因此，对于有意从事深度学习研究与应用实践的研究人员、工程师、爱好者，本书是首选之作。

教授

中国信息协会大数据分会副会长
中国最畅销的《云计算》书籍作者

前 言



深度学习是当今最流行和最受关注的信息技术之一。本书的特点是理论与实践兼备，以举例的方式介绍神经网络和卷积神经网络的原理，然后，通过实例讲解、实战操作的方式，讲解深度学习的具体应用，有利于读者动手能力的培养和解决问题能力的提升。

本书旨在降低深度学习的门槛，成为通用、普及性强的深度学习书籍，方便深度学习爱好者快速上手，帮助读者快速入门、进阶、精通深度学习相关的技术，适用于研究人员、高校教师、深度学习爱好者、研究生、高年级本科生。

本书共 15 章，分为 6 篇，第 1 篇是深度学习基础篇，包括第 1~2 章，分别是绪论和国内外深度学习技术研发现状及其产业化趋势。第 2 篇是深度学习理论篇，包括第 3~4 章，分别讲解神经网络和卷积神经网络。第 3 篇是深度学习工具篇，包括第 5~6 章，分别是深度学习工具 Caffe 和 Pylearn2 的介绍和使用详解。第 4 篇是深度学习实践篇（入门与进阶），包括第 7~10 章，分别讲解基于深度学习解决手写数字识别、图像识别、物体图像识别、人脸识别等常见的应用问题。第 5 篇是深度学习实践篇（高级应用），包括第 11~14 章，分别讲解基于深度学习的高级人脸识别算法 DeepID、表情识别、年龄估计和人脸关键点检测。第 6 篇是深度学习总结与展望篇，包括第 15 章，介绍对深度学习的总结和展望。

本书所有章节中的示例内容具有高度的可重现性，读者按照本书的步骤进行操作，编写对应的程序，就能重现相应的实验结果。

本书有助于读者快速入门深度学习的研究和应用实践，帮助读者在较短的时间内掌握相关的技术，并从事基于深度学习的应用开发。然而，深度学习的理论和技术发展极其迅速，尤其是深度学习的理论，需要长期的研究和积淀才能全面理解、深入掌握。本书仅仅介绍了深度学习的基本原理，更多全面、深入、前沿的原理和技术，需要读者自行钻研、主动探究。

本书由张重生和王朋友合著。硕士研究生裴宸平、于珂珂也参与了部分章节的撰写工作，在此致谢。本书的出版，得到了国家自然科学基金(41401466)、省属高校基本科研业务费(xxjc20140005)、河南省科技攻关项目(132102210188)、河南大学研究生院“大数据汇聚与分析”双语课程建设项目(Y1513004)的资助。

感谢中国信息协会大数据分会副会长刘鹏教授，IEEE Fellow、亚里士多德大学 Ioannis Pitas 教授为本书作序。感谢电子工业出版社董亚峰先生对本书的大力支持和无私帮助。

笔者自认才疏学浅，仅略知深度学习之皮毛。书中错谬之处在所难免，如蒙读者诸君不吝告知(本书作者邮箱：chongsheng.zhang@yahoo.com，微信号：A13938613173)，将不胜感激。

张重生

2016年11月

目 录



深度学习基础篇

第 1 章 绪论	2
1.1 引言	2
1.1.1 Google 的深度学习成果	2
1.1.2 Microsoft 的深度学习成果	3
1.1.3 国内公司的深度学习成果	3
1.2 深度学习技术的发展历程	4
1.3 深度学习的应用领域	6
1.3.1 图像识别领域	6
1.3.2 语音识别领域	6
1.3.3 自然语言理解领域	7
1.4 如何开展深度学习的研究和应用开发	7
参考文献	11
第 2 章 国内外深度学习技术研发现状及其产业化趋势	13
2.1 Google 在深度学习领域的研发现状	13
2.1.1 深度学习在 Google 的应用	13
2.1.2 Google 的 TensorFlow 深度学习平台	14
2.1.3 Google 的深度学习芯片 TPU	15
2.2 Facebook 在深度学习领域的研发现状	15
2.2.1 Torchnet	15
2.2.2 DeepText	16
2.3 百度在深度学习领域的研发现状	17
2.3.1 光学字符识别	17

2.3.2	商品图像搜索	17
2.3.3	在线广告	18
2.3.4	以图搜图	18
2.3.5	语音识别	18
2.3.6	百度开源深度学习平台 MXNet 及其改进的 深度语音识别系统 Warp-CTC	19
2.4	阿里巴巴在深度学习领域的研发现状	19
2.4.1	拍立淘	19
2.4.2	阿里小蜜——智能客服 Messenger	20
2.5	京东在深度学习领域的研发现状	20
2.6	腾讯在深度学习领域的研发现状	21
2.7	科创型公司（基于深度学习的人脸识别系统）	22
2.8	深度学习的硬件支撑——NVIDIA GPU	23
	参考文献	24

深度学习理论篇

第 3 章	神经网络	30
3.1	神经元的概念	30
3.2	神经网络	31
3.2.1	后向传播算法	32
3.2.2	后向传播算法推导	33
3.3	神经网络算法示例	36
	参考文献	38
第 4 章	卷积神经网络	39
4.1	卷积神经网络特性	39
4.1.1	局部连接	40
4.1.2	权值共享	41
4.1.3	空间相关下采样	42
4.2	卷积神经网络操作	42
4.2.1	卷积操作	42
4.2.2	下采样操作	44
4.3	卷积神经网络示例：LeNet-5	45
	参考文献	48

深度学习工具篇

第 5 章 深度学习工具 Caffe	50
5.1 Caffe 的安装	50
5.1.1 安装依赖包	51
5.1.2 CUDA 安装	51
5.1.3 MATLAB 和 Python 安装	54
5.1.4 OpenCV 安装 (可选)	59
5.1.5 Intel MKL 或者 BLAS 安装	59
5.1.6 Caffe 编译和测试	59
5.1.7 Caffe 安装问题分析	62
5.2 Caffe 框架与源代码解析	63
5.2.1 数据层解析	63
5.2.2 网络层解析	74
5.2.3 网络结构解析	92
5.2.4 网络求解解析	104
参考文献	109
第 6 章 深度学习工具 Pylearn2	110
6.1 Pylearn2 的安装	110
6.1.1 相关依赖安装	110
6.1.2 安装 Pylearn2	112
6.2 Pylearn2 的使用	112
参考文献	116

深度学习实践篇 (入门与进阶)

第 7 章 基于深度学习的手写数字识别	118
7.1 数据介绍	118
7.1.1 MNIST 数据集	118
7.1.2 提取 MNIST 数据集图片	120
7.2 手写数字识别流程	121
7.2.1 模型介绍	121
7.2.2 操作流程	126

7.3 实验结果分析	127
参考文献	128
第8章 基于深度学习的图像识别	129
8.1 数据来源	129
8.1.1 Cifar10 数据集介绍	129
8.1.2 Cifar10 数据集格式	129
8.2 Cifar10 识别流程	130
8.2.1 模型介绍	130
8.2.2 操作流程	136
8.3 实验结果分析	139
参考文献	140
第9章 基于深度学习的物体图像识别	141
9.1 数据来源	141
9.1.1 Caltech101 数据集	141
9.1.2 Caltech101 数据集处理	142
9.2 物体图像识别流程	143
9.2.1 模型介绍	143
9.2.2 操作流程	144
9.3 实验结果分析	150
参考文献	151
第10章 基于深度学习的人脸识别	152
10.1 数据来源	152
10.1.1 AT&T Facedatabase 数据库	152
10.1.2 数据库处理	152
10.2 人脸识别流程	154
10.2.1 模型介绍	154
10.2.2 操作流程	155
10.3 实验结果分析	159
参考文献	160

深度学习实践篇 (高级应用)

第 11 章 基于深度学习的人脸识别——DeepID 算法	162
11.1 问题定义与数据来源	162
11.2 算法原理	163
11.2.1 数据预处理	163
11.2.2 模型训练策略	164
11.2.3 算法验证和结果评估	164
11.3 人脸识别步骤	165
11.3.1 数据预处理	165
11.3.2 深度网络结构模型	168
11.3.3 提取深度特征与人脸验证	171
11.4 实验结果分析	174
11.4.1 实验数据	174
11.4.2 实验结果分析	175
参考文献	176
第 12 章 基于深度学习的表情识别	177
12.1 表情数据	177
12.1.1 Cohn-Kanade (CK+) 数据库	177
12.1.2 JAFFE 数据库	178
12.2 算法原理	179
12.3 表情识别步骤	180
12.3.1 数据预处理	180
12.3.2 深度神经网络结构模型	181
12.3.3 提取深度特征及分类	182
12.4 实验结果分析	184
12.4.1 实现细节	184
12.4.2 实验结果对比	185
参考文献	188
第 13 章 基于深度学习的年龄估计	190
13.1 问题定义	190
13.2 年龄估计算法	190

13.2.1	数据预处理	190
13.2.2	提取深度特征	192
13.2.3	提取 LBP 特征	196
13.2.4	训练回归模型	196
13.3	实验结果分析	199
	参考文献	199
第 14 章	基于深度学习的人脸关键点检测	200
14.1	问题定义和数据来源	200
14.2	基于深度学习的人脸关键点检测的步骤	201
14.2.1	数据预处理	201
14.2.2	训练深度学习网络模型	206
14.2.3	预测和处理关键点坐标	207
	参考文献	212

深度学习总结与展望篇

第 15 章	总结与展望	214
15.1	深度学习领域当前的主流技术及其应用领域	214
15.1.1	图像识别	214
15.1.2	语音识别与自然语言理解	215
15.2	深度学习的缺陷	215
15.2.1	深度学习在硬件方面的门槛较高	215
15.2.2	深度学习在软件安装与配置方面的门槛较高	216
15.2.3	深度学习最重要的问题在于需要海量的 有标注的数据作为支撑	216
15.2.4	深度学习的最后阶段竟然变成枯燥、机械、 及其耗时的调参工作	217
15.2.5	深度学习不适用于数据量较小的数据	218
15.2.6	深度学习目前主要用于图像、声音的识别和 自然语言的理解	218
15.2.7	研究人员从事深度学习研究的困境	219
15.3	展望	220
	参考文献	220

深度学习基础篇

第 1 章

绪 论



1.1 引言

深度学习 (Deep Learning)^[1], 是当今人工智能/机器学习领域研究和应用开发的热点。在“大数据+深度学习”的共同推动下, 深度学习在 ImageNet 图像分类竞赛、语音理解、图像识别、视频分析、无人驾驶汽车领域都取得了重要突破。与传统方法不同, 深度学习首先通过大规模的迭代实验 (调参实验) 逼近所能达到的最高识别准确率, 然后使用对应的 (参数和) 模型对新样本 (图像、声音等) 提取关键特征, 并基于该特征, 利用已训练好的分类模型预测新样本的类别。目前, Google、Facebook、Microsoft 等国际巨头, 以及百度、腾讯、阿里巴巴、京东等国内互联网巨头都已投入巨资布局深度学习, 并将其作为重要的研发方向。

1.1.1 Google 的深度学习成果

2015 年 10 月, Google (谷歌) 旗下 DeepMind 公司研发了人工智能围棋程序, 该程序主要使用深度学习的技术, 整体上包含离线学习和在线对弈两个过程, 其中离线学习主要利用大量已有棋谱进行训练“价值网络”去计算局面优劣, 训练“策略网络”去选择下子位置; 在线对弈主要利用“价值网络”计算当前棋面优劣, 利用“策略网络”计算当前应该选择的下子位置。2015 年, 阿尔法围棋 (AlphaGo) 以 5:0 的总比分击败欧洲围棋冠军樊麾; 紧接着, 2016 年 3 月, 以

4:1 的总比分击败世界围棋冠军、职业九段选手李世石^[2]。

而在此之前,2011年谷歌就成立了由人工智能和机器学习顶级学者吴恩达(Andrew Ng)领衔的“Google Brain”项目,这个项目利用谷歌的分布式计算框架训练深度人工神经网络。该项目的主要成果是使用包含16000个CPU核的并行计算平台,使用基于深度学习算法训练超过10亿个神经元的深度神经网络^[3],该系统能够在没有任何先验知识的前提下,自动学习YouTube网站上海量的视频数据,训练深度神经网络。吴恩达目前是斯坦福大学计算机科学系和电子工程系副教授、人工智能实验室主任,并担任百度公司首席科学家,负责百度研究院的百度大脑计划。

1.1.2 Microsoft 的深度学习成果

2012年,微软首席研究官Rick Rashid在“21世纪计算大会”上的英文演讲被实时翻译成与他音色很接近的中文演讲,该功能主要借助于基于深度学习技术实现的自动同声传译系统^[4],自动同声传译过程主要是语音识别、机器翻译和语音合成。

1.1.3 国内公司的深度学习成果

2013年,百度成立了由知名学者余凯领导的百度深度学习研究院(Institute of Deep Learning, IDL),主要目标是将深度学习应用于语音识别和图像识别、智能检索等领域。现在,基于深度学习,百度的图像搜索更加准确,百度翻译更加专业,语音识别效果令人十分满意。目前,许多基于深度学习的产品已经面市,例如百度识别APP,该APP主要功能是图像识别和智能检索,其中拍照购物和通过照片匹配度来交友都是该APP中比较有特色的功能。百度在“小度机器人”和无人驾驶汽车领域等都取得了重要进展。小度机器人能够通过对话等自然的交互方式,准确理解用户意图,并与用户进行信息和服务等的交流。

阿里巴巴的“拍立淘”是基于“大数据+深度学习+图像处理”的构思开发的,网购用户通过手机拍照,利用“拍立淘”就能在淘宝中找到非常类似的产品,其搜索准确度和用户满意度非常高。

LinkFace(脸云科技)在2014年开创了基于深度学习的人脸检测算法,支持人脸检测、人脸识别、人脸关键点检测等全套技术,在FDDB数据集上的人

脸识别准确率高达 99.5%。图森 (<http://www.tusimple.com/>) 通过深度学习引擎, 研发了图像识别和语义分析技术, 为企业搭建了自己的图片识别服务, 根据企业的实际业务设计了分类标签系统, 精准描述企业图片分类需求。该公司还研发了基于摄像头的智能驾驶解决方案。

1.2 深度学习技术的发展历程

深度学习是基于人工神经网络发展起来的一项技术, 而神经网络的起源可以追溯到 20 世纪 40 年代。神经网络的基本组成单元为神经元, 1943 年心理学家 W.S.McCulloch 和数理逻辑学家 W.Pitts 首次提出神经元计算模型, 并讲述了神经元的形式化数学描述和网络结构方法, 验证了单个神经元具有逻辑功能^[5], 从此打开了人工神经网络研究领域的大门。

人工神经网络在全世界科研工作者的持续努力下不断发展, 1957 年, Frank Rosenblatt 首次提出“感知器”的概念^[6], 该项工作在当时得到了广泛应用, 在许多研究领域都取得了一定的成果。但由于当时计算条件的限制, 并没有能够持续很长时间。此外, 还有一个主要原因是 1969 年 Marvin Minsky 和 Seymour Papert 出版的书籍 *Perceptron*, 该书主要论证了“感知器”不能解决异或问题(也就是“异或门”), 而且当时的计算机硬件满足不了神经网络的复杂计算。此后, 神经网络的发展陷入低潮, 在很长一段时间都没有得到重视。

尽管如此, Geoffrey Hinton 一直在坚持神经网络领域的研究, 1986 年, DE Rumelhart, G.E. Hinton 和 R.J. Williams 共同在 *Nature* 上发表论文^[7], 提出误差后向传播 (Error Back Propagation, EBP) 算法, 该算法解决了“感知器”的“异或”问题, 降低了神经网络的计算复杂度, 在解决相同问题的前提下, 基于误差后向传播算法的神经网络的效果要优于简单“感知器”。误差后向传播算法在各个领域得到广泛应用, 并延续至今, 重新掀起了神经网络高潮。1989 年, Yann Lecun 成功将误差后向传播算法应用到手写邮政编码的识别任务上^[8]。Yann Lecun 在人工神经网络的基础上提出了“卷积神经网络”, 目前卷积神经网络在图像和视频分析等领域都取得了很好的成果。

深度神经网络是模仿人脑机制构建的具有学习和分析解决问题的神经网络, 它是由输入层、若干隐层和输出层构成的多层网络结构。相邻层之间的神经元通