

关于深度学习的专题性著作，也是掌握卷积神经网络内涵的进阶之书

Deep Learning
Mastering Convolutional Neural Networks from Beginner

深度学习

卷积神经网络从入门到精通

李玉鑑 张婷 单传辉 刘兆英 等著



机械工业出版社
China Machine Press

■ ■ ■ 智能系统与技术丛书

Deep Learning
Mastering Convolutional Neural Networks from Beginner

深度学习

卷积神经网络从入门到精通

李玉鑑 张婷 单传辉 刘兆英 等著



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

深度学习：卷积神经网络从入门到精通 / 李玉鑑等著. —北京：机械工业出版社，2018.7
(智能系统与技术丛书)

ISBN 978-7-111-60279-8

I. 深… II. 李… III. 学习系统 IV. TP273

中国版本图书馆 CIP 数据核字 (2018) 第 137837 号

深度学习：卷积神经网络从入门到精通

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号 邮政编码：100037）

责任编辑：张梦玲

责任校对：李秋荣

印刷：北京市兆成印刷有限责任公司

版次：2018 年 7 月第 1 版第 1 次印刷

开本：186mm × 240mm 1/16

印张：27

书号：ISBN 978-7-111-60279-8

定价：79.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

客服热线：(010) 88379426 88361066

投稿热线：(010) 88379604

购书热线：(010) 68326294 88379649 68995259

读者信箱：hzit@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问：北京大成律师事务所 韩光 / 邹晓东

主要作者介绍

李玉鑑（鉴）

北京工业大学教授，博士生导师。华中理工大学（现名为华中科技大学）本科毕业，中国科学院数学研究所硕士毕业，中国科学院半导体研究所博士毕业，北京邮电大学博士后出站。曾在中国科学院生物物理所工作，对意识的本质问题关注过多年，并在《21世纪100个交叉科学难题》上发表《揭开意识的奥秘》一文，提出了解决意识问题的认知相对论纲领，对脑计划和类脑研究具有宏观指导意义。长期围绕人工智能的核心目标，在神经网络、自然语言处理、模式识别和机器学习等领域开展教学、科研工作，发表国内外期刊、会议论文数十篇，是本书和《深度学习导论及案例分析》的第一作者。

P R E F A C E

前 言

随着谷歌的 AlphaGo、IBM 的 Watson 和百度的小度机器人等智能产品的问世，人工智能成为大众热烈讨论的焦点。深度学习作为其中的核心技术之一，经过学术界与工业界的积极推动，已经被广泛应用于计算机视觉、语音识别和自然语言处理等诸多领域。如果读者想了解深度学习的总体概况，可参考作者编写的《深度学习导论及案例分析》^①一书。

本书专注讨论深度学习中应用非常广泛的模型——卷积神经网络，该模型特别适用于图像分类和识别、目标分割和检测以及人工智能游戏方面，受众对象包括计算机、自动化、信号处理、机电工程、应用数学等相关专业的研究生、教师以及算法工程师和科研工作者。

卷积神经网络是一种特殊的多层感知器或前馈神经网络，具有局部连接、权值共享的特点，其中大量神经元按照一定方式组织起来对视野中的交叠区域产生反应。其前身是日本学者 Fukushima 在感受野概念的基础上提出的神经认知机模型。利用神经认知机的思想，LeCun 等人在 1998 年提出了卷积神经网络的现代雏形 LeNet。2012 年，Krizhevsky 等人取得了卷积神经网络研究的重大突破，提出了著名的 AlexNet。AlexNet 在 ImageNet 的大规模图像分类竞赛中取得优异成绩，为深度学习的全面推广立下了汗马功劳。随后，卷积神经网络模型如雨后春笋般出现，如 VGGNet、GoogLeNet、SPPNet、ResNet、DenseNet、Faster R-CNN、YOLO、SSD、FCN、PSPNet、Mask R-CNN、SiameaseNet、SqueezeNet、DCGAN、NIN，以及在人工智能游戏中用到的深度强化模型等。

本书的最大特色是对卷积神经网络进行由浅入深的分类描述，依次包括：现代雏形、突破模型、应变模型、加深模型、跨连模型、区域模型、分割模型、特殊模型、强化模型和顶尖成就。这种分类框架是在模型概述和预备知识的基础上逐步展开的，既方

^① 此书已由机械工业出版社出版，书号为 ISBN 978-7-111-55075-4。——编辑注

便读者入门学习，又有助于读者深入钻研。

本书的另一大特色是结合 Caffe 或 TensorFlow 的代码来说明各种卷积神经网络模型的具体实现过程，并通过应用案例说明其价值和意义所在。典型的应用案例包括：字符识别、交通标志识别、交通路网提取、大规模图像分类、人脸图像性别分类、图像目标检测、图像语义分割、图像实例分割、人脸图像生成、Flappy Bird 智能体、AlphaGo 的仿效围棋程序等。读者可以通过运行各个应用案例的程序代码和实验数据，检验其演示效果。

与其他深度学习的书籍相比，本书对卷积神经网络的内容涵盖更为广泛、模型讨论更为深入、应用实践更为细致。特别是，还总结了一些运行卷积神经网络的配置技巧和操作经验。比如，在运行 Mask R-CNN 的时候，需要先安装读取 COCO 数据集的程序，然后再进行训练或测试。在运行 SSD 的时候，可视化结果只给出了类别编号而没有给出类别名，作者对此已进行了修改，以方便读者按照书中所示代码显示相应的类别名。把这些经过摸索得到的技巧和经验分享给读者，对提高读者的深度学习技术水平，无疑具有很好的加速作用。

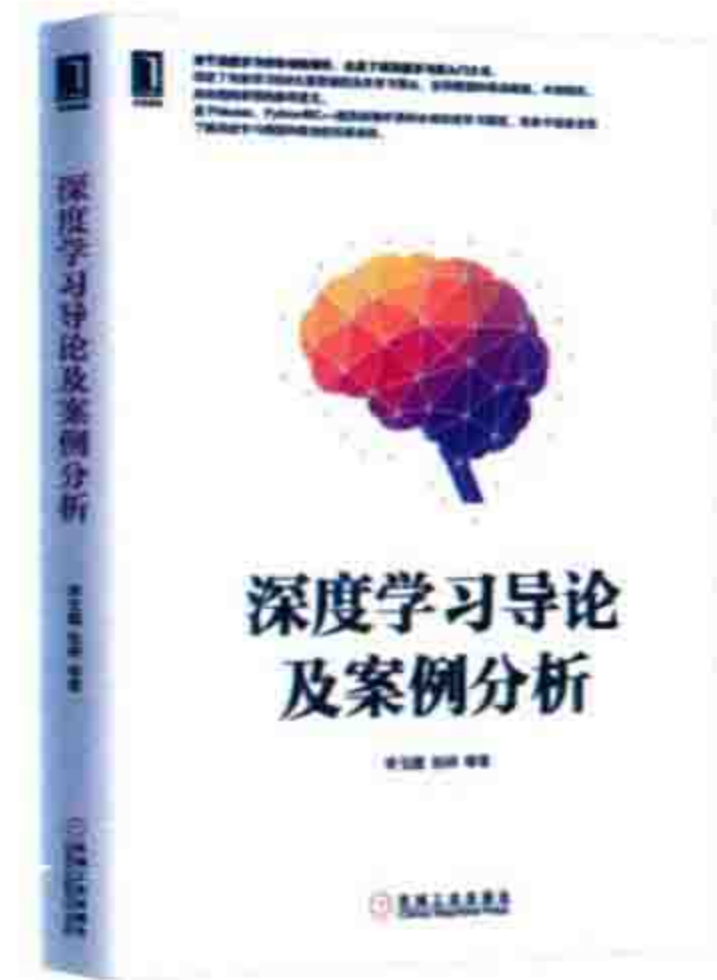
本书是集体努力的成果，主要作者包括北京工业大学的李玉鑑、张婷、单传辉、刘兆英、聂小广和欧军。他们对全书的内容进行了精心的布局、认真的编写和细致的整理。同时，曾少锋、刘博文、穆红章、余华擎和方皓达等人在文献资料、实现代码和软件工具的收集方面也提供了积极的帮助。此外，华章公司的温莉芳副总经理和张梦玲编辑对本书的排版提出了许多宝贵的意见。最后，需要特别感谢家人的支持，他们也在不知不觉中以各种方式对此书出版做出了贡献。

限于作者水平，本书难免在内容取材和结构编排上有不妥之处，希望读者不吝赐教，提出宝贵的批评和建议，我们将不胜感激。

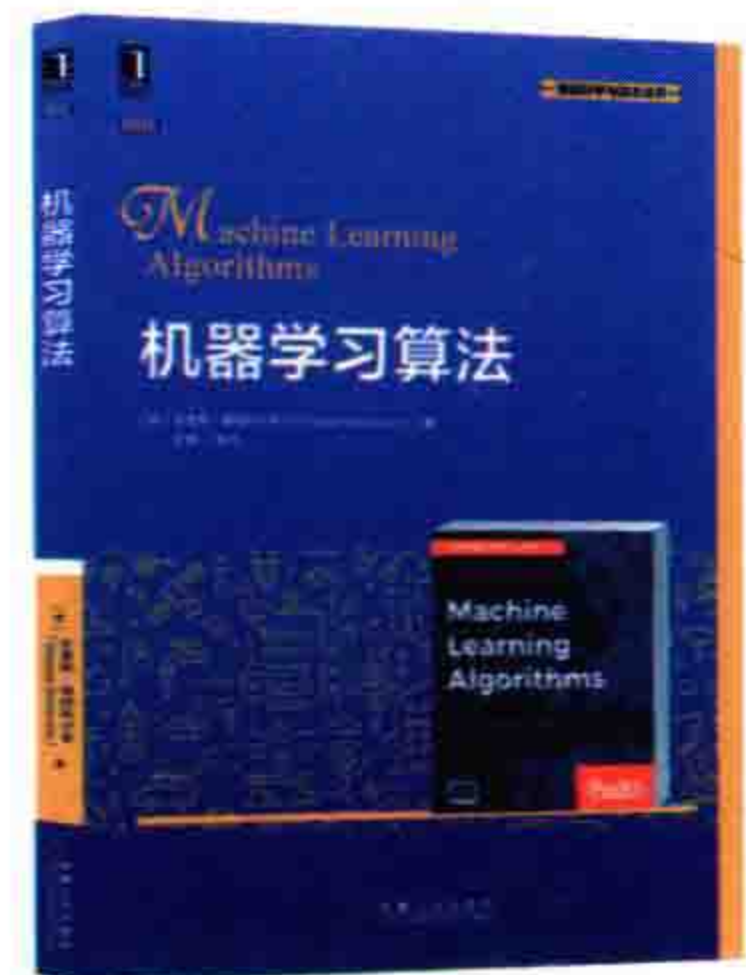
作者

2018年4月于北京工业大学

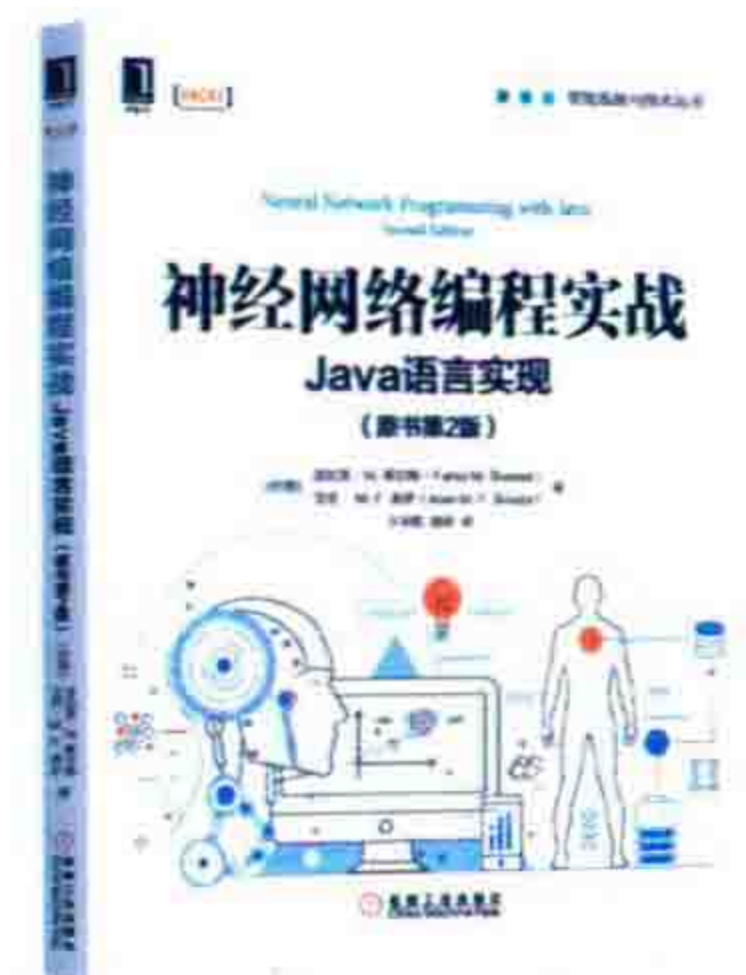
推荐阅读



书号：978-7-111-55075-4
定价：59.00元



书号：978-7-111-59513-7
定价：69.00元



书号：978-7-111-60012-1
定价：59.00元

目 录

前言

第 1 章 概述 1

1.1 深度学习的起源和发展 1

1.2 卷积神经网络的形成和演变 4

1.3 卷积神经网络的应用和影响 6

1.4 卷积神经网络的缺陷和视图 9

1.5 卷积神经网络的 GPU 实现和
cuDNN 库 10

1.6 卷积神经网络的平台和工具 10

1.7 本书的内容结构和案例数据 13

1.7.1 内容结构 13

1.7.2 案例数据 15

第 2 章 预备知识 22

2.1 激活函数 22

2.2 矩阵运算 23

2.3 导数公式 24

2.4 梯度下降算法 25

2.5 反向传播算法 26

2.5.1 通用反向传播算法 27

2.5.2 逐层反向传播算法 28

2.6 通用逼近定理 31

2.7 内外卷积运算 31

2.8 膨胀卷积运算 32

2.9 上下采样运算 33

2.10 卷积面计算 34

2.11 池化面计算 36

2.12 局部响应归一化 36

2.13 权值偏置初始化 37

2.14 丢失输出 37

2.15 丢失连接 38

2.16 随机梯度下降算法 39

2.17 块归一化 39

2.18 动态规划算法 40

第 3 章 卷积神经网络的现代

雏形——LeNet 41

3.1 LeNet 的原始模型 41

3.2 LeNet 的标准模型 43

3.3 LeNet 的学习算法 44

3.4 LeNet 的 Caffe 代码实现及说明 46

3.5 LeNet 的手写数字识别案例 54

3.6 LeNet 的交通标志识别案例 58

3.6.1 交通标志数据集的格式
转换 58

3.6.2 交通标志的识别分类 60

3.7 LeNet 的交通路网提取案例	63	6.1.1 VGGNet 的模型结构	118
3.7.1 交通路网的人工标注	64	6.1.2 VGGNet 的 TensorFlow	
3.7.2 交通路网的图像块分类	67	代码实现及说明	120
3.7.3 交通路网的图像块分类		6.1.3 VGGNet 的物体图像分类	
LeNet	69	案例	129
3.7.4 交通路网的自动提取		6.2 结构更深的卷积网络 GoogLeNet	130
代码及说明	71	6.2.1 GoogLeNet 的模型结构	130
3.7.5 交通路网的自动提取程序		6.2.2 GoogLeNet 的 TensorFlow	
运行结果	75	代码实现及说明	136
第 4 章 卷积神经网络的突破		6.2.3 GoogLeNet 的鲜花图像	
模型	78	分类案例	149
4.1 AlexNet 的模型结构	78	第 7 章 卷积神经网络的跨连	
4.2 AlexNet 的 Caffe 代码实现及		模型	154
说明	82	7.1 快道网络 HighwayNet	154
4.3 AlexNet 的 Caffe 大规模图像分类		7.2 残差网络 ResNet	155
案例及演示效果	95	7.2.1 ResNet 的模型结构	155
4.4 AlexNet 的 TensorFlow 代码		7.2.2 ResNet 的 Caffe 代码实现	
实现及说明	97	及说明	157
4.5 AlexNet 的 TensorFlow 大规模		7.2.3 ResNet 的大规模图像分类	
图像分类案例及演示效果	103	案例	163
4.6 AlexNet 的改进模型 ZFNet	107	7.3 密连网络 DenseNet	169
第 5 章 卷积神经网络的应变		7.3.1 DenseNet 的模型结构	169
模型	109	7.3.2 DenseNet 的 Caffe 代码	
5.1 SPPNet 的模型结构	109	实现及说明	171
5.2 SPPNet 的 Caffe 代码实现及		7.3.3 DenseNet 的物体图像分类	
说明	112	案例	174
5.3 SPPNet 的大规模图像分类		7.4 拼接网络 CatNet	178
案例及演示效果	114	7.4.1 CatNet 的模型结构	178
第 6 章 卷积神经网络的加深		7.4.2 CatNet 的 Caffe 代码	
模型	118	实现及说明	179
6.1 结构加深的卷积网络 VGGNet	118	7.4.3 CatNet 的人脸图像性别	
		分类案例	183

第 8 章 卷积神经网络的区域**模型** 190

8.1 区域卷积网络 R-CNN 190

8.2 快速区域卷积网络 Fast
R-CNN 1918.3 更快区域卷积网络 Faster
R-CNN 1938.3.1 Faster R-CNN 的模型
结构 1938.3.2 Faster R-CNN 的 Tensor
Flow 代码实现及说明 1968.3.3 Faster R-CNN 的图像目标
检测案例及演示效果 216

8.4 你只看一次网络 YOLO 220

8.4.1 YOLO 的模型结构 220

8.4.2 YOLO 的 TensorFlow 代码
实现及说明 2268.4.3 YOLO 的图像目标检测
案例及演示效果 239

8.5 单次检测器 SSD 242

8.5.1 SSD 的模型结构 242

8.5.2 SSD 的 TensorFlow 代码
实现及说明 2458.5.3 SSD 的图像目标检测
案例及演示效果 260**第 9 章 卷积神经网络的分割****模型** 266

9.1 全卷积网络 FCN 266

9.1.1 FCN 的模型结构 266

9.1.2 FCN 的 Caffe 代码
实现及说明 2699.1.3 FCN 的图像语义和几何
分割案例 272

9.2 金字塔场景分析网络 PSPNet 277

9.2.1 PSPNet 的模型结构 277

9.2.2 PSPNet 的 TensorFlow
代码实现及说明 2829.2.3 PSPNet 的图像语义分割
案例及演示效果 2919.3 掩膜区域卷积网络 Mask
R-CNN 2949.3.1 Mask R-CNN 的模型
结构 2949.3.2 Mask R-CNN 的 Keras 和
TensorFlow 代码实现及
说明 2979.3.3 Mask R-CNN 的图像实例
分割案例及演示效果 318**第 10 章 卷积神经网络的特殊****模型** 325

10.1 孪生网络 SiameseNet 325

10.1.1 SiameseNet 的模型
结构 32510.1.2 SiameseNet 的 Caffe
代码实现及说明 32610.1.3 SiameseNet 的手写数字
验证案例 328

10.2 挤压网络 SqueezeNet 331

10.2.1 SqueezeNet 的模型
结构 33110.2.2 SqueezeNet 的 Caffe
代码实现及说明 334

10.2.3	SqueezeNet 大规模图像 分类案例	337	11.4.4	笨笨鸟网络的演示 效果	370
10.3	深层卷积生成对抗网络		第 12 章	卷积神经网络的顶尖 成就——AlphaGo	371
	DCGAN	339	12.1	人工智能棋类程序简介	371
10.3.1	DCGAN 的模型结构	339	12.2	AlphaGo 的设计原理	373
10.3.2	DCGAN 的 TensorFlow 代码实现及说明	340	12.2.1	总体思路	373
10.3.3	DCGAN 的 CelebA 人脸 图像生成案例	345	12.2.2	训练流程	374
10.4	网中网 NIN	348	12.2.3	搜索过程	377
10.4.1	NIN 的模型结构	348	12.3	AlphaGo Zero 的新思想	380
10.4.2	NIN 的 Caffe 代码 实现及说明	350	12.4	仿效 AlphaGo 的围棋程序 案例 MuGo	383
10.4.3	NIN 大规模图像分类 案例	353	12.4.1	MuGo 的开发环境	383
第 11 章	卷积神经网络的强化 模型	356	12.4.2	MuGo 的代码实现及 说明	386
11.1	强化学习的基本概念	356	12.4.3	MuGo 的学习训练过程	401
11.2	深度强化学习网络的学习算法	358	12.4.4	MuGo 的演示效果	403
11.3	深度强化学习网络的变种模型	359	附录 A	Caffe 在 Windows 上的 安装过程	406
11.4	深度强化学习网络的 Flappy Bird 智能体案例	361	附录 B	Caffe 在 Linux 上的安装 过程	409
11.4.1	笨笨鸟网络的开发 环境和工具包	362	附录 C	TensorFlow 在 Windows 上的安装过程	412
11.4.2	笨笨鸟网络的代码 实现及说明	363	附录 D	TensorFlow 在 Linux 上的安装过程	414
11.4.3	笨笨鸟网络的学习训练 过程	367	参考文献	416	

概 述

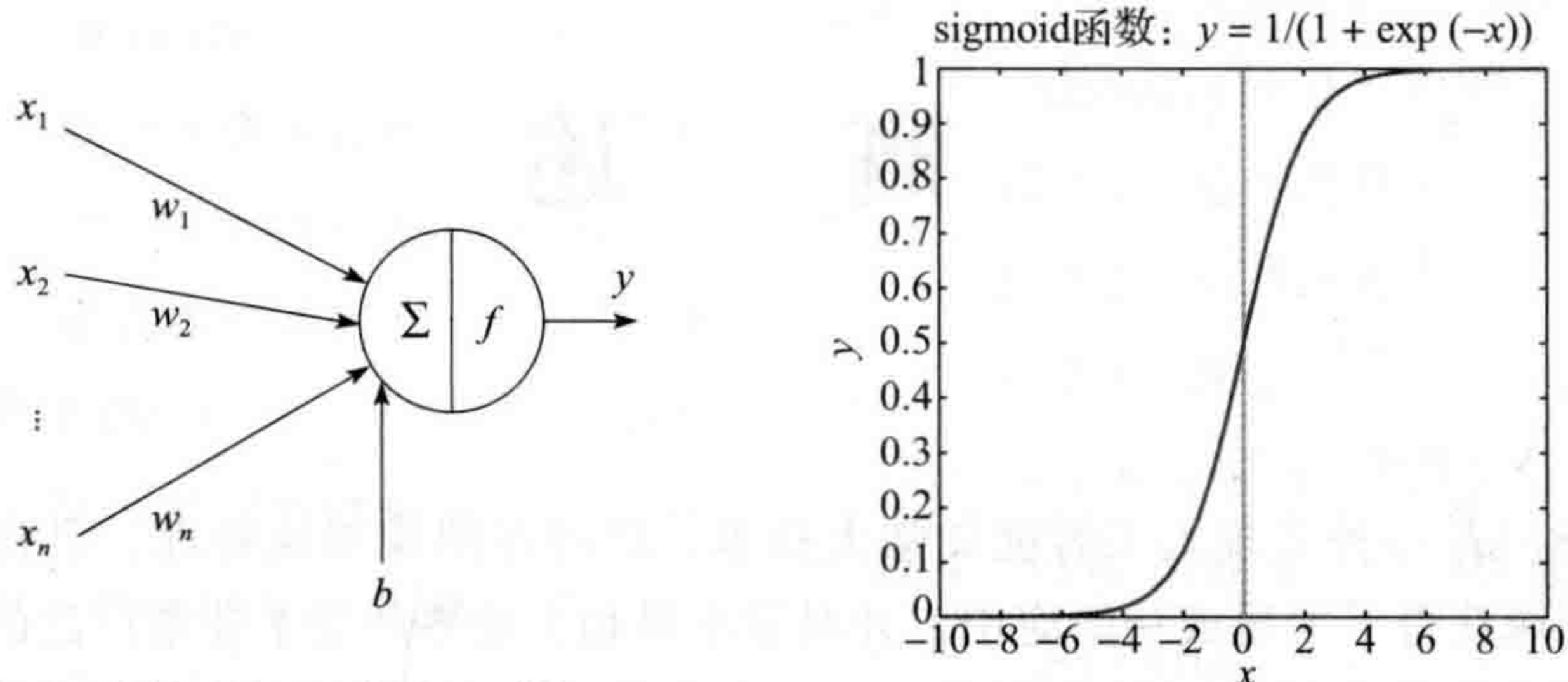
深度学习是一种实现人工智能的强大技术，已经在图像视频处理、语音处理、自然语言处理等领域获得了大量成功的应用，并对学术界和工业界产生了非常广泛的影响。卷积神经网络是深度学习中最为核心的模型，2012 年以来极大地推进了图像分类、识别和理解技术的发展。而且通过与其他技术相结合，卷积神经网络还可用于设计实现游戏智能体 Q 网络、围棋程序 AlphaGo，以及语音识别和机器翻译软件等各种应用系统，所取得的成就已经使人工智能迈进了盛况空前、影响深远的新时代。本章主要介绍深度学习的起源和发展，说明卷积神经网络的形成和演变，分析卷积神经网络的应用和影响，讨论卷积神经网络的缺陷和视图，总结卷积神经网络的平台和工具，并概括本书的内容结构及案例数据。

1.1 深度学习的起源和发展

深度学习的概念起源于人工神经网络，本质上是指一类对具有深层结构的神经网络进行有效训练的方法。神经网络是一种由许多非线性计算单元（或称神经元、节点）组成的分层系统，通常网络的深度就是其中不包括输入层的层数。

最早的神经网络是心理学家 McCulloch 和数理逻辑学家 Pitts 在 1943 年建立的 MP 模型^[1]，如图 1.1 所示。MP 模型实际上只是单个神经元的形式化数学描述，具有执行逻辑运算的功能，虽然不能进行学习，但开创了人工神经网络研究的时代。1949 年，Hebb 首先对生物神经网络提出了有关学习的思想^[2]。1958 年，Rosenblatt 提出了感知器模型及其学习算法^[3]。在随后的几十年间，尽管神经网络的研究出现过一段与 Minsky 对感知器的批评有关的低潮期^[4]，但依然在逐步向前推进，并产生了许多神经网络的新模型^[5-10]。到 20 世纪八九十年代，这些新模型终于引发了神经网络的重生，并掀起了对神经网络研究的世界性高潮^[11]。其中最受欢迎的模型至少包括：Hopfield 神经网络^[8]、波耳兹曼机^[9]和多层感知器^[10]。最早的深度学习系统也许就是那些通过数据分组处理方法训练的多层感知器^[12]。多层感知器，在隐含层数大于 1 时常称为深层感知器，实际上是一种由多层节点有向图构成的前馈神经网络^[13]，其中每一个非

输入节点是具有非线性激活函数的神经元，每一层与其下一层是全连接的。此外，Fukushima提出的神经认知机可能是第一个具有“深度”属性的神经网络^[14-16]，并且也是第一个集成了“感受野”思想的神经网络^[17-18]，以便有效地对视觉输入的某些特性起反应。更重要的是，神经认知机促成了卷积神经网络结构的诞生和发展^[19]。而卷积神经网络作为一种判别模型，极大地推进了图像分类、识别和理解技术的发展，在大规模评测比赛中成绩卓著^[20]，盛誉非凡。



a) MP神经元的数学模型，包含 n 个输入 x_1, x_2, \dots, x_n ， n 个权值 w_1, w_2, \dots, w_n ，1个偏置 b ，以及1个输出 y

b) f 是一个非线性变换，称为激活函数，通常选为sigmoid函数

图 1.1

在训练神经网络方面，反向传播无疑是最常用、最著名的算法^[10, 21]。这是一种有监督学习算法，需要教师指导信号。也就是说，应提供一组训练样本，对给定的输入，指明相应的输出。然而，直到20世纪80年代末期，反向传播似乎还只是对浅层网络有效，尽管原理上也应对深层网络有效。浅层网络主要是指具有1个隐含层的神经网络，如图1.2所示。深层网络则主要是指具有2个及以上隐含层的神经网络，如图1.3所示。在早期的应用中，大多数多层感知器都只用1个或很少的隐含层，增加隐含层几乎没有什么经验上的收益。这似乎可以从神经网络的单隐层感知器逼近定理中找到某种解释^[22, 23]，该定理指出，只要单隐层感知器包含的隐含神经元足够多，就能够在闭区间上以任意精度逼近任何一个多变量连续函数。直到1991年的时候，关于多层感知器在增加层数时为什么难学习的问题，才开始作为一个深度学习的基本问题，得到了完全的理解。

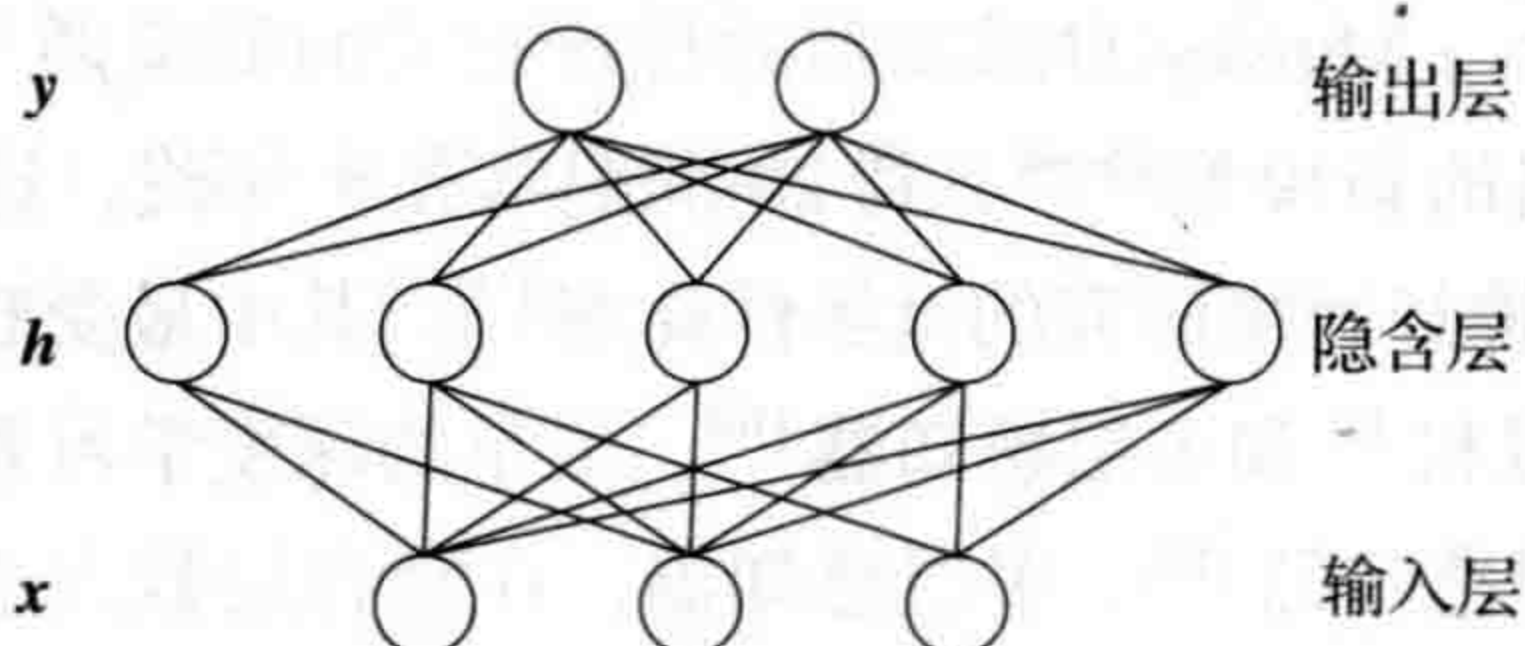


图 1.2 浅层（单隐层）神经网络

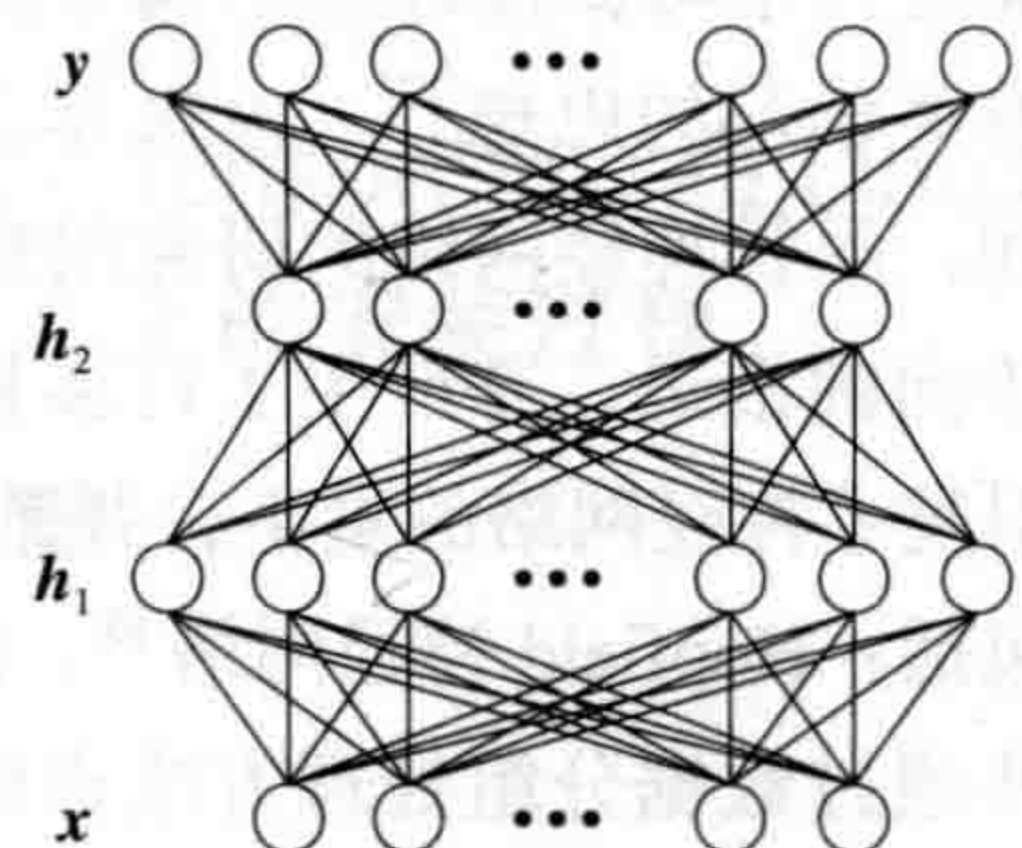


图 1.3 深层神经网络

1991年, Hochreiter正式指出, 典型的深层网络存在梯度消失或爆炸问题, 从而明确确立了深度学习的一个里程碑^[24]。该问题是说, 累积反向传播误差信号在神经网络的层数增加时会出现指数衰减或增长的现象, 从而导致数值计算快速收缩或越界。这就是深层网络很难用反向传播算法训练的主要原因。需要指出的是, 梯度消失或爆炸问题又称为长时滞后问题, 在循环神经网络中也会出现^[25]。

为了在一定程度上克服梯度消失或爆炸问题, 1990 ~ 2000年, Hochreiter的深邃思想推动了若干新方法的探索^[26-28]。但除了卷积神经网络以外^[29], 训练深层网络的问题直到2006年才开始得到严肃认真的对待。一个重要的原因是, 1995年之后支持向量机的快速发展减缓了神经网络的有关工作进展^[30]。

普遍认为, 深度学习正式发端于2006年, 以Hinton及其合作者发表的两篇重要论文为标志: 一篇发表在《Neural Computation》上, 题目为“A fast learning algorithm for deep belief nets”^[31]; 另一篇发表在《Science》上, 题目为“Reducing the dimensionality of data with neural networks”^[32]。从那以后, 大量的深度学习模型开始重新受到广泛关注, 或如雨后春笋般迅速发展起来, 其中主要包括受限波耳兹曼机(Restricted Boltzman Machine, RBM)^[33]、深层自编码器(deep AutoEncoder, deep AE)^[32]、深层信念网络(deep belief net)^[31]、深层波耳兹曼机(Deep Boltzman Machine, DBM)^[34]、和积网络(Sum-Product Network, SPN)^[35]、深层堆叠网络(Deep Stacked Network, DSN)^[36]、卷积神经网络(Convolutional Neural Network, CNN)^[19]、循环神经网络(Recurrent Neural Network, RNN)^[25]、长短期记忆网络(Long Short-Term Memory network, LSTM network)^[27]、强化学习网络(Reinforcement Learning Network, RLN)^[37]、生成对抗网络(Generative Adversarial Network, GAN)^[38]等。通过结合各种有效的训练技巧, 比如最大池化(max pooling)^[39]、丢失输出(dropout)^[40]和丢失连接(dropconnect)^[41], 这些深度学习模型取得了许多历史性的突破和成就, 例如手写数字识别^[32]、ImageNet分类^[20]和语音识别^[42]。而这些历史性的突破和成就, 使深度学习很快在学术界掀起了神经网络的一次新浪潮。其中最主要的原因, 当然是深度学习在解决大量实际问题时所表现的性能超越了机器学习的其他替代方法, 例如支持向量机^[30]。

在理论上, 一个具有浅层结构或层数不够深的神经网络虽然在节点数足够大时也可能充分逼近地表达任意的多元非线性函数, 但这种浅层表达在具体实现时往往由于需要太多的节点而无法实际应用。一般说来, 对于给定数目的训练样本, 如果缺乏其他先验知识, 人们更期望使用少量的计算单元来建立目标函数的“紧表达”, 以获得更好的泛化能力^[43]。而在网络深度不够时, 这种紧表达可能根本无法建立起来, 因为理论研究表明, 深度为 k 的网络能够紧表达的函数在用深度为 $k-1$ 的网络来表达时有时需要的计算单元会呈指数增长^[44]。这种函数表达的潜在能力说明, 深层神经网络(又称深度神经网络)在一定的条件下可能具有非常重要的应用前景。随着深度学习的兴起, 这种潜在能力开始逐步显现出来, 特别是对卷积神经网络的全面推广应用, 使得这种潜在能力几乎得到了淋漓尽致的发挥。

1.2 卷积神经网络的形成和演变

卷积神经网络最初是受到视觉系统的神经机制启发、针对二维形状的认识设计的一种生物物理模型，在平移情况下具有高度不变性，在缩放和倾斜情况下也具有一定的不变性。这种生物物理模型集成了“感受野”的思想，可以看作一种特殊的多层感知器或前馈神经网络，具有局部连接、权值共享的特点，其中大量神经元按照一定方式组织起来对视野中的交叠区域产生反应。1962年，Hubel和Wiesel通过对猫的视觉皮层细胞的研究，提出了感受野的概念^[17-18]。1979年，日本学者Fukushima在感受野概念的基础上，提出了神经认知机模型^[14-16]，该模型被认为是实现的第一个卷积神经网络。1989年，LeCun等人首次使用了权值共享技术^[45]。1998年，LeCun等人将卷积层和下采样层相结合，设计卷积神经网络的主要结构，形成了现代卷积神经网络的雏形（LeNet）^[19]。2012年，卷积神经网络的发展取得了历史性的突破，Krizhevsky等人采用修正线性单元（Rectified Linear Unit, ReLU）作为激活函数提出了著名的AlexNet，并在大规模图像评测中取得了优异成绩^[46]，成为深度学习发展史上的重要拐点。

在理论上，卷积神经网络是一种特殊的多层感知器或前馈神经网络。标准的卷积神经网络一般由输入层、交替的卷积层和池化层、全连接层和输出层构成，如图1.4所示。其中，卷积层也称为“检测层”，“池化层”又称为下采样层，它们可以被看作特殊的隐含层。卷积层的权值也称为卷积核。虽然卷积核一般是需要训练的，但有时也可以是固定的，比如直接采用Gabor滤波器^[47]。作为计算机视觉领域最成功的一种深度学习模型，卷积神经网络在深度学习兴起之后已经通过不断演化产生了大量变种模型。

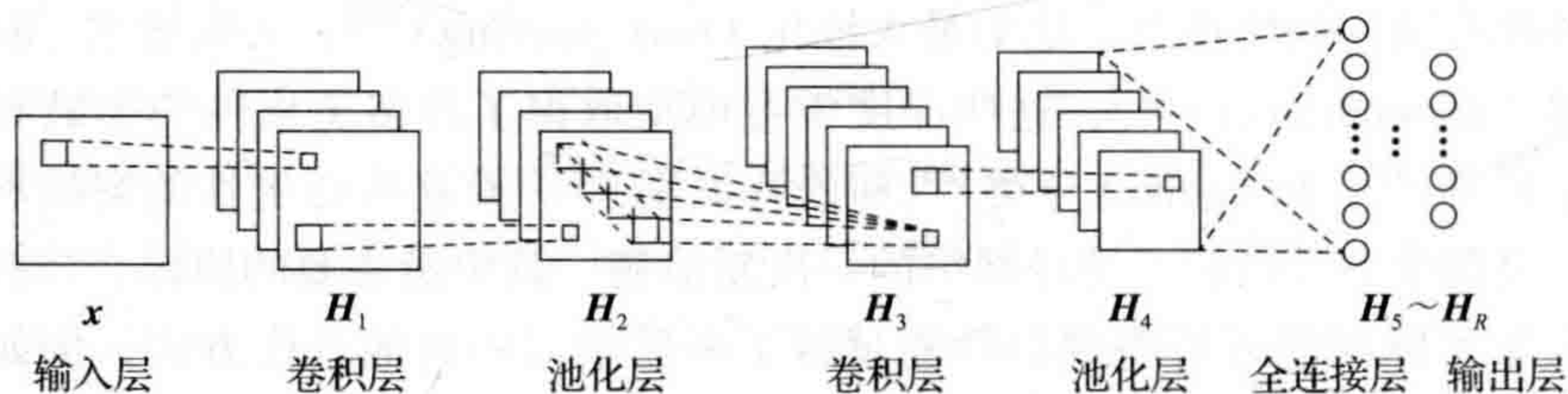


图 1.4 标准卷积神经网络

从结构的角度看，卷积神经网络起初只能处理黑白或灰度图像，变种模型通过把红、绿、蓝3个颜色通道作为一个整体输入已能直接处理彩色图像^[46]，有些还可以直接处理多帧图像甚至连续图像^[48]。同时，变种模型可以使用多个相邻的卷积层或多个相邻的池化层，也可以使用重叠池化和最大池化，还可以使用修正线性单元、渗漏修正线性单元（Leaky ReLU, LReLU）、参数修正线性单元（Parametric ReLU, PReLU）或指数线性单元（Exponential Linear Unit, ELU）取代sigmoid单元作为激活函数^[46, 49-51]，也可以在输出层采用软最大函数softmax替代sigmoid函数以产生伪概率。此外，卷积神经网络可以设计成孪生结构（siamese architecture），把原始数据映射到目标空间，产生对几何扭曲的鲁棒性^[52]。最后，

卷积神经网络可以设计成快道结构，允许信息通过快道无阻碍地跨越多层流动，使得用梯度下降训练非常深的网络变得更加容易^[53]。

从卷积核的角度看，卷积神经网络可以通过采用非常小的卷积核，比如 1×1 和 3×3 大小，被加深成一个更深的网络，比如16层或19层的VGGNet^[54]。如果采用参数修正线性单元代替修正线性单元，可以把VGGNet发展成MSRANet^[55]。而且，卷积神经网络通过使用小卷积核在保持总体计算代价的条件下增加深度和宽度，并与“摄入模块（inception module）”进行集成，可以用来建立谷歌网络（GoogLeNet）^[56]。此外，卷积神经网络通过使用微型多层感知器代替卷积核，还可以被扩展成更为复杂的网络，例如“网中网（Network In Network, NIN）”^[57]。

从区域的角度看，区域卷积神经网络（Region-based CNN, R-CNN）可以用来抽取区域卷积特征，并通过区域提议进行更加鲁棒的定位和分类^[58]。空间金字塔池化网络（Spatial Pyramid Pooling Net, SPPNet）可以克服其输入大小固定的缺点，办法是在最后一个卷积层和第一个全连接层之间插入一个空间金字塔池化层^[59]。不管输入的大小如何，空间金字塔池化层都能够产生固定大小的输出，并使用多尺度空间箱（spatial bin）代替滑动窗口对在不同尺度上抽取的特征进行池化。虽然与R-CNN相比，空间金字塔池化网络具有能够直接输入可变大小图像的优势，但是它们需要一个多阶段的管道把特征写入硬盘，训练过程较为麻烦。为了解决这个训练问题，可以在R-CNN中插入一个特殊的单级空间金字塔池化层（称为感兴趣区池化层，ROI pooling layer），并将其提取的特征向量输入到一个最终分化成两个兄弟输出层的全连接层，再构造一个单阶段多任务损失函数对所有网络层进行整体训练，建立快速区域卷积神经网络（Fast R-CNN）^[60]，其优点是可以通过优化一个单阶段多任务损失函数进行联合训练。为了减少区域提议的选择代价，可以插入一个区域提议网络与Fast R-CNN共享所有卷积层，进一步建立更快速区域卷积神经网络（Faster R-CNN），产生几乎零代价的提议预测对象（或称为目标、物体）边界及有关分数^[61]。为了获得实时性能极快的对象检测速度，可以把输入图像划分成许多网格，并通过单个网络构造的整体检测管道，直接从整幅图像预测对象的边框和类概率建立YOLO模型，只需看一遍图像就能知道对象的位置和类别^[62]。为了更准确地定位对象，还可以在多尺度特征图的每个位置上，使用不同长宽比的缺省框建立单次检测器（SSD）来取代YOLO^[63]。此外，采用空间变换模块有助于卷积神经网络学到对平移、缩放、旋转和其他扭曲更鲁棒的不变性^[64]。最后，可以把Faster R-CNN扩展成掩膜区域卷积神经网络（Mask R-CNN），在图像中有效检测对象的同时，还能够对每个对象实例生成一个高质量的分割掩膜^[65]。

从优化的角度看，许多技术可以用来训练卷积神经网络，比如丢失输出^[40, 66]、丢失连接^[41]和块归一化（batch normalization）^[67]。丢失输出是一种减小过拟合的正则化技术，而丢失连接是丢失输出的推广。块归一化（或批量归一化）则是按迷你块大小对某些层的输入进行归一化处理的方法。此外，残差网络（Residual Network, ResNet）采用跨越2~3层的连接策略也是一种重要的优化技术，可以用来克服极深网络的训练困难。借助残差学习能

够快速有效地成功训练超过 150 层甚至 1000 层的深层卷积神经网络，它在 ILSVRC & COCO 2015 的多项任务评测中发挥了关键作用^[68]，全部取得了第一名的突出成绩。最后，为了优化模型的结构，还可以采用火焰模块（fire module）建立卷积神经网络的挤压模型 SqueezeNet^[69]，也可以结合深度压缩（deep compression）技术进一步减少网络的参数^[70]。

从模型演变的角度看，卷积神经网络的发展脉络如图 1.5 所示。从图中可以看出，现代卷积网络以 LeNet 为雏形，在经过 AlexNet 的历史突破之后，演化生成了很多不同的网络模型，主要包括：加深模型、跨连模型、应变模型、区域模型、分割模型、特殊模型和强化模型等。加深模型的代表是 VGGNet-16、VGGNet-19 和 GoogLeNet；跨连模型的代表是 HighwayNet、ResNet 和 DenseNet；应变模型的代表是 SPPNet；区域模型的代表是 R-CNN、Fast R-CNN、Faster R-CNN、YOLO 和 SSD；分割模型的代表是 FCN、PSPNet 和 Mask R-CNN；特殊模型的代表是 SiameseNet、SqueezeNet、DCGAN、NIN；强化模型的代表是 DQN 和 AlphaGo。

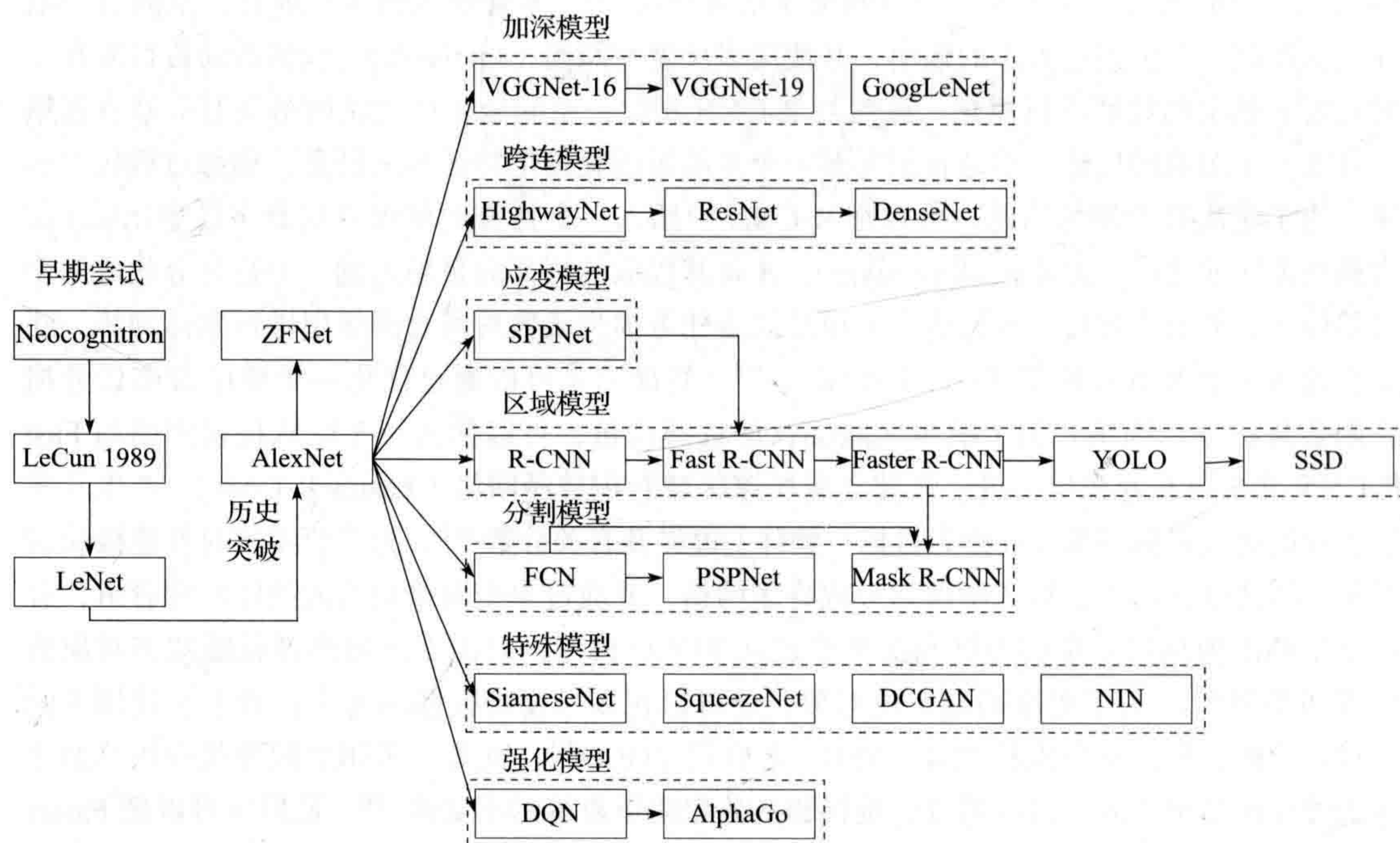


图 1.5 卷积神经网络的演变

1.3 卷积神经网络的应用和影响

自从卷积神经网络在深度学习领域闪亮登场之后，很快取得了突飞猛进的进展，不仅显著提高了手写字符识别的准确率，而且屡屡在图像分类与识别、目标定位与检测等大规模数据评测竞赛中名列前茅、战绩辉煌。此外，卷积神经网络在人脸验证、交通标志识别、视